National Toxicology Program's Report
of the

Organized by the
National Institute of Environmental Health Sciences, NIH
National Toxicology Program

NTP NIEHS

Sponsored by the
U.S. Environmental Protection Agency

and the

National Institute of Environmental Health Sciences, NIH
National Toxicology Program

National Toxicology Program's Report
of the
Endocrine Disruptors Low-Dose Peer Review

August 2001*

*This report represents the views and expert opinions of the Low-Dose Peer Review Panel that met October 10-12, 2000, in Research Triangle Park, NC. Public comments received in response to this report are included as Appendix C.

National Toxicology Program, U.S. Department of Health and Human Services
National Institute of Environmental Health Sciences, National Institutes of Health
Research Triangle Park, NC

This report is available on the National Toxicology Program (NTP) web site:
http://ntp-server.niehs.nih.gov/htdocs/liason/LowDoseWebPage.html
Hard copies of this report can be obtained by contacting:
NTP Office of Liaison and Scientific Review
NIEHS, NIH
P.O. Box 12233, MD A3-01
Research Triangle Park, NC 27709
919-541-0530 (phone)
919-541-0295 (fax)
liaison@starbase.niehs.nih.gov (e-mail)

TABLE OF CONTENTS

Executive Summary

Purpose and Background

At the request of the U.S. Environmental Protection Agency (EPA), the National Toxicology Program (NTP)/National Institute of Environmental Health Sciences (NIEHS) organized and conducted an independent and open peer review aimed at evaluating the scientific evidence on reported low-dose effects and dose-response relationships for endocrine disrupting chemicals in mammalian species that pertain to assessments of effects on human health. The peer review took place in Research Triangle Park, North Carolina, on October 10-12, 2000. The members of the peer review organizing committee are listed in Table 1.

The purpose of this meeting was to establish a sound scientific foundation upon which the U.S. EPA could determine what aspects, if any, of its standard guidelines for reproductive and developmental toxicity testing need to be modified to detect and characterize low-dose effects of endocrine disruptors. Results from this review may also influence how other national and international agencies select doses, endpoints, animal models, and testing regimens for reproductive and developmental studies of endocrine active agents. In particular, the NTP is interested in evaluating the scientific underpinnings of dose-response relationships for reproductive toxicants. For this peer review, "low-dose effects" referred to biological changes that occur in the range of human exposures or at doses that are lower than those typically used in the EPA's standard testing paradigm for evaluating reproductive and developmental toxicity. The U.S. EPA's current recommended methods are described in the document "Health Effects Test Guidelines OPPTS 870.3800 Reproduction and Fertility Effects" (EPA 712-C-98-208, August 1998). The focus of this review was on "biological change" rather than on "adverse effect" because, in many cases, the long-term health consequences of altered endocrine function during development have not been fully characterized.

The peer review panel (the Panel) included individuals from academia, government, and industry with expertise in receptor/molecular biology, experimental and clinical endocrinology, reproductive and developmental toxicology, statistics, and mathematical modeling. The Panel was divided into five subpanels: Bisphenol A, Other Environmental Estrogens and Estradiol, Androgens and Antiandrogens, Biological Factors and Study Design, and Statistics and Dose-Response Modeling. Table 2 identifies the members of each subpanel.

This peer review used a unique and novel approach to resolve a controversial but very important environmental health issue. Fifteen principal investigators of primary research groups active in this field were asked to provide their individual animal data on selected parameters for independent statistical re-analysis by the Statistics Subpanel prior to the meeting. The Organizing Committee requested the raw data on specific parameters in 59 different studies. The selected studies are listed by principal investigator in Table 3 and the requested parameters from each study are given in Table 4. Data were willingly submitted from 49 of the 59 selected studies. In general, the primary reasons that certain requested data sets were not provided was that the data were not available in an electronic format as specified by the Statistics Subpanel or the raw data were in the possession of collaborators and could not be provided in the requested time frame. Studies for which requested data sets were not submitted by principal investigators for independent review by the Statistics Subpanel were used as background information by the Panel. In addition to submitting their raw data, principal investigators

were asked to provide for each study responses to a list of 23 questions (Table 5) on issues relevant to the evaluation of endocrine low-dose studies; these questions addressed animal source and specification, animal husbandry, chemical characterization, administration of test agent, treatment of controls, evaluation of endpoints, and methods of data analysis. Investigators from these research groups were also available at the meeting to give formal presentations of their findings and to have informal discussions with individual subpanels. Because of the extreme rigor of this evaluation process and the extensive analyses of raw data performed by the Statistics Subpanel, unpublished studies were also included in this peer review.

The selected studies included treatments with bisphenol A, diethylstilbestrol, ethinyl estradiol, nonylphenol, octylphenol, genistein, methoxychlor, 17β–estradiol, and vinclozolin, or effects of diet or intrauterine position. Exposure periods included *in utero*, neonatal, pubertal, adult, *in utero* through neonatal, *in utero* through puberty, and *in utero* through adult. Requested parameters included organ weights (prostate, testis, epididymis, seminal vesicle, preputial gland, uterus, and ovary), perinatal measures (e.g., anogenital distance), pubertal measures (e.g., age at vaginal opening, first estrus, preputial separation, and testis descent), and other relevant factors (e.g., daily sperm production, sperm count, serum hormone levels, lymphocyte proliferation in response to anti-CD3, histopathology, estrous cyclicity, receptor binding, estrogen receptor levels, gene expression, and volume of sexually dimorphic nuclei of the preoptic area of the hypothalamus). To conduct this evaluation within a reasonable time frame, the focus of this review was on reproductive and developmental effects. The extensive literature on dioxin and dioxin-like compounds was excluded because EPA was finalizing its extensive and rigorous reevaluation of dioxin risk. Phthalate esters were also excluded because separate evaluations on these compounds were being conducted by the NTP Center for the Evaluation of Risks to Human Reproduction. A future workshop may focus on low-dose effects of dioxin-like compounds.

The Statistics Subpanel analyzed the raw data from 39 of the 49 submitted studies over a 6-week period and provided results from these analyses to the other subpanels prior to the peer review meeting. These analyses provide greater insight on the experimental data than is typically apparent in most peer-reviewed research articles, consequently, the statisticians' report was critical for each of the subpanel reviews. The Dose-Response Modeling group provided theoretical dose-response models based on mechanisms of receptor-mediated processes, as well as empirical dose-response models of endocrine-related effects prior to the meeting. Several important statistical issues were identified by the subpanel and are addressed in their report; these include study sensitivity (power), adjustment for litter effects, pooling of control groups, exclusion of statistical outliers, accounting for body weight differences on organ weight effects, appropriateness of the selected statistical methodology, and data heterogeneity across dose groups. All of these matters, plus experimental design and conduct issues, were taken into consideration by each of the subpanels in their evaluations of the individual studies during the peer review. The statisticians and modelers participated in the other subpanel reviews to ensure that their analyses and models were appropriately used by the subpanels.

The Panel evaluated data from the major, selected studies that support the presence or absence of low-dose effects in laboratory animals and that would be relevant for human health assessments. Low-dose effects analyzed by the Panel should be considered as effects occurring at NOELs (no-observed-effect levels) since this review did not distinguish adverse versus non-adverse effects.

However, the Panel did compare, when appropriate, its analyses to existing NOAELs (no-observed-adverse-effect levels) or LOAELs (lowest-observed-adverse-effect levels) reported by EPA or others. The Panel was also asked to consider biological and mechanistic data that might influence the plausibility of low-dose effects and to identify study design issues or other biological factors that might account for differences in study outcomes. Conclusions from the Panel on the existence of low-dose effects and the shape of the dose-response curve for endocrine active substances in the low-dose region were based on the totality of available knowledge. The specific questions and issues formulated by the Organizing Committee for the subpanels to address in this peer review are given in Table 6.

This unique scientific peer review provided an extraordinarily rigorous, open, transparent, and objective evaluation of the scientific evidence showing the presence or absence of low-dose effects of endocrine disrupting agents and an opportunity for participation by all stakeholders. The subpanels' independently prepared reports follow the Executive Summary. Highlights of the subpanels' findings are given below.

Peer Review Subpanel Findings

Bisphenol A

Based on EPA's estimate that the lowest-observed-adverse-effect level (LOAEL) for oral exposure to bisphenol A in rats is 50 mg/kg/day, the Subpanel used 5 mg/kg/day as a cutoff dose for low-dose effects, regardless of the route or duration of exposure or the age/life stage at which exposure occurred.

- Several studies provide credible evidence for low-dose effects of bisphenol A; these include increased prostate weight in male mice at six months of age and advanced puberty in female mice after *in utero* exposure to 2 or 20 µg/kg/day, and low-dose effects on uterine growth and serum prolactin levels that occurred in F344 rats but not in Sprague-Dawley rats exposed to 0.5 mg/kg/day. The latter findings demonstrate a clear difference in sensitivity to the estrogenic effects of bisphenol A in these two strains of rats.

- Several large studies in rats and mice, including multigenerational studies in Sprague-Dawley rats, found no evidence for a low-dose effect of bisphenol A despite the considerable strength and statistical power those studies represent.

- For those studies that included DES exposure groups, those that showed an effect with bisphenol A showed a similar low-dose effect with DES (e.g., prostate and uterus enlargement in mice), while those that showed no effect with bisphenol A also found no effect with DES.

- Discrepancies in experimental outcome among studies showing positive and negative effects of bisphenol A may have been due to different diets with differing background levels of phytoestrogens, differences in strains of animals that were used, differences in dosing regimen, and differences in housing of animals (singly versus group). Although some studies attempted to replicate previous findings, body weights and prostate weights of controls differed between these studies. Studies also differed in the extent of analysis of dosing solutions.

- The Subpanel concluded that "there is credible evidence that low doses of BPA [bisphenol A] can cause effects on specific endpoints. However, due to the inability of other credible studies in several different laboratories to observe low dose effects of BPA, and the consistency of these negative studies, the Subpanel is not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding."

- Data are insufficient to establish the shape of the dose-response curve for bisphenol A in the low dose region, and the mechanism and biological relevance of reported low dose effects are unclear.

- The Subpanel identified areas for additional research that would clarify uncertainties about low-dose effects of bisphenol A, these include:
  1) additional low-dose studies, including the development and use of sensitive and easily measured molecular endpoints, following *in utero* or early neonatal exposure to conclusively establish low-dose effects of bisphenol A as a general, reproducible phenomenon;
  2) pharmacokinetic data in multiple species and strains of animals to characterize fetal uptake, metabolism, and elimination of bisphenol A and its metabolites;
  3) mechanistic data on estrogen receptor occupancy during critical periods of development, effects of specific receptor antagonists, and responses in estrogen-receptor knock-out mice;
  4) additional studies on intrauterine position effects;
  5) characterization of genetic and epigenetic factors that affect responses to bisphenol A and hormones in general, e.g., factors that lead to strain and species differences in sensitivity;
  6) mechanistic studies on the effects of bisphenol A on regulation of transcriptional activity, from gestation through adulthood.

Other Environmental Estrogens and Estradiol

The subpanel developed an operational definition for "low-dose effects" that was based on the dose-response data for the selected endpoints for each agent under evaluation. Low-dose effects were considered to be occurring when a nonmonotonic dose-response resulted in significant effects below the presumed NOEL expected by the traditional testing paradigm.

- Low-dose effects were clearly demonstrated for estradiol and several other estrogenic compounds. The shape of the dose-response curves for effects of estrogenic compounds varies with the endpoint and the dosing regimen. Theoretical models based on mechanisms of receptor-mediated processes, as well as empirical models of endocrine-related effects, produced dose-response shapes that were either low-dose linear, or threshold-appearing, or non-monotonic (e.g., U-shaped or inverted U-shaped). Low-dose effects of the estrogenic agents evaluated by the Subpanel include the following:

- Estradiol (ovarian steroid with greatest estrogenic activity) - Low-dose effects include changes in serum prolactin, LH, and FSH in ovariectomized rats at a dose of approximately 3 μg/kg/day.

- Diethylstilbestrol (DES, a non-steroidal synthetic estrogen that had been used to prevent spontaneous abortions and to enhance cattle weight gain) - DES is a transplacental carcinogen in

humans. There is clear evidence of a low-dose effect on prostate size by DES (at 0.02 μg/kg) in mice.

- Genistein (isoflavone derived from soy) - Low dose effects were observed in F1 offspring following dietary exposure to 25 ppm, these include a decrease in the volume of sexually dimorphic nuclei of the preoptic area (SDN-POA) of the hypothalamus in male rats (approaching female-like volumes), changes in mammary gland tissue in male rats, and an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3.

- Methoxychlor (insecticide) - Classic estrogenic activity occurs in F1 rats following *in utero* and perinatal exposure to 5 mg/kg/day or higher doses. Low-dose immune system effects occur in F1 offspring following dietary exposure to 10 ppm methoxychlor (approximately equal to 1 mg/kg/day).

- Nonylphenol (industrial compound identified in drinking water supplies) – Low-dose effects in F1 rats following dietary exposure to 25 ppm include a decrease in SDN-POA in males, an increase in relative thymus weight, an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3, and a prolonged estrus in females.

- Octylphenol (an intermediate for the production of surfactants) - There was no evidence of low-dose effects in a five-dose multigeneration study in rats.

- Areas of future research include:
  1) multiple dose studies and modeling of dose-response relationships,
  2) need for replication of low-dose findings in other studies or in other laboratories,
  3) determination of the toxicological significance of volume changes in SDN-POA in male rats and the relationship between estrogenic activity and stimulation of lymphocyte proliferation.

Androgens and Antiandrogens

The Subpanel's review focused on low-dose effects of vinclozolin, a fungicide that is an androgen receptor antagonist. NOAELs for vinclozolin were established from studies in rats; these levels are 6 mg/kg/day for acute dietary exposure and 1.2 mg/kg/day from chronic dietary exposure. No studies have been conducted on vinclozolin at doses below its NOAEL.

- Exposure of pregnant rats to vinclozolin at six doses ranging from 3.125 to 100 mg/kg/day results in reduced anogenital distance (female-like), increased incidence of areolas, and permanently reduced ventral prostate weight in male offspring. For these effects, the dose-response curves appeared linear to the lowest dose tested. Reproductive tract malformations and reduced ejaculated sperm numbers were observed only at the two highest doses. Thus, dose-response relationships are not equivalent among endpoints affected by exposure to vinclozolin.

- Antiandrogens have been shown to act as androgen receptor antagonists, inhibitors of 5α–reductase activity, and/or inhibitors of steroidogenesis. In addition to vinclozolin, other agents (or their metabolites) that have been identified as antiandrogens include p,p'-DDT (insecticide), flutamide and Casodex (pharmaceuticals developed to treat prostate cancer), finasteride (pharmaceuticals

developed to treat benign prostate hyperplasia), methoxychlor (pesticide), procymidone (fungicide), linuron (herbicide), ketoconazole (fungicide), and certain phthalate esters (plasticizers). For finasteride, which acts as a 5α–reductase inhibitor, the dose-response for reduction in anogenital distance (linear) was different than that for increased hypospadias (threshold-appearing).

- There are no data available on low-dose effects of environmental chemicals that act as androgen mimics.

- Future research needs include the following:
  1) further testing of the hypothesis that the dose-response for antiandrogens is linear to the NOAEL/LOAEL,
  2) development of mechanism-based assays for the detection of androgen mimics,
  3) development and utilization of molecular and biochemical markers as sensitive indicators of low-dose effects of androgenic and antiandrogenic agents,
  4) characterization of dose-response relationships for androgenic and antiandrogenic agents in different species and in multiple strains,
  5) development of dosimetry/mechanistic models for exposures occurring during *in utero* and early neonatal development.

Biological Factors and Study Design

- Several factors may account for discrepant findings on low-dose effects of particular endocrine active agents, these include:
  1) intrauterine position, although not essential for the detection of low-dose effects, may be important in evaluating variability in response;
  2) strain and substrain differences in response, which could occur due to genetic differences or selective breeding to maintain high rates of fecundity and growth;
  3) diet with varying background levels of phytoestrogens and differences in caloric intake might influence reproductive parameters;
  4) differences in caging (e.g., stainless steel, polycarbonate), bedding material, or housing (group versus individual) could influence study outcomes;
  5) seasonal variation, which has been reported to affect sex ratios in rodents.

- Comments on the multigeneration test. The traditional multigeneration reproduction study protocol includes exposure of animals through most critical windows of sexual differentiation in the F1 generation and an assessment of the F2 generation through postnatal day 21. This protocol provides substantial information on reproductive effects, but limited information on developmental effects. Frequently, litter size is reduced on postnatal day 4 (usually to 4 males and 4 females) and litter size is further reduced at weaning (postnatal day 21) so that only one animal/sex/litter is held until adulthood. The reduction in number of treated animals evaluated may provide inadequate power to detect low incidence responses (e.g., reproductive tract malformations). Further, a number of sensitive or subtle endocrine-related endpoints are not routinely evaluated, and evaluations of F2 pups on or around postnatal day 21 may not reveal effects on reproductive tract organs that are not yet fully developed. This concern is underscored by the fact that certain endocrine active chemicals were negative in standard multigeneration and prenatal studies.

- Additional design factors for future studies:
  1) Because of clear species and strain differences in sensitivity, animal model selection should be based on responsiveness to endocrine active agents of concern (i.e. responsive to positive controls), not on convenience and familiarity.
  2) Pharmacokinetic data need to be routinely generated, using appropriately sensitive methods, to characterize the dosimetry of the test chemical or its metabolites in target tissues.
  3) Caution is needed in implementing experimental designs to reduce animal variability (e.g., controlled feeding, individual housing), because factors such as body weight and stress can influence reproductive endpoints.
  4) The biological/toxicological relevance of specific endpoints affected by endocrine active agents would benefit from measuring functional parameters or collecting mechanistic data on related biomarkers of effect.
  5) The long-term health consequences of early changes induced by endocrine active agents, e.g., prostate enlargement or accelerated uterine development, need to be determined.
  6) Windows of susceptibility to endocrine disrupting chemicals need to be identified from mechanistic data, and empirical tests need to include exposures at those times.

Overall Conclusions

- Low-dose effects, as defined for this review, were demonstrated in laboratory animals exposed to certain endocrine active agents. The effects are dependent on the compound studied and the endpoint measured. In some cases where low-dose effects have been reported, the findings have not been replicated. The toxicological significance of many of these effects has not been determined.

- The shape of the dose-response curves for these effects varies with the endpoint and dosing regimen, and may be low-dose linear, threshold-appearing, or non-monotonic.

- The traditional multigeneration reproduction study protocol has not revealed major reproductive or developmental effects in laboratory animals exposed to endocrine active agents at doses approaching their NOAELs set by the standard testing paradigm. However, few multigenerational studies have been conducted over expanded dose ranges, and endpoints such as cancer of reproductive organs or neurobehavioral effects are generally not evaluated in multigenerational studies.

- The Panel recommended additional research to replicate previously reported key low-dose findings, to characterize target tissue dosimetry during critical periods of development, to identify sensitive molecular markers that would be useful in understanding mechanistic events associated with low-dose effects, and to determine the long-term health consequences of low-dose effects of endocrine active agents.

- The findings of the Panel indicate that the current testing paradigm used for assessments of reproductive and developmental toxicity should be revisited to see if changes are needed regarding dose selection, animal model selection, age when animals are evaluated, and the endpoints being measured following exposure to endocrine active agents.

Table 1. Peer Review Organizing Committee

William Allaben
National Center for Toxicological Research
Food and Drug Administration

Christopher De Rosa
Agency for Toxic Substances and Disease Registry

Penny Fenner-Crisp
US Environmental Protection Agency, currently at International Life Sciences Institute

Lynn Goldman
Johns Hopkins University

Sandra Inkster
US Consumer Products Safety Commission

Jim Kariya
US Environmental Protection Agency

Robert Kavlock
US Environmental Protection Agency

George Lucier
National Institute of Environmental Health Sciences (retired)
National Institutes of Health

Ronald Melnick (Chair)
National Institute of Environmental Health Sciences
National Institutes of Health

Eisuke Murono
Centers for Disease Control and Prevention

Mary Wolfe
National Institute of Environmental Health Sciences
National Institutes of Health

Roxanne Hall (meeting coordinator)
National Institute of Environmental Health Sciences
National Institutes of Health

Table 2. Subpanels: Chairs, Rapporteurs, Panelists

Bisphenol A

George Stancel (Chair), University of Texas at Houston
Gail Prins (Rapporteur), University of Illinois at Chicago
Ralph Cooper, US Environmental Protection Agency
Warren Foster, Health Canada
Jun Kanno, National Institute of Health Sciences – Japan
John Faust, California Environmental Protection Agency

Other Environmental Estrogens and Estradiol

Michael Gallo (Chair), UMDNJ-Robert Wood Johnson Medical School
Kenneth Reuhl, (Rapporteur), Rutgers University
Mari Golub, California Environmental Protection Agency
Claude Hughes, UCLA School of Medicine
Richard Lyttle, Wyeth-Ayerst Research
Lynne McGrath, Schering-Plough Research Institute
Patricia Whitten, Emory University

Androgens and Antiandrogens

Shuk-Mei Ho (Chair), University of Massachusetts Medical School
Terry Brown (Rapporteur), Johns Hopkins University School of Public Health
George Daston, The Procter & Gamble Company
Mitch Eddy, National Institute of Environmental Health Sciences
Lorenz Rhomberg, Gradient Corporation
Elizabeth Wilson, University of North Carolina at Chapel Hill

Biological Factors and Study Design

John Moore (Chair), Sciences International, Inc.
Julian Leakey (Rapporteur), National Center for Toxicological Research
Sue Barlow, Consultant
Paul Foster, Chemical Industry Institute of Toxicology
Robert Luebke, US Environmental Protection Agency
Robert Maronpot, National Institute of Environmental Health Sciences
Cory Teuscher, University of Illinois at Urbana-Champaign

Statistics and Dose-Response Modeling

Joseph Haseman (Co-chair, Statistics), National Institute of Environmental Health Sciences
John Bailer, Miami Universtiy of Ohio
Ralph Kodell, National Center for Toxicological Research
Richard Morris, Analytical Sciences, Inc.
Kenneth Portier, University of Florida
Michael Kohn (Co-chair, Modeling), National Institute of Environmental Health Sciences
Hugh Barton, US Environmental Protection Agency
Jim Cogliano, US Environmental Protection Agency
Rory Connolly, Chemical Industry Institute of Toxicology
Robert Delongchamp, National Center for Toxicological Research

Table 3. Selected Studies from Principal Investigators

John Ashby, Zeneca Central Toxicological Laboratory, United Kingdom

1. Ashby, J., H. Tinwell, et al. (1999). "Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate glad of CF1 mice exposed in utero." Regulatory Toxicology and Pharmacology **30**: 156-166.

2. Ashby, J., H. Tinwell, et al. (2000). "Current issues in Mutation Research. DNA adducts, estrogenicity and rodent diets." Mutation Research **(in press)**.

3. Ashby, J., H. Tinwell, et al. (2000). "Uterotrophic activity of a "phytoestrogen-free" rat diet." Environmental Health Perspectives **108**(1): A12-A13.

4. Ashby, J., H. Tinwell. (2000). "Activity of bisphenol A in pregnant SD and Alpk rats: preliminary data." **(Unpublished Abstract)**.

5. Odum, J., P. A. Lefevre, et al. (1997). "The rodent uterotrophic assay: critical protocol features, studies with nonylphenols and comparison with a yeast estrogenicity assay." Regulatory Toxicology and Pharmacology **25**: 176-188.

6. Odum, J. and J. Ashby (1999). "Neonatal exposure of male rats to nonylphenol has no effect on the reproductive tract." Toxicological Science **(in press)**.

7. Odum, J., I. T. G. Pyrah, et al. (1999). "Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays." Regulatory Toxicology and Pharmacology **29**: 184-195.

8. Odum, J., I. T. G. Pyrah, et al. (1999). "Effects of p-nonylphenol and diethylstilbestrol on the alderley park rat: comparison of mammary gland and uterus sensitivity following oral gavage or implanted mini-pumps." Journal of Applied Toxicology **19**: 367-378.

9. Tinwell, H., R. Joiner, et al. (2000). "Uterotrophic activity of bisphenol A in the immature mouse." Regulatory Toxicology and Pharmacology **(in press)**.

Barry Delclos, National Center for Toxicological Research

1. Delclos, K. B., T. J. Bucci, et al. (2000). "Effects of dietary genistein exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

2. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary ethinyl estradiol exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

3. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

4. Germolec, D. R., J. A. Munson, et al. (2000). “Immunotoxicity of genistein in male and female Sprague Dawley rats.” **(Unpublished Final Report-Executive Summary)**.

5. Germolec, D. R., J. A. Munson, et al. (2000). “Immunotoxicity of methoxychlor in male and female Sprague Dawley rats.” **(Unpublished Final Report-Executive Summary)**.

6. Germolec, D. R., J. A. Munson, et al. (2000). “Immunotoxicity of nonylphenol in male and female Sprague Dawley rats.” **(Unpublished Final Report-Executive Summary)**.

7. Laurenzana, E. M., C. C. Weis, et al. (2000). “Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats.” In prep **(Unpublished Abstract)**.

8. Laurenzana, E. M., C. C. Weis, et al. (2000). “Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression.” In prep **(Unpublished Abstract)**.

9. Meredith, J. M., C. Bennett, et al. (2000). “Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats.” Society for Neuroscience Abstracts **(in press)**.

10. Scallet, A. C., C. Bennett, et al. (1999). “Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure.” Society for Neuroscience Abstracts **25**: 227.

John O’Connor, DuPont Haskell Laboratory

1. Biegel, L. B., J. C. Cook, et al. (1998). “Effects of 17ß-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats.” Toxicological Sciences **44**: 143-154.

2. Biegel, L. B., J. A. Flaws, et al. (1998). “90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17ß-estradiol.” Toxicological Sciences **44**: 116-142.

3. Cook, J. C., L. Johnson, et al. (1998). “Effects of dietary 17ß-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats.” Toxicological Sciences **44**: 155-168.

4. O'Connor, J. C., S. R. Frame, et al. (1998). “Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17ß-estradiol.” Toxicological Sciences **44**: 169-184.

Frederick vom Saal, University of Missouri

1. Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor

in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October.

2. Howdeshell, K. L., A. K. Hotchkiss, et al. (1999). "Exposure to bisphenol A advances puberty." Nature **401**: 763-764.

3. Howdeshell, K. L. and F. S. vom Saal (2000). "Developmental exposure to bisphenol A: interaction with endogenous estradiol during pregnancy in mice." American Zoologist **40**(3). (in press).

4. Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives **105**(1): 70-76.

5. Palanza, P., S. Parmigiani, et al. (1999). "Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and o,p'-DDT alters aggressive behavior of male and female house mice." Pharmacology Biochemistry and Behavior **64**(4): 665-672.

6. Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." **(in press)**.

7. Timms, B. G., S. L. Petersen, et al. (1999). "Prostate gland growth during development is stimulated in both male and female rat fetuses by intrauterine proximity to female fetuses." Journal of Urology **161**: 1694-1701.

8. vom Saal, F. S., B. G. Timms, et al. (1997). "Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses." Proceedings of the National Academy of Sciences **94**: 2056-2061.

9. vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health **14** (1/2): 239-260.

10. vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November.
11. Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health **15**: 12-25.

Nira Ben-Jonathan, University of Cincinnati

1. Khurana, S., S. Ranmal, et al. (2000). "Exposure of newborn male and female rats to environmental estrogens: delayed and sustained hyperprolactinemia and alterations in estrogen receptor expression." Endocrinology **(in press)**.

2. Long, X., R. Steinmetz, et al. (2000). "Strain differences in vaginal responses to the xenoestrogen bisphenol A." Environmental Health Perspectives **108**(3): 243-247.

3. Steinmetz, R., N. G. Brown, et al. (1997). "The environmental estrogen bisphenol A stimulates prolactin release in vitro and in vivo." Endocrinology **138**(5): 1780-1786.

4. Steinmetz, R., N. Mitchner, et al. (1998). "The xenoestrogen bisphenol A induces growth, differentiation and c-fos gene expression in the female reproductive tract." Endocrinology **139**(6): 2741-2747.

Ibrahim Chahoud, Freie Universitaet Berlin

1. Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." **(3 Abstracts)**.

Robert Chapin, DuPont Haskell Laboratory, formerly at National Institute of Environmental Health Sciences

1. Chapin, R. E., M. W. Harris, et al. (1997). "The effects of perinatal/juvenile methoxychlor exposure on adult rat nervous, immune, and reproductive system function." Fundamental and Applied Toxicology **40**: 138-157.

2. Chapin, R. E. (1999). Study of rats exposed to methoxychlor. The Toxicology Forum: dose-response considerations for potential endocrine active substances, Washington, DC, April.

3. Chapin, R. E., J. Delaney, et al. (1999). "The effects of 4-nonylphenol in rats: a multigeneration reproduction study." Toxicological Sciences **52**: 80-91.

Makoto Ema, National Institute of Health Science - Japan

1. Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." **(Unpublished Study Report)**.

Earl Gray, U.S. Environmental Protection Agency

1. Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health **15**: 48-64.

2. Wolf, C., J. Ostby, et al. (2000). "Effects of prenatal testosterone propionate on the sexual development of male and female rats: a dose-response study." Biology of Reproduction **62**(Supplement): 247.

Ping Lee, Medical College of Wisconsin

1. Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine **9**(1): 105-111.

Retha Newbold, National Institute of Environmental Health Sciences

1. Newbold, R. (1995). "Cellular and molecular effects of developmental exposure to diethylstilbestrol: implications for other environmental estrogens." Environmental Health Perspectives **103**(7): 83-87.

2. Newbold, R. R., W. N. Jefferson, et al. (1997). "Uterine carcinoma in mice treated neonatally with Tamoxifen." Carcinogenesis **18**(12): 2293-2298.

3. Newbold, R. R., R. B. Hanson, et al. (1998). "Increased tumors but uncompromised fertility in the female descendants of mice exposed developmentally to diethylstilbestrol." Carcinogenesis **19**(9): 1655-1663.

4. Newbold, R. R., R. B. Hanson, et al. (2000). "Proliferative lesions and reproductive tract tumors in male descendants of mice exposed developmentally to diethylstilbestrol." Carcinogenesis **21**(7): 1355-1363.

5. Newbold, R. R. (2000). "Dose related changes in male reproductive tract tissues following prenatal exposure to diethylstilbestrol." **(Unpublished Abstract)**.

6. Newbold, R. R., E. P. Banks, et al. (2000). "Low doses of diethylstilbestrol during development result in permanent alterations in the reproductive tract." **(Unpublished Abstract)**.

Jimmy Spearow, University of California at Davis

1. Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science **285**: 1259-1261.

2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August.

Rochelle Tyl, Research Triangle Institute

1. Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology **30**: 81-95.

2. Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 **(Draft Final Report)**.

John Waechter, Dow Chemical Company

1. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences **50**: 36-44.

2. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology **30**: 130-139.

Frank Welsch, Chemical Industry Institute of Toxicology

1. Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology **(in press)**.

2. Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences **54**(Supplement): 256A.

3. Welsch, F., B. A. Elswick, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A on female offspring of Sprague-Dawley rats." Toxicological Sciences **54**(Supplement): 256A.

4. Welsch, F., B. A. Elswick, et al. (2000). "Lack of effects of perinatal exposure to low doses of bisphenol A on male rat offspring ventral prostate glands." In prep **(Unpublished Abstract)**.

Table 4. Selected Studies: Requested Parameters

(Studies shown in **bold** indicate data were received by the Statistics Subpanel)

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
1	**Ashby, John**	**Ashby, J., H. Tinwell, et al. (1999). "Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate glad of CF1 mice exposed in utero." Regulatory Toxicology and Pharmacology 30: 156-166.**	**Bisphenol A** **DES**	In Utero	Testis, Epididymis, Seminal Vesicle, Prostate Testis, Epididymis, Seminal Vesicle, Prostate		Vaginal Opening Vaginal Opening	Daily Sperm Production Daily Sperm Production
2,3	**Ashby, John**	**Ashby, J., H. Tinwell, et al. (2000). "Current issues in Mutation Research. DNA adducts, estrogenicity and rodent diets." Mutation Research (in press).** **Ashby, J., H. Tinwell, et al. (2000). "Uterotrophic activity of a "phytoestrogen-free" rat diet." Environmental Health Perspectives 108(1): A12-A13.**	**3 diets**	Pubertal	Uterus			
4	**Ashby, John**	**Ashby, J., H. Tinwell. (2000). "Activity of bisphenol A in pregnant SD and Alpk rats: preliminary data." (Unpublished Abstract).**	**Bisphenol A** **Ethinyl Estradiol**	In Utero		Anogenital Distance Anogenital Distance	Vaginal Opening, Preputial Separation Vaginal Opening, Preputial Separation	
5	**Ashby, John**	**Odum, J., P. A. Lefevre, et al. (1997). "The rodent uterotrophic assay: critical protocol features, studies with nonylphenols and comparison with a yeast estrogenicity assay." Regulatory Toxicology and Pharmacology 25: 176-188.**	**Estradiol** **17β-Estradiol Benzoate** **Ethinyl Estradiol** **Nonylphenol**	Pubertal	Uterus Uterus Uterus Uterus		 Vaginal Opening Vaginal Opening Vaginal Opening	

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
6	**Ashby, John**	**Odum, J. and J. Ashby (1999). "Neonatal exposure of male rats to nonylphenol has no effect on the reproductive tract." Toxicological Science (in press).**	**Nonylphenol**	Perinatal	Testis, Epididymis, Seminal Vesicle, Ventral Prostate			
7	**Ashby, John**	**Odum, J., I. T. G. Pyrah, et al. (1999). "Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays." Regulatory Toxicology and Pharmacology 29: 184-195.**	**p-Nonylphenol** **DES**	Pubertal	Uterus Uterus			Mammary Gland Differentiation Mammary Gland Differentiation
8	**Ashby, John**	**Odum, J., I. T. G. Pyrah, et al. (1999). "Effects of p-nonylphenol and diethylstilbestrol on the alderley park rat: comparison of mammary gland and uterus sensitivity following oral gavage or implanted mini-pumps." Journal of Applied Toxicology 19: 367-378.**	**p-Nonylphenol** **DES**	Pubertal	Uterus Uterus			Mammary Gland Differentiation Mammary Gland Differentiation
9	**Ashby, John**	**Tinwell, H., R. Joiner, et al. (2000). "Uterotrophic activity of bisphenol A in the immature mouse." Regulatory Toxicology and Pharmacology (in press).**	**Bisphenol A**	Pubertal	Uterus			Uterine Histopathology Cell Proliferation
1	**Delclos, Barry**	**Delclos, K. B., T. J. Bucci, et al. (2000). "Effects of dietary genistein exposure during development on male and female CD rats." In prep (Unpublished Abstract).**	**Genistein**	In Utero-Adult	Testis, Prostate	Birth weight, Anogenital Distance	Vaginal Opening, Preputial Separation	Mammary Gland Pathology
2	**Delclos, Barry**	**Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary ethinyl estradiol exposure during development on male and female CD rats." In prep (Unpublished Abstract).**	**Ethinyl Estradiol**	In Utero-Adult	Ovary, Testis, Prostate	Birth Weight, Anogenital Distance	Vaginal Opening, Preputial Separation	

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
3	**Delclos, Barry**	**Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep (Unpublished Abstract).**	**Nonylphenol**	In Utero-Puberty	Testis, Epididymis, Seminal Vesicle, Prostate		Vaginal Opening, Preputial Separation	Epididymal Sperm Count
4	**Delclos, Barry**	**Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of genistein in male and female Sprague Dawley rats." (Unpublished Final Report).**	**Genistein**	In Utero-Adult				Proliferation in Response to Anti-CD3 & Bone Marrow Colony Forming Unit Assays
5	**Delclos, Barry**	**Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of methoxychlor in male and female Sprague Dawley rats." (Unpublished Final Report).**	**Methoxychlor**	In Utero-Adult				Proliferation in Response to Anti-CD3 & Bone Marrow Colony Forming Unit Assays
6	**Delclos, Barry**	**Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of nonylphenol in male and female Sprague Dawley rats." (Unpublished Final Report).**	**Nonylphenol**		In Utero-Adult			Proliferation in Response to Anti-CD3 & Bone Marrow Colony Forming Unit Assays
7	**Delclos, Barry**	**Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats." In prep (Unpublished Abstract).**	**Nonylphenol**	In Utero				Neonatal Serum Testosterone
8	**Delclos, Barry**	**Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression." In prep (Unpublished Abstract).**	**Genistein** **Nonylphenol** **Ethinyl Estradiol**	In Utero-Adult				Testosterone Metabolism, ERα Testosterone Metabolism, ERα Testosterone Metabolism, ERα

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
9,10	**Delclos, Barry**	**Meredith, J. M., C. Bennett, et al. (2000). "Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats." Society for Neuroscience Abstracts (in press).** **Scallet, A. C., C. Bennett, et al. (1999). "Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure." Society for Neuroscience Abstracts 25: 227.**	**Nonylphenol** **Genistein** **Ethinyl Estradiol**	In Utero-Adult				Sexually Dimorphic Nucleus of the Preoptic Area of the Hypothalamus (SDN-POA) SDN-POA SDN-POA
1	**O'Connor, John**	**Biegel, L. B., J. C. Cook, et al. (1998). "Effects of 17ß-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats." Toxicological Sciences 44: 143-154.**	**17 β-Estradiol**	Adult; In Utero-Adult				Serum Hormone Concentrations (P1 & F1): Estradiol, Progesterone, Follicle Stimulating Hormone, Luteinizing Hormone, Prolactin Estrous Cyclicity (P1 & F1)
2	**O'Connor, John**	**Biegel, L. B., J. A. Flaws, et al. (1998). "90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17ß-estradiol." Toxicological Sciences 44: 116-142.**	**Estradiol**	Adult; In Utero-Adult	Epididymis, Accessory Sex Glands, Testis, Uterus, Ovary (P1)	Anogenital Distance (F1)	Vaginal Opening, Preputial Separation (F1)	Histopathology: Ovaries and Uterus (P1) Mammary Gland Cell Proliferation (P1 & F1) Uterine Markers (P1 & F1) Reproductive Indices (P1)

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
3	**O'Connor, John**	**Cook, J. C., L. Johnson, et al. (1998). "Effects of dietary 17ß-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats." Toxicological Sciences 44: 155-168.**	**Estradiol**	Adult; In Utero-Adult	Testis, Epididymis			Hormone Serum Concentrations (P1 & F1): Estradiol, Prolactin, Testosterone, Follicle Stimulating Hormone, Luteinizing Hormone Sperm (P1& F1) Sertoli Cells (F1)
4	**O'Connor, John**	**O'Connor, J. C., S. R. Frame, et al. (1998). "Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17ß-estradiol." Toxicological Sciences 44: 169-184.**	**Estradiol**	Adult	Uterus, Testis, Epididymis, Seminal Vesicle, Prostate			Uterine Stromal Cell Proliferation & Epithelial Cell Height Uterine Estrogen Receptor Concentration Female & Male Serum Reproductive Hormone Concentrations: Estradiol, Prolactin, Follicle Stimulating Hormone, Luteinizing Hormone, Testosterone, Dihydro-Testosterone
1	**vom Saal, Fred**	**Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October.**	**DES** **Methoxychlor**	In Utero	Uterus Uterus			

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
2,3	vom Saal, Fred	Howdeshell, K. L., A. K. Hotchkiss, et al. (1999). "Exposure to bisphenol A advances puberty." Nature **401**: 763-764. Howdeshell, K. L. and F. S. vom Saal (2000). "Developmental exposure to bisphenol A: interaction with endogenous estradiol during pregnancy in mice." American Zoologist **40**(3). (in press).	Bisphenol A	In Utero			Vaginal Opening, Time to 1st Estrus	Female Wean Weight, Male Wean Weight
4	**vom Saal, Fred**	**Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives 105(1): 70-76.**	**Bisphenol A** **Octylphenol**	In Utero	Prostate Prostate			
5	vom Saal, Fred	Palanza, P., S. Parmigiani, et al. (1999). "Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and o,p'-DDT alters aggressive behavior of male and female house mice." Pharmacology Biochemistry and Behavior **64**(4): 665-672.	DES DDT	In Utero	Preputial, Testis Preputial, Testis			
6	**vom Saal, Fred**	**Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." (in press).**	**Ethinyl Estradiol**	In Utero	Prostate at 50 days and 5-months-old			Daily Sperm Production at 50 days and 5-months-old
7	vom Saal, Fred	Timms, B. G., S. L. Petersen, et al. (1999). "Prostate gland growth during development is stimulated in both male and female rat fetuses by intrauterine proximity to female fetuses." Journal of Urology **161**: 1694-1701.	None-Intrauterine Position					Prostate Budding (Figure 2)

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
8	vom Saal, Fred	vom Saal, F. S., B. G. Timms, et al. (1997). "Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses." Proceedings of the National Academy of Sciences **94**: 2056-2061.	DES Estradiol	In Utero	Prostate (Figure 4) Prostate (Figure 2)			
9	**vom Saal, Fred**	**vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health 14 (1/2): 239-260.**	**Bisphenol A** **Octylphenol**	In Utero	Seminal Vesicle, Epididymis Testis, Preputial Glands Seminal Vesicle, Epididymis Testis, Preputial Glands			Daily Sperm Production Daily Sperm Production
10	**vom Saal, Fred**	**vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November.**	*None-Intrauterine Position*	In Utero	Prostate			
11	**vom Saal, Fred**	**Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health 15: 12-25.**	None-Intrauterine Position *Methoxychlor*	In Utero	Prostate Prostate, Seminal Vesicles, Testis,			Male Fetal Serum Estradiol Concentration Androgen Binding

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
2	Ben-Jonathan, Nira	Long, X., R. Steinmetz, et al. (2000). "Strain differences in vaginal responses to the xenoestrogen bisphenol A." Environmental Health Perspectives **108**(3): 243-247.	Bisphenol A Estradiol Controls	Adult				DNA synthetic response in vaginal epithelia (F344 & SD rats), ^{3}H-BPA (F344 & SD rats) DNA synthetic response in vaginal epithelia (F344 & SD rats) Estradiol-ER binding (F344 & SD rats)
3	Ben-Jonathan, Nira	Steinmetz, R., N. G. Brown, et al. (1997). "The environmental estrogen bisphenol A stimulates prolactin release in vitro and in vivo." Endocrinology **138**(5): 1780-1786.	β-Estradiol Bisphenol A	Young Adult				Serum Prolactin (F344 & SD rats), Prolactin Regulating Factor (F344 & SD rats) Serum Prolactin (F344 & SD rats), Prolactin Regulating Factor (F344 & SD rats)
4	Ben-Jonathan, Nira	Steinmetz, R., N. Mitchner, et al. (1998). "The xenoestrogen bisphenol A induces growth, differentiation and c-fos gene expression in the female reproductive tract." Endocrinology **139**(6): 2741-2747.	Bisphenol A	Young Adults				Cell Proliferation of Uterus and Vagina, *c-fos* Expression
1	**Chahoud, Ibrahim**	**Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." (3 Abstracts).**	**Bisphenol A** **Ethinyl Estradiol**	In Utero	Prostate Prostate	Anogenital Distance Anogenital Distance	Vaginal Opening, Preputial Separation Vaginal Opening, Preputial Separation	Daily Sperm Productions, Testosterone Levels, and Estrouscycle Daily Sperm Productions, Testosterone Levels, and Estrouscycle

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
1	**Chapin, Robert**	**Chapin, R. E., M. W. Harris, et al. (1997). "The effects of perinatal/juvenile methoxychlor exposure on adult rat nervous, immune, and reproductive system function." Fundamental and Applied Toxicology 40: 138-157.**	**Methoxychlor**	In Utero-Pubertal	Prostate, Testis, Epididymis, Seminal Vesicle, Ovary, Uterus	Anogenital Distance	Vaginal Opening, Preputial Separation	
2	**Chapin, Robert**	**Chapin, R. E. (1999). Study of rats exposed to methoxychlor. The Toxicology Forum: dose-response considerations for potential endocrine active substances, Washington, DC, April.**	**Methoxychlor**	In Utero-Perinatal	Testis, Epididymis, Seminal Vesicle, Prostate			Sperm Count
3	**Chapin, Robert**	**Chapin, R. E., J. Delaney, et al. (1999). "The effects of 4-nonylphenol in rats: a multigeneration reproduction study." Toxicological Sciences 52: 80-91.**	*Nonylphenol*	In Utero-Adult	Vagina, Uterus, Testis, Epididymis, Ventral Prostate (F1,F2, & F3) Ovary (F2)	Anogenital Distance, (F1,F2, & F3)	Vaginal Opening (F1,F2, & F3) Preputial separation (F1) Testis Descent (F2 & F3)	Estrouscycle Length (F1 & F2) Sperm Count (F2)
1	**Ema, Makoto**	**Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." (Unpublished Study Report).**	**Bisphenol A**		Prostate, Testis, Epididymis, Seminal Vesicle, Ovary, Uterus	Anogenital Distance	Vaginal Opening, Preputial Separation	Sperm Count
1	**Gray, Earl**	**Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health 15: 48-64.**	**Vinclozolin**	In Utero-Perinatal	Prostate, Epididymis, Seminal Vesicle, Testis,	Anogenital Distance, Areola/Nipple		Sperm Count, Permanent Nipples, Hypospadias, Agenesis of Ventral Prostate, Ectopic Testis, Epididymal Granulomas, Epididymal Agenesis

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
2	Gray, Earl	Wolf, C., J. Ostby, et al. (2000). "Effects of prenatal testosterone propionate on the sexual development of male and female rats: a dose-response study." Biology of Reproduction 62(Supplement): 247.	*Study not through US EPA clearance. Received abstract.					
1	**Lee, Ping**	**Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine 9(1): 105-111.**	**Nonylphenol**	Perinatal	Testis, Epididymis, Seminal Vesicle, Ventral Prostate	Anogenital Distance		
1	Newbold, Retha	Newbold, R. (1995). "Cellular and molecular effects of developmental exposure to diethylstilbestrol: implications for other environmental estrogens." Environmental Health Perspectives **103**(7): 83-87.	DES	In Utero				Reproductive Capacity
2	**Newbold, Retha**	**Newbold, R. R., W. N. Jefferson, et al. (1997). "Uterine carcinoma in mice treated neonatally with Tamoxifen." Carcinogenesis 18(12): 2293-2298.**	**Tamoxifen**	Perinatal				Uterine Tumors
3	**Newbold, Retha**	**Newbold, R. R., R. B. Hanson, et al. (1998). "Increased tumors but uncompromised fertility in the female descendants of mice exposed developmentally to diethylstilbestrol." Carcinogenesis 19(9): 1655-1663.**	**DES**	In Utero; Perinatal				Uterine Tumors (F2)

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
4	**Newbold, Retha**	**Newbold, R. R., R. B. Hanson, et al. (2000). "Proliferative lesions and reproductive tract tumors in male descendants of mice exposed developmentally to diethylstilbestrol." Carcinogenesis 21(7): 1355-1363.**	**DES**	In Utero; Perinatal				Testis Proliferative Lesions (F2)
5	**Newbold, Retha**	**Newbold, R. R. (2000). "Dose related changes in male reproductive tract tissues following prenatal exposure to diethylstilbestrol." (Unpublished Abstract).**	**DES**	Adult	Seminal Vesicle, Prostate, Testis, Coagulating Gland			
6	**Newbold, Retha**	**Newbold, R. R., E. P. Banks, et al. (2000). "Low doses of diethylstilbestrol during development result in permanent alterations in the reproductive tract." (Unpublished Abstract).**	**DES**	Perinatal	Uterus			Uterine Epithelial Cell Height, Epithelial Cell Number, Gland Number, Induction of Lactoferrin (an estrogen-responsive protein)
1	**Spearow, Jimmy**	**Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science 285: 1259-1261.**	**Estradiol**	Pubertal	Testis in 4 Mouse Strains			Sperm Maturation
2	**Spearow, Jimmy**	**Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August.**	**Estradiol** **Estradiol Benzoate**	Pubertal	Testis, Vesicular Gland Uterus			Sperm Maturation

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
1	**Tyl, Rochelle**	**Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology 30: 81-95.**	**Octylphenol**	Adult; In Utero-Adult	Epididymis, Seminal Vesicle, Prostate, Testis, (F0, F1-day 21 as well, & F2-day 21 as well) Ovary, Uterus (F0 & F1-day 21 as well)	Vaginal Opening, Preputial Separation (F1 & F2)		Daily Sperm production (F0, F1, & F2)
2	**Tyl, Rochelle**	**Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 (Draft Final Report).**	**Bisphenol A**	Adult; In Utero-Adult	Epididymis, Seminal Vesicle, Prostate, Testis, (F0, F1-day 21 as well, & F2-day 21 as well) Ovary, Uterus (F0 & F1-day 21 as well)	Vaginal Opening, Preputial Separation (F1 & F2)		Daily Sperm Production (F0, F1, & F2)
1	**Waechter, John**	**Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences 50: 36-44.**	**Bisphenol A** **DES**	In Utero	Prostate, Seminal Vesicle, Epididymis, Testis Prostate, Seminal Vesicle, Epididymis, Testis			Daily Sperm Production, Sex Ratio Daily Sperm Production, Sex Ratio
2	**Waechter, John**	**Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology 30: 130-139.**	**Bisphenol A** **DES**		Prostate, Seminal Vesicle, Epididymis, Testis Prostate, Seminal Vesicle, Epididymis, Testis			Daily Sperm Production, Sex Ratio Daily Sperm Production, Sex Ratio
1	**Welsch, Frank**	**Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology (in press).**	**Bisphenol A**	In Utero-Perinatal	Prostate			

Selected Studies Ref. #	P.I.	Study	Chemical	Exposure Category	Parameter			
					Organ Weight	Perinatal	Puberty	Other
2	**Welsch, Frank**	**Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A.**	**Bisphenol A** **DES**	In Utero-Perinatal	Prostate, Seminal Vesicle, Epididymis, Testis Prostate, Seminal Vesicle, Epididymis, Testes	Anogenital Distance Anogenital Distance	Preputial Separation Preputial Separation	
3	**Welsch, Frank**	**Welsch, F., B. A. Elswick, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A on female offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A.**	**Bisphenol A** **DES**	In Utero-Perinatal		Anogenital Distance Anogenital Distance	Vaginal Opening, Time to 1st Estrus Vaginal Opening, Time to 1st Estrus	
4	**Welsch, Frank**	**Welsch, F., B. A. Elswick, et al. (2000). "Lack of effects of perinatal exposure to low doses of bisphenol A on male rat offspring ventral prostate glands." In prep (Unpublished Abstract).**	**Bisphenol A**	In Utero-Perinatal	Prostate, Seminal Vesicle, Epididymis, Testis			

Table 5. Issues Relative to the Evaluation of Endocrine Low-Dose Studies

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives		
2) Species, strain, and source of animals		
3) Diet/source		
4) Caging protocols (single or multiple housing)		
5) Assignment of treatment groups to cage location on racks		
6) Bedding/source		
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses		
8) Age and weight of animals at start and end of study		
9) Method of assigning animals to dosed and control groups		
10) Type of control groups? Concurrent with dosed groups?		
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle		
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study		
13) Was there any "culling" of litters? If so, When? How much?		

What was the method of selection? Was any cross fostering done? If so, please provide details		
14) Survival information: were there any early deaths or notable “competing risks”?		
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially “confounding variables” that should be considered in the data analysis		
16) Was the study done in a single “replicate” with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details		
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.		
18) Were animals and tissue samples examined in a blinded fashion?		
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.		
21) What statistical techniques were used to evaluate the data and why?		
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details		

Table 6. Subpanel Questions and Issues

Subpanels: Bisphenol A, Other Environmental Estrogens and Estradiol, Androgens and Antiandrogens

1. What is the extent of empirical evidence demonstrating low-dose effects of chemical X on reproductive and developmental endpoints from studies in mammalian species? For this meeting, "low-dose effects" refer to biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in EPA's standard toxicity testing paradigm. Within and across studies, describe the specificity, consistency, and strength of the evidence with consideration of the timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

2. What is the extent of empirical evidence demonstrating the lack of low-dose effects of chemical X? Within and across studies, describe the specificity, consistency, and strength of the evidence with consideration of the timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

3. If possible, identify differences in study design or biological factors that might account for the observed differences in study outcomes.

4. How do the findings from studies of low-dose effects on reproductive and developmental outcomes using chemical X compare with those for other endocrine active chemicals? Describe the specificity and consistency of the evidence, and as possible, identify the similarities and/or differences in study design, chemical activity, species or strain, etc. that might explain the observed outcomes.

5. Describe the available and relevant pharmacokinetic, biologic, and other mechanistic information that strengthen or weaken the plausibility of low-dose effects. How did this information impact on the subpanel's overall conclusions? Describe the shape of the dose-response curves in the low-dose region using empirical data as well as biologically based dose-response models.

6. Based on the totality of available knowledge, what is the subpanel's overall conclusion regarding whether chemicals can cause hormone-related effects on reproductive and developmental endpoints at doses lower than those typically used in the standard toxicological dose-setting paradigm?

7. Are there specific knowledge gaps for which additional research relative to the low-dose question is needed? If possible, suggest ways to address those gaps.

Subpanel on Biological Factors and Study Design

1. Within the scientific literature positive and negative findings are reported relative to whether endocrine disrupting chemicals have low-dose effects on reproductive and developmental endpoints in mammalian species. For this meeting, "low-dose effects" refer to biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in EPA's standard toxicity testing paradigm. Identify similarities and/or differences in study design (*e.g.*, diet) or biological factors (*e.g.*, strain, intrauterine position) that might explain the observed outcomes. Describe the specificity, consistency, and strength of the evidence with consideration of the timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are divergent data sets equally supported by appropriate statistical analyses?

2. Based on the totality of available knowledge, what is the subpanel's overall conclusion regarding whether chemicals can cause hormone-related effects on reproductive and developmental endpoints at doses lower than those typically used in the standard toxicological dose-setting paradigm? Describe the available and relevant information regarding study design, biological factors, and mechanistic data that strengthens or weakens the plausibility of low-dose effects. How did this information impact on the subpanel's overall conclusions?

3. Are there specific knowledge gaps for which additional research is needed relative to the question of study design and/or biological factors affecting study outcomes? If possible, suggest ways to address those gaps.

Tuesday, October 10, 2000

7:30 AM	**Registration**	**(Foyer, Empire A)**
8:30 AM	**Welcome**	**(Empire ABC)**
	Kenneth Olden, NIEHS and Penelope Fenner-Crisp, U.S. EPA	
	Background and Charge to Panel *Ronald Melnick, NIEHS, Organizing Committee Chair*	
	Review Process *George Lucier, NIEHS, Peer Review Chair*	
9:00 AM	**Body of Knowledge Presentation and Discussion** *John O'Connor, DuPont*	
10:00 AM	**Break**	
10:30 AM	**Body of Knowledge Presentation and Discussion** *Frederick vom Saal, University of Missouri*	
11:30 AM	**Body of Knowledge Presentation and Discussion** *John Ashby, Zeneca, United Kingdom*	
12:30 PM	**Lunch**	
1:30 PM	**Body of Knowledge Presentation and Discussion** *Barry Delclos, National Center for Toxicological Research*	
2:30 PM	**Summary of Other Bodies of Knowledge Studies** *Jim Kariya, U.S. EPA* *Principal Investigators:* *Nira Ben-Jonathan, University of Cincinnati* *Ibrahim Chahoud, Freie Universitat Berlin* *Robert Chapin, DuPont* *Makoto Ema, National Institute of Health Science-Japan* *Earl Gray, U.S. EPA* *Ping Lee, Medical College of Wisconsin* *Retha Newbold, NIEHS* *Jimmy Spearow, University of California at Davis* *Rochelle Tyl, Research Triangle Institute* *John Waechter, Dow Chemical Company* *Frank Welsch, Chemical Industry Institute of Toxicology*	
3:30 PM	**Break**	
3:45 PM	**Summary of Other Bodies of Knowledge** *(continued)*	
4:30 PM	**Report from the Statistics and Dose-Response Modeling Subpanel - Statistical Evaluations** *Joseph Haseman, NIEHS, Subpanel Co-Chair*	
5:00 PM	**Report from the Statistics and Dose-Response Modeling Subpanel - Theoretical Modeling** *Michael Kohn, NIEHS, Subpanel Co-Chair*	
5:20 PM	**Dinner**	
6:50 PM	**Public Comments** *Moderator - Lynn Goldman, Johns Hopkins University School of Public Health*	

8:00 PM	**Subpanels: Initial Meeting**	
	1. Bisphenol A	**(Bull Durham Room)**
	2. Other Environmental Estrogens and Estradiol	**(Royal A)**
	3. Androgens and Anti-Androgens	**(Crown B)**
	4. Biological Factors and Study Design	**(Royal B)**

Wednesday, October 11, 2000

8:30 AM - 5:00 PM	**Subpanel Meetings**	
	1. Bisphenol A	**(Empire C)**
	2. Other Environmental Estrogens and Estradiol	**(Empire D)**
	3. Androgens and Anti-Androgens	**(Empire A)**
	4. Biological Factors and Study Design	**(Empire E)**
10:00 AM	**Break**	
12:00 PM	**Lunch**	
2:30 PM	**Break**	

Thursday, October 12, 2000

8:30 AM	**Subpanel Meetings**	
	1. Bisphenol A	**(Imperial I)**
	2. Other Environmental Estrogens and Estradiol	**(Imperial II)**
	3. Androgens and Anti-Androgens	**(Imperial VI)**
	4. Biological Factors and Study Design	**(Imperial VII)**
10:00 AM	**Break**	
10:30 AM	**Presentation and Discussion of Subpanel Reports** *Moderators - George Lucier, NIEHS and* *Lynn Goldman, Johns Hopkins University* *School of Public Health*	**(Empire DE)**
12:00 PM	**Lunch**	
1:00 PM	**Presentation and Discussion of Subpanel Reports** *(continued)*	
2:30 PM	**Break**	
3:00 PM	**Presentation and Discussion of Subpanel Reports** *(continued)*	
5:00 PM	**Closing Remarks** *Penelope Fenner-Crisp, U.S. EPA* *Christopher Portier, NIEHS*	

Chapter 1:
Report of the Bisphenol A Subpanel

Chair
George Stancel, University of Texas Health Science Center at Houston

Rapporteur
Gail Prins, University of Illinois at Chicago

Facilitator
Penelope Fenner-Crisp, U.S. Environmental Protection Agency

Panelists
Ralph Cooper, U.S. Environmental Protection Agency
Warren Foster, Bureau of Chemical Hazards, Health Canada
Jun Kanno, National Institute of Health Sciences-Japan
John Faust, California Office of Environmental Health Hazard Assessment

Statistics and Dose-Response Modeling Subpanel Representatives
Joseph Haseman, NIEHS (statistics)
Robert Delongchamp, National Center for Toxicological Research (modeling)

Background

Prior to the meeting the members of the Subpanel were provided with a selected set of background references and a set of "selected studies" which had been selected by the members of the organizing committee. The raw data for some, but not all, of the selected studies was provided in advance to the Statistics and Dose-Response Modeling Subpanels who independently analyzed the data. Drs. Joseph Haseman (Statistics Subpanel) of NIEHS and Robert Delongchamp (Modeling Subpanel) of the National Center for Toxicological Research then served as members of the Bisphenol A Subpanel at the meeting.

Based on an analysis of the selected studies, the Subpanel was then asked to address a set of 7 questions.

1. What is the extent of empirical evidence demonstrating low-dose effects of bisphenol A (BPA) on reproductive and developmental endpoints from studies in mammalian species? For this meeting, "low dose effects" refer to biological changes that occur at environmentally relevant exposure levels or a doses that are lower than those typically used in EPA's standard toxicity testing paradigm. Within and across studies, describe the specificity, consistency, and strength of the evidence with consideration of the timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

2. What is the extent of empirical evidence demonstrating the lack of low-dose effects of BPA? Within and across studies, describe the specificity, consistency, and strength of the evidence with consideration of the timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

3. If possible, identify differences in study design or biological factors that might account for the observed differences in study outcomes.

4. How do the findings from studies of low-dose effects on reproductive and developmental outcomes using chemical X compare with those for other endocrine active chemicals? Describe the specificity and consistency of the evidence, and if possible, identify the similarities and/or differences in study design, chemical activity, species or strain, etc. that might explain the observed outcomes.

5. Describe the available and relevant pharmacokinetic, biologic, and other mechanistic information that strengthen or weaken the plausibility of low-dose effects of BPA. How did this information impact the Subpanel's overall conclusions?
   - Describe the shape of the dose-response curves for BPA in the low-dose region using empirical data as well as biologically based dose-response models.

6. Based on the totality of available knowledge, what is the Subpanel's overall conclusion regarding whether BPA can cause hormone-related effects on reproductive and

developmental endpoints at doses lower than those typically used in the standard toxicological dose-setting paradigm?

7. Are there specific knowledge gaps for which additional research relative to the low-dose question for BPA is needed? If possible, suggest ways to address those gaps.

(NOTE: Because the Subpanel was not provided with an explicit level to be considered "low dose", considerable time was spent for an on site discussion about what dose should be used as a cut-off for this level to address questions (1) – (7) above. In the initial discussion, it was noted that a NOAEL had not been found in rodent studies used for setting a reference dose for BPA. A 1982 NTP Technical Report (CAS no. 80-05-7) found "no convincing evidence that bisphenol A was carcinogenic for F344 rats or B6C3F1 mice of either sex". Nevertheless, since the incidence of testis tumors was significantly elevated in the low dose (1000 ppm) male rat group, the EPA apparently used these data to support their conclusion that this dose, which corresponds to an oral dose of 50 mg/kg/day based on typical food consumption rates, represents a LOAEL for BPA Applying the standard use of uncertainty factors of 10 for interspecies variability, intraspecies variability, and subchronic to chronic comparison, the Subpanel originally decided to use the oral reference dose of 50 μg/kg/day as the "low dose" cut-off.

However, after a lengthy Subpanel discussion on this point, members of the Organizing Committee instructed the Subpanel to consider 5 mg/kg/day or less as representing a "low dose" of BPA, and during the course of the meeting instructions were similarly given to consider 1 μg/kg/day or less of diethylstilbestrol (DES) as a low dose of that chemical. [Note added in proof. The Organizing Committee based this directive on the definition of "low dose" used in this peer review as "doses that are lower than those typically used in EPA's standard toxicity testing paradigm". This would generally be interpreted as doses in the range of a NOAEL, or in the absence of a NOAEL, the use of ~ LOAEL/10. Given the oral LOAEL of 50 mg/kg/day in rats as noted above, this approximation corresponds to a value of 5 mg/kg/day, and this was the rationale used by the Organizing Committee.])

It should be emphasized that the Subpanel used the 5 mg/kg/day dose level as instructed by the Organizing Committee to define the low-dose cutoff for BPA, **regardless of the route or duration of administration or the age of the animal used for a particular protocol**. This is an important point, since the concentration of a chemical reached at a tissue site, can vary widely following administration of identical doses by different routes and over different durations. This is especially true when one route is oral and the other is parenteral. The Subpanel also did not distinguish the duration of administration of BPA, i.e., the number of days for which a dose of 5 mg/kg/day was administered, the age at which exposure occurred, or the developmental stage for in utero exposure. Thus, in some cases BPA was administered by injection or silastic implants, while in others it was given via feed, and in others it was given by gavage. Additionally, some exposures were in utero, while others were during neonatal or adult life.)

Question 1. What is the extent of empirical evidence demonstrating low-dose effects of BPA on reproductive and developmental endpoints from studies in mammalian species? Within and across studies, describe the specificity, consistency, and strength of the evidence with considerations of the timing of exposure, when the endpoints were measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

Studies from vom Saal and Colleagues.

Several studies in mice provided evidence for a low dose effect(s) of BPA. These include the following. The report by Nagel et al (1997), *EHP* 105:70-76, demonstrates an increase in absolute prostate weight of CF-1 male mouse offspring at 6 months of age following administration of 2 and 20 µg/kg/day BPA to pregnant mothers. The statistics Subpanel reanalyzed this data and found it to be significant at the level of $p < 0.05$, and the BPA Subpanel found this data to be credible. However, the changes in body weight reported in this study were found to be unusual in that the body weight appears decreased in the low dose group, but this finding is not replicated in other studies from the same group, where in fact, a low dose of BPA *increases* body weight (Howdeshell et al, *vide infra*). In addition, a study from another group (Ashby, J., H. Tinwell, et al., "Lack of effects of low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF1 mice exposed in utero", *Reg. Toxicol. Pharmacol. 30*:156-166, 1999) did not observe BPA-related effects on body weight.

A subsequent paper from the same laboratory (vom Saal et al., *Tox. Industrial Hlth* 14:239-260) using other measurements of the same group of 6 month old animals, found a small (approximately 20%) decrease in sperm efficiency (daily sperm production per gram of testis weight) in the 20 µg/kg/day dose group (but not the 2 µg/kg/day group). The authors reported this decrease as significant ($p < 0.05$), but this level of significance was not confirmed by the Statistics Subpanel's revaluation which found only $p < 0.10$ (see final report of Statistics Subpanel). The Subpanel did not consider the issue of whether or not a 20% decrease in daily sperm production efficiency, even if real, is likely to have biological or toxicological significance. It was also noted that statistically significant changes in preputial gland weight, seminal vesicle weight, and testis weight were *not* observed in this study.

The above 2 studies were performed with CF-1 mice obtained from the colony at the University of Missouri. That colony is no longer available, but in an oral presentation at the Low Dose Peer Review, data were presented showing an effect of in utero exposure of 10 µg/kg/day to pregnant CD-1 mice on enlargement of the prostate in male offspring. Since the information from this study was not provided prior to the peer review, neither the statistics nor BPA Subpanels had the opportunity to independently analyze the raw data.

In several reports (Howdeshell et al., *Nature* 401:763-764 and Howdeshell and vom Saal, *Am. Zoologist* 40: in press) in utero exposure to 2.4 µg/kg/day BPA advanced puberty in female CF-1 mice (measured as the number of days between vaginal opening and first vaginal oestrus), although there was no change in the age of vaginal opening). This effect was observed only for females positioned between two females during intrauterine development. In the same study, increases in body weight at weaning of both male and female CF-1 were observed following in utero exposure in animals located between 2 females or between 1 male and 1 female, but not

between animals located between 2 males, during intrauterine growth. It was noted at the Peer Review Meeting that the authors did not provide the raw data to the Statistics Subpanel for re-analysis for either of these studies, so there was no way for that Subpanel to independently confirm the reported positional effects.

These observations seemed counterintuitive to the Subpanel since animals positioned between 2 females would be expected to be exposed in vivo to higher levels of endogenous estrogens from neighboring fetuses than animals positioned between 2 males. While counterintuitive, the Subpanel could not rule out the possibility of endocrine signaling loops, e.g., some type of "feed forward" loop, not currently documented. [Note added in proof. The Statistics Subpanel found in data provided in advance of a paper to be presented by vom Saal's group in Berlin in November, 2000, that no consistent positional effects on body weight were observed in castrated mice treated at 3 months of age with testosterone or 5alpha-dihydrotestosterone. See the Statistics Subpanel report for further details.] The Bisphenol A Subpanel did not discuss whether or not these reported changes in body weight at weaning or time between vaginal opening and vaginal oestrus are likely to have biologically or toxicologically significant implications.

Studies from Ben-Jonathan and Colleagues.

In a set of studies using Fisher 344 rats, another group demonstrated (Khurana et al., *Endocrinology* in press) that sc injections of 5-10 mg/kg/day of BPA on neonatal days 1-5 altered plasma levels of prolactin and developmental patterns of this hormone in the plasma of both male and female animals between 15 and 30 days of age. BPA also produced more modest changes in hypothalamic and pituitary levels of estrogen receptor mRNA levels measured by RT-PCR.

In other studies, this same group investigated the effects of administering BPA to young adult rats (7-8 weeks of age) and observed the following. 1) silastic implants yielding BPA release estimated to be approximately 0.5 mg/kg/day for 3 days increase uterine epithelial cell height and uterine weight in one strain of rats (F344) but not another (SD), and 2) silastic implants releasing BPA at an estimated daily rate of 0.3 – 0.5 mg/kg/day increase serum prolactin levels, again in F344 but not SD rats. In another study with 7-8 week old rats, this group reported increases in uterine DNA replication and c-*fos* gene expression in F344 rats – although no statistically significant effects were noted in the low dose range of 5 mg/kg/day or less used by the Subpanel – there was a trend toward a response over the entire dose range, including the low doses. Interestingly, this latter study displayed an apparent monotonic dose response curve over the entire dose range studied. It is also noteworthy, that collectively these studies illustrated a clear difference in sensitivity to the effects of BPA in the two strains of rats.

On balance, the Subpanel found the set of studies provided from Dr. Ben-Jonathan's group to be very credible, and consistent. At the same time, however, it should be stressed that these studies did NOT find any low dose effects in SD animals, and that the "low dose" effects of BPA seen in F344 animals were observed at what would be considered the "high edge" of the low dose range. Furthermore, BPA administration in all of these studies was via sc injection or release from silastic implants, and both these routes of administration would be expected to have far higher levels of bioavailability than oral administration of BPA.

[Note. It should be mentioned that while the Bisphenol A Subpanel found the data from Dr. Ben-Jonathan's group to be credible and consistent, these data were not provided to the Statistics Subpanel for their independent re-analysis.]

Low dose effects observed in other studies.

Several other reports contained data that reported effects in the low dose range of BPA. In one study from Welsch's group an effect of BPA on ventral prostate weight was observed in male SD rat offspring exposed in utero, although a subsequent study with a different sampling strategy with larger N values did not repeat this observation. Furthermore, the effect that was observed did not show a clear dose-response relationship.

In multigenerational studies in SD rats with a very large number of endpoints, statistically significant ($p < 0.05$) increases in ovarian weight were found in certain BPA-treated groups in a study by Tyl and significantly decreased anogenital distances were found some BPA-treated groups in a study by Ema. However, the Subpanel felt these observations displayed erratic dose response relationships, and some changes were eliminated when the values in question were subject to corrections for body weight differences between groups. Given the large number of endpoints in these studies, the Subpanel felt these miscellaneous observations might simply represent coincidental effects.

Summary. There are several reports of low dose effects of BPA which the Subpanel finds credible as outlined above, especially in studies with CF-1 mice from vom Saal's group, and this group presented similar data for CD-`1 mice in an oral presentation at the meeting. Data from Ben-Jonathan's group was considered very credible by the Subpanel, and this data is also potentially important because it provides very strong evidence for strain differences in the sensitivity to BPA which is consistent with known differences in SD and F344 rats with regard to sensitivity to endpoints of estrogenic action. These latter studies, however, were performed at the very high end of the low dose range. Also, they all utilized either sc injection or release from silastic implants as the route of administration of BPA, and the Subpanel questions whether similar results would be obtained if this chemical was administered by the oral route which would be more toxicologically relevant. These findings would thus provide greater support for a low-dose effect if they could be repeated at lower doses and/or following oral administration. In conclusion, there is credible evidence for low dose effects of BPA, and the conclusions reported in the above several studies are supported by appropriate statistical analyses. This evidence is limited to one dose level in a small number of reports and it is thus difficult to generalize about specificity, consistency, or strength of the evidence relating to the timing of exposure, biological endpoints measured or their functional significance, or their sensitivity.

Question 2. What is the extent of empirical evidence demonstrating the lack of low-dose effects of BPA? Within and across studies, describe the specificity, consistency, and strength of evidence with consideration of timing of exposure, when the endpoint was measured, and sensitivity of the endpoint. Are conclusions supported by appropriate statistical analyses?

A number of studies have provided evidence demonstrating the lack of low dose effects of BPA. Since low dose effects were NOT observed in these studies, the experimental details are not reviewed as extensively as those in Question #1 above – rather, interested readers may refer directly to the Selected Studies chosen by the Organizing Committee. This lack of discussion here is simply due to the lack of any observed effects. The Subpanel felt that the selected studies provided fell into 3 categories.

First there were 3 very large studies, conducted as GLP studies, that failed to show low dose effects of BPA. These included several multigenerational studies (by Tyl and Ema) in rats that examined a large number of endpoints, and a large study (by the Cagen group) using CF-1 mice specifically designed to be conducted exactly as one of the vom Saal studies.

Second, a large study from the Welsch group that used multiple pups per litter exposed during pregnancy found no BPA effects on prostate weight or on other endpoints.

Finally, a number of studies from Ashby's group using both mice and rats, including some that used the same CF-1 strain of mice and were designed to replicate the vom Saal studies, did not observe low dose effects of BPA.

The Subpanel explicitly noted that this collection of studies covered the reported window of sensitivity of exposure to BPA reported in studies referred to in Question #1, that they involved very long duration exposures over multiple generations, that they used both mice and rats, and that they administered BPA by several routes. Several of these studies also went so far as to include DES as a known estrogen so as to repeat the experimental design of some of the vom Saal studies. (Note: these studies did not observe an effect of DES on the endpoints measured as reported in studies from vom Saal's group, but the Subpanel did not address the question of whether one would or would not expect to see an effect of DES on the endpoints measured in either set of studies. The Subpanel did note, however, that if DES was included in a given study, its effect generally "mirrored" the effect of BPA, i.e., either the two produced a similar effect or neither produced any effect.) As a group these studies are very consistent, the conclusions are supported by appropriate statistical analyses, and the Statistics Subpanel confirmed the lack of BPA effects for the studies noted above, except for the second Welsch designed to investigate sampling design for which that Subpanel was not provided the raw data . Collectively, these studies found no evidence for a low dose effect of BPA, despite the considerable strength and statistical power they represent, which the Subpanel considered especially noteworthy.

Question 3. If possible, identify differences in study design or biological factors that might account for the observed differences in study outcomes.

There were a number of differences between the studies that provided evidence demonstrating either a low dose effect of BPA or the lack of such an effect.

1. Some of the no effect studies were multigenerational, and animals were thus chronically exposed to BPA. Animals in these studies may have adapted so that they did not show a response to BPA at some "critical" time. However, there were other studies that failed to

observe a low dose effect of BPA which utilized an exposure paradigm designed to reproduce the short-term exposure studies from the vom Saal group and these also failed to show a low dose effect of BPA.

2. Diets were not identical in studies that observed and did not observe low dose effects. In particular, the Subpanel wishes to note that the background level of estrogens, e.g., from dietary sources, may have been different and this could have contributed to the positive effect of the BPA via previously unrecognized mechanisms (*vide infra*). For example, vom Saal studies used a diet (Purina 5001) different from that used by the Cagen and Tyl studies (Purina 5002), and Thigpen reported (background data) that the 5002 diet had soy/phytoestrogen levels approximately half of those found in the 5001 diet. The Subpanel thought the possible contribution of dietary estrogens should be considered in light of the report that intrauterine position plays an important role (presumably due to small differences in exposure to endogenous estrogens) in determining whether developmental exposure to BPA in utero produces biological effects in the adult animal. These studies reported that IUP effects played a role in the response of both female and male animals for low dose effects of BPA, i.e., 2F animals (animals between 2 females) show the greatest response following developmental exposure to BPA.

3. The Subpanel also felt that differences in the strains of mice used could in theory have contributed to different responses to low doses of BPA. For example, while the studies of both Cagen's and Ashby's groups used CF-1 mice in attempts to replicate the experimental format used in the vom Saal studies, the CF-1 mice used by the Missouri group had been raised in a closed colony since 1979. While the studies of both Cagen and Ashby also used CF-1 mice, these were from true outbred colonies.

4. Careful examination of the raw data indicates that certain parameters in the control animals were different in studies that observed and did not observe low dose effects of BPA. In particular, the control BW and prostate weights differ between some studies, e.g., some of the Ashby studies and the vom Saal studies. This raises the theoretical possibility that tissues may have already been maximally stimulated by estrogens and /or that the differences in body and prostate weights could be indicative of different levels of maturation in the animals used in the two studies.

5. The routes of administration of BPA varied across studies, and this was felt to be potentially most significant for studies using sc injection or release from silastic implants versus those studies using oral administration. Even within the oral dosing groups there were differences, e.g., in different studies BPA was given in drinking water, by gavage, or in oil via micropipettes.

6. There may have been some differences in housing of the animals between different studies. For example, housing of animals singly versus group housing for different periods of time. This could, in theory, affect the outcome of studies since there are known effects of housing due to the phenomenon of male dominance.

7. Different bedding was used by Ashby, vs that used by vom Saal, although Cagen used the same bedding as the Missouri group. Thus, while bedding could be a potential factor (as a source of possible exposure to estrogenic substances) the Subpanel thinks this is less likely to be a potential factor than other differences.

8. The sample size was significantly different between studies that did and did not report low dose effects of BPA. However, the Subpanel did not think this particular concern was likely to be the principle basis for the differences because the negative studies had the larger number of animals. This further emphasizes that the different studies reviewed by the Subpanel indeed observed different outcomes.

9. There were differences between studies with respect to whether or not there was an analysis of BPA purity in the starting material, and in the concentrations actually present in dosing solutions. The Cagen study included an analysis of the BPA used as well as an analysis of the dosing solutions for their actual BPA content. The Ashby studies did not perform chemical analyses of BPA during the actual study, but did determine the stability of BPA solutions (stable). Vom Saal did not perform chemical analyses of BPA preparations used, but did determine the estrogenic potency. Ashby and vom Saal obtained BPA from the same supplier (Aldrich), although the Ashby sample was obtained from an Aldrich distribution source in the UK while the vom Saal material was from a source in the U.S. In contrast, Cagen used BPA obtained from a different supplier (Dow).

   The Subpanel found no specific reason to suspect that there were differences in the estrogenic activity, impurities, or other properties between the BPA batches used in these different studies, or in the preparation of dosing solutions. Nevertheless, the studies that reported a positive low dose effect did not specifically analyze their dosing solutions or starting material, and one has to recognize this is always a potential confounder. Thus, without analyses done at the time of the actual study, one cannot unequivocally rule out potential effects of contaminants or errors in the preparation of dosing solutions.

Question 4. How do the findings from studies of low-dose effects on reproductive and developmental outcomes using BPA compare with those for other endocrine active chemicals? Describe the specificity and consistency of the evidence, and if possible, identify the similarities and/or differences in study design, chemical activity, species or strain, etc. that might explain the observed outcomes.

Due to the large amount of material on BPA assigned by the Organizing Committee, and the level of discussion of that material, the Subpanel did not have sufficient time for an in depth analysis of studies on other endocrine active chemicals. The Subpanel felt that it could only briefly consider DES, since some of the BPA (both those reporting and not reporting low dose effects of BPA) also studied DES. As noted previously, a representative of the Organizing Committee instructed the Subpanel to consider a dose of 1 μg/kg/day as a "low dose" of DES.

There are credible reports that DES may produce low dose effects, including studies from both the vom Saal and Newbold laboratories. Similar to the vom Saal findings in the male reproductive tract, Newbold's lab found low-dose stimulatory effects on female reproductive tract endpoints. Studies from the Welsch group also showed a decreased body weight at 10 µg/kg/day of DES in drinking water and an effect on the vaginal opening (advanced) in the female pups. Data from another study provided by the Organizing Committee as background information (Gupta) also showed a low-dose BPA effect on prostate, which was blocked by the antiestrogen ICI 182,780. However, other credible studies have not observed such low dose effects of DES or other estrogens. Thus the limited time available to discuss this question, as well as the conflicting results, did not really allow the Subpanel's discussion of this point to contribute much to the central issue of low dose effects of BPA.

(As a point of reference, several members of the Subpanel, who are familiar with uterotrophic assays, expressed the opinion during discussions that an oral dose of 0.2 µg/kg/day DES is approximately the lowest dose at which they would expect to observe an effect of the synthetic estrogen following this route of administration.)

Question 5. Describe the available and relevant pharmacokinetic, biologic, and other mechanistic information that strengthen or weaken the plausibility of low dose effects of BPA. How did this information impact on the Subpanel's overall conclusions? Describe the shape of the dose response curves in the low-dose region using empirical data as well as biologically based dose-response models.

The selected studies included some *in vitro* data which indicate that BPA is not bound as extensively to serum proteins as estradiol. This data alone would suggest that BPA might preferentially exit the plasma space (relative to the endogenous hormone) to enter target cells. However, this hypothetical possibility has not been established *in vivo*, which is the critical issue since decreased plasma binding might also be expected to enhance both renal and hepatic clearance and such an effect would be expected to decrease BPA concentrations at cellular receptor sites.

There is also data available that the bioavailability of BPA is likely to be significantly less following oral vs parenteral administration because of first pass hepatic metabolism, especially since the glucuronides generated by metabolism do not have appreciable affinity for the classical estrogen receptor.

A very important issue for thorough analyses of developmental effects is fetal uptake via transplacental transfer, of non-metabolized BPA, as well as fetal biotransformation and accumulation of BPA and metabolites. Extensive information on this important point was not available in the selected studies, but from comments offered to the Subpanel by audience members, these issues may be contentious.

An in-depth knowledge of *directly* measured pharmacokinetic parameters such as bioavailability, bioaccumulation, transplacental transfer and fetal accumulation, clearance, volume of distribution, half-life, and the complete spectrum of metabolites formed, is essential to

understand the toxicology of BPA and is particularly important because of the low affinity, as measured in vitro, of this chemical for estrogen receptors. This is also important because BPA and other endocrine active chemicals could in theory affect the metabolism of endogenous steroid hormones via induction or inhibition of P450s or glucuronyl transferases, or other mechanisms.

With respect to the shape of the dose curve, there are not sufficient doses of BPA that have been reported to elicit low dose effects to establish the shape of the dose response curve for this chemical. Parenthetically, the Subpanel noted that in the studies with F344 rats, which contained 1-2 doses of BPA in the low dose range (5 mg/kg/day or less), showed a monotonic dose response curve (Ben-Jonathan studies).

In addition, the Subpanel wishes to emphasize that the large number of negative data points in the literature make it impossible to perform any sensible dose response modeling for BPA in the low dose range at this time.

The paucity of BPA pharmacokinetic data led to the Subpanel's view that this is an area that represents a critical data gap. Extensive pharmacokinetic information was not provided in the selected studies and background information. This, coupled with the available time, precluded the Subpanel from undertaking a rigorous consideration of pharmacokinetic issues. Thus, if definitive and reproducible pharmacokinetic information on BPA is available, it should be thoroughly analyzed in the context of low dose effects, and if such data is not available, it should be a high priority for future work. This was emphasized by the striking difference in response of two rat strains (F344 and SD) to BPA in the selected studies, and because of the clear evidence of genetic effects on hormonal responsiveness reported in the literature and described in part during the oral sessions at the meeting.

Question 6. Based on the totality of available knowledge, what is the Subpanel's overall conclusion regarding whether BPA can cause hormone-related effects on reproductive and developmental endpoints at doses lower than those typically used in the standard toxicological dose-setting paradigm?

There is credible evidence that low doses of BPA can cause effects on specific endpoints. However, due to the inability of other credible studies in several different laboratories to observe low dose effects of BPA, and the consistency of these negative studies, the Subpanel is not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding. In addition, for those studies in which low dose effects have been observed, the mechanism(s) is uncertain (i.e., hormone related or otherwise) and the biological relevance is unclear.

(Note: The Subpanel wishes to emphasize that the above is a consensus statement. The Subpanel expended a considerable amount of time and effort developing the above statement to answer the question posed by the Organizing Committee about our "overall conclusion", and the true sense of our overall conclusion is accurately presented from the *entirety* of the above statement. Thus, the presentation of only a portion of the above statement would not accurately

represent the content or spirit of our conclusion. Consequently, the Subpanel will not endorse anything but the above statement in its complete form as presented here.)

Question 7. Are there specific knowledge gaps for which additional research relative to the low-dose question is needed? If possible, suggest ways to address those gaps.

There are numerous knowledge gaps which limit the ability to assess low dose effects of BPA. Some of the specific items suggested during the Subpanel discussion are listed below, although it should be noted that this listing is not intended to be all-inclusive.

1. Studies should be performed with multiple doses of BPA in the low dose range, especially following oral administration during in utero or early neonatal development. If experimental paradigms can be developed to conclusively establish low dose effects of the chemical as a general, reproducible phenomenon, these should be used to obtain sufficient data points to perform credible physiologically based pharmacokinetic modeling.

2. Extensive pharmacokinetic data should be obtained in multiple species including CF-1 and CD-1 mice and in F344 and SD rats. This data is intrinsically important and is also required for modeling studies.

3. There is no meaningful data on the occupancy of estrogen receptors following exposure of animals to BPA in the low dose range, especially during critical periods of development. Data on the occupancy of receptors, in the reproductive tract, pituitary, and brain, following exposure of animals to low and high doses of BPA would be very valuable and is essential to rigorously address mechanistic questions.

4. The use of pharmacological (e.g., specific receptor antagonists) and genetic (e.g., knock out animals) approaches would provide important information about the mechanism of BPA effects, especially the role (if any) of estrogen receptors in any observed effects.

5. Further studies on the intrauterine position effect are suggested to fill existing knowledge gaps. It would be valuable to establish the generality and reproducibility of this effect, as well as establishing unequivocally the endogenous hormone levels as a function of intrauterine position and the site of their production. In related areas, it may also be important to carefully examine the effect(s) of minor differences in background levels of estrogens (e.g., provided by different feeds, due to genetic variation and species differences, etc.).

6. Genetic and epigenetic factors that affect responses to BPA and hormones in general are important areas that deserve further study. These include not only factors that affect hormone and receptor levels, but also factors in “intermediate” steps in hormone action which could lead to observed differences in sensitivity (e.g., such differences in “intermediate” or “down stream” effects have been suggested from studies comparing F344 and SD rats).

7. Mechanistic studies of BPA action that span the full course of in utero development, neonatal life, puberty, and adulthood would provide important data not currently available.

8. Given recent advances in understanding the basic mechanisms of steroid receptor actions, ligand specific effects of BPA on the transcriptional activity of receptors, recruitment and activation of co-activators and co-repressors, regulation of transcription by protein-protein vs DNA-binding mechanisms, and non-genomic actions of BPA might aid our understanding of the actions of this chemical at all dose levels. These studies would be especially important to determine if effects of BPA are mediated through classical hormone regulated pathways, or whether other mechanisms are operable.

9. One of the most critical needs is to search for other possible markers and specific endpoints (e.g., in addition to, or instead of, gross measures such as organ weights) that can be used to reproducibly investigate low dose effects of BPA. The development of easily measured and sensitive molecular endpoints, especially endpoints that can be assessed shortly after exposure, are critical needs which would greatly aid our ability to resolve current questions about low dose effects of BPA.

Chapter 2:
Report of the Other Environmental Estrogens and Estradiol Subpanel

Chair
Michael Gallo, UMDNJ-Robert Wood Johnson Medical School

Rapporteur
Kenneth Reuhl, Rutgers University

Facilitator
Lynn Goldman, Johns Hopkins University School of Public Health

Panelists
Mari Golub, California-EPA
Claude Hughes, UCLA School of Medicine
Richard Lyttle, Wyeth-Ayerst Research
Lynne McGrath, Schering-Plough Research Institute
Patricia Whitten, Emory University

Statistics and Dose-Response Modeling Subpanel Representatives
John Bailer, Miami University of Ohio (statistics)
Michael Kohn, NIEHS (modeling)

Introduction:

Estrogens are classically defined as those compounds (endogenous and exogenous) that induce the state of estrus in the immature or ovariectomized female rat or mouse. Variations on this theme have been used since the discovery and characterization of the estrogen specific binding protein called the estrogen receptor. In elegant experiments it has been shown that the estrogen action of estradiol and other estrogens is mediated by the estrogen receptor (ERα). Hence, many compounds are being classified as estrogens because they bind to the estrogen receptor and turn on estrogen responsive genes. In addition, the recent discovery of a second high affinity estrogen binding protein (ERβ) has added new complexities to the general principle of induction of estrus. The ERα and ERβ are not distributed equally in the same tissues or cells. Hence, even the definition of a classic ER agonist, antagonist and partial agonist is in flux.

The subpanel addressed the question of *what is an estrogen* and how the definition should flavor the deliberations. The members also addressed the question of potency of a compound as an estrogen when compared to estradiol. The area of relative potency and mechanisms of action were beyond the scope of this meeting but are clearly areas for further research. The discussions focused on three major areas: 1) criteria to evaluate effects, 2) relevant parameters to be considered, and 3) gaps in the databases.

The criteria to evaluate effects centered on the question of the ability to lead to disease, or do the effects lead to a persistent and detectable change in cells, tissues or organs. Importantly, the question of whether the effect(s) were part of a continuum of toxicity, or simply a manifestation of a physiologic response that rapidly reversed with no permanent effect, was addressed. The overarching question of *what is a low dose* is imbedded in the above questions.

The subpanel considered several key parameters in their deliberations. The first and foremost parameter was the compound in question. What was known about the chemistry and biological effects of the compound as tested. Secondly, the questions focused on the specifics of the protocol such as the age and time of exposure and/or examination of the animals, the length of exposure of the animals, the dose range tested, what was the lowest dose tested, what was the lowest effect level, was there an effect at the lowest dose tested, and the species, strain, sex and number of animals tested. An important question that remained in front of the subpanel was *what is a low dose.* The consensus was that a low dose might be compound specific based on background exposure, body burden, or chemical class. It was generally agreed that a single number could not be used as the low dose for all compounds. The robustness of each study was considered using the above criteria. Dose-response curve models (See Figure 1) were discussed that allowed the subpanel members to characterize the general biological effects of the compounds. The subpanel developed an operational definition for “low-dose effects” that was based on the dose-response data for the selected endpoints for each agent under evaluation. Low-dose effects were considered to be occurring when a nonmonotonic dose-response resulted in significant effects below the presumed NOEL (no-observed-effect level) expected by the traditional testing paradigm.

Qualitative and quantitative data gaps were identified for all the compounds examined.

Compounds Examined:

The compounds selected for examination were considered to be representative of several classes of estrogenic xenobiotics, and estradiol. These compounds include: genistein (an isoflavone soy derivative), diethylstilbestrol [DES] (the prototypical non-steroidal estrogen), octyl-, and nonylphenols (environmental chemicals), methoxychlor [an estrogenic methoxylated derivative of p,p'-DDT] (an organochlorine insecticide) and estradiol (the primary ovarian estrogen). For each compound the subpanel made a judgment on five criteria: 1) empirical evidence for a low dose effect, 2) empirical evidence for lack of a low dose effect, 3) relevant PB/PK and/or mechanistic information, 4) overall evaluation of hormonal activity at "low doses", and 5) data gaps.

Genistein is an isoflavone derived from soy. As such genistein is a general dietary component of humans, including newborns and children, and animals. There is a large body of experimental and clinical literature relating to the biological actions of genistein. A low dose effect at 25ppm in the diet is achievable in human mammary tissue, and in mammary tissue, brain and lymphocytes of CD rats. The latter effect involved an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3. At 5ppm in the diet there was a trend toward a decrease in volume of sexually dimorphic nuclei (SDN) of the medial preoptic area (POA) of the hypothalamus in male rats that returns toward normal at 625ppm. The volume of SDN-POA is approximately 5-10 times larger in male rats than in female rats. However, in F1 male rats that had been exposed to genistein the SDN volume was intermediate between control male and control female rats. The physiological consequence of a change in volume of SDN-POA in male rat pups has to be more fully examined and elucidated. The lack of data below the 5ppm dietary concentration of genistein was considered as evidence that an effect at the lowest dose tested had been demonstrated. Little mechanistic data and/or physiologically based-pharmacokinetic (PB/PK) information were made available to the panel. However, there is a great deal of information in the open literature. The overall evaluation of genistein is that hormonal activity was demonstrated at low doses (5-25ppm) and the effects were seen in the CNS, mammary tissue and WBCs. There are several data gaps that should be addressed regarding genistein. There should be clarification of the activity at the lower end of the dose-response curve (<25ppm in the diet), what mechanisms are involved in genistein action (test in the ERKO α and β), and test in studies using estrogen antagonists. The subpanel was unanimous in its recommendation that the "low dose" studies must be replicated.

Diethylstilbestrol (DES) is the prototypical non-steroidal estrogen. The compound was synthesized in the mid-1930s, and introduced in the 1940s as a drug to prevent spontaneous abortions, and the late 1940s as a caponizing agent in chickens and a supplement in cattle to enhance weight gain. DES was banned for use in pregnancy and as an indirect food additive in the 1960s but is still available for some medical and veterinary uses. DES is an animal carcinogen, as well as a transplacental carcinogen in humans. There is an enormous literature on the biology and toxicology of DES. The transplacental carcinogenicity and the effects on the neonatal mouse urogenital system are unique. The most recent literature is addressing questions

of effects on the developing male reproductive system. The importance of DES is underscored in that it serves as the model for several other estrogenic chemicals. Understanding the mechanism of action of this potent drug and toxicant will lead to further insights into the actions of other like compounds. There is very clear empirical evidence of a low dose effect on prostate size at 0.02μg/kg bw in CF-1 mice and supportive evidences in CD-1 mice. Additionally there is evidence of behavioral changes in CD-1 mice at this dose. There is evidence from dose-response studies that a no effect level exists at 0.002μg/kg bw in mice. PB/PK and mechanistic information exists but was not reviewed by the subpanel. Overall, hormonal activity was observed at low doses (see above) but non-hormonal effects may exist at lower doses. Data gaps exist in several areas. The low dose effects should be remodeled. A plausible mechanism and the studies to support it should be developed to validate the findings of enlarged prostates in treated males. Additionally, a META analysis should be conducted on the four major studies that have been completed on DES.

Alkylphenols are industrial compounds that have been identified in drinking water supplies and wastewaters. Several investigators have demonstrated the estrogenic potential of this class of compounds *in vitro* and *in vivo.* Two compounds were evaluated by the subpanel; nonylphenol and octylphenol. The "low dose" of nonylphenol is questionable. Renal toxicity occurred at the 200ppm in the parent generation of SD-rats of a multi-generation study, and increased relative uterine weight occurred at the same concentration in F1 females. Higher doses (2000 ppm) induced changes in testes and prostate weight, decreased live births and prolonged estrus cyclity. Several other changes occurred in the F1 offspring of the 25ppm group including an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3, an increased relative thymus weight, a decreased volume of SDN-POA in males, and a prolonged estrus in females. A pattern of change in SDN-POA volume in males following treatment with nonylphenol was similar to that observed for genistein. As stated above the SDN-POA changes are difficult to interpret at the moment, as are the anti-CD3 findings. The lack of a low dose effect appears to be at approximately 5ppm, but the immune markers have not been tested at that level. There was no PB/PK information available to the subpanel on nonylphenol. The data gaps for nonylphenol are similar to those stated above. What is the meaning and mechanistic basis for the changes in SDN-POA? Immune-markers should be evaluated at levels less than 25 ppm. More data are needed at the level of ng/kg/day, the human exposure level.

Octylphenol is another member of the class of alkylphenols. The compound is primarily used as an intermediate for the production of surfactants. The literature for low dose evaluation of this compound is limited to one major study. There was no evidence of a low dose effect in a five dose multigeneration study in rats. Only the highest dose (2000 ppm) induced toxicological changes. At doses ranging from 0.02-200ppm no effects were observed. Little or no PB/PK or mechanistic data exists for this compound. The binding of octylphenol to the ERα is 3 to 7 orders of magnitude less than estradiol to the same receptor. Hence, there is no evidence that octylphenol induces hormonal activity at low doses. The only data gap identified by the subpanel was the absence of a confirmatory study in another laboratory species.

Methoxychlor is a chlorinated diphenylethane insecticide that is chemically related to p,p'-DDT. The major difference between methoxychlor and DDT is the substitution of methoxy-groups for the chlorines in the para-positions of the phenyl rings. The estrogenicity of methoxychlor in

rodents has been known for several decades. The compound binds to the ERα with a relatively low affinity compared to estradiol, but can induce uterotrophism in immature rodents. Despite the many studies on the mechanism(s) of action of methoxychlor few studies detailed the dose-response relationship for estrogenic action. The primary multi-dose study with methoxychlor examined six doses ranging from 0.05 to 150mg/kg/d given shortly before birth and through neonatal day 7, at which time the pups were dosed directly. Effects were seen at all doses with the exception of the 0.5 and below. At the effective doses there were a wide range of changes in estrogen sensitive organs. The lowest effective dose was 5/mg/kg/d. Hence it appears that at levels less than 5mg/kg/d (0.05 and 0.5mg/kg/d) none of the effects seen at higher doses were reported. Mechanistic studies have been carried out in several laboratories. Serum concentrations and milk concentrations mimic the administered doses. Additionally, the active metabolites also mirror the parent compounds in relative concentrations. Overall the classic estrogenic activity is limited to doses greater than 5mg/kg/d, but some immune effects (increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3 in F1 female SD rats) have been reported at 10ppm in the diet. Data gaps that should be addressed to complete the profile on methoxychlor are a further evaluation on the anti-CD3 alterations. Are these changes related to estrogen action or are other pathways affected. A second data gap, which may be more important, is the comparison between technical grade and pure methoxychlor.

Estradiol is the ovarian steroid with the greatest estrogenic activity. It binds the ER with the greatest affinity of the compounds evaluated and induces all the classic effects that are termed estrogenic. The basic biology of estradiol defines the feedback pathways of the steroid-driven endocrine system. Estradiol has been widely studied and the molecular mechanisms of action are very well understood. The studies evaluated by the Subpanel addressed the question of low-dose activity keeping in mind that this is an extremely potent steroid that is difficult to evaluate in an *in vivo* system because of the homeostatic mechanisms of the test systems. The ovariectomized CD rat in the Tier 1 (EDSTAC Protocol) studies had reproducible changes in serum prolactin at 3µg/kg/d with an associated increase with administered dose. Several other changes in hormonally active tissue were reported, as was changes in LH and FSH as a function of dose and blood level. An apparent no effect level was attained in a 90-day feeding study with no changes being observed at the three lowest doses tested (3, 170 and 700µg/kg/d). Mechanistic studies were not conducted in this bioassay. However, the best metric for endocrine changes was the blood estradiol level rather than administered dose. Overall, the TIER1 dietary study demonstrated the three types of dose-response relationships based on particular endpoints, tissue responses and time points. The remaining data gap for estradiol in this test system is to determine the shape of the dose-response curve at the low effect and high no effect levels.

GENERAL RECOMMENDATIONS

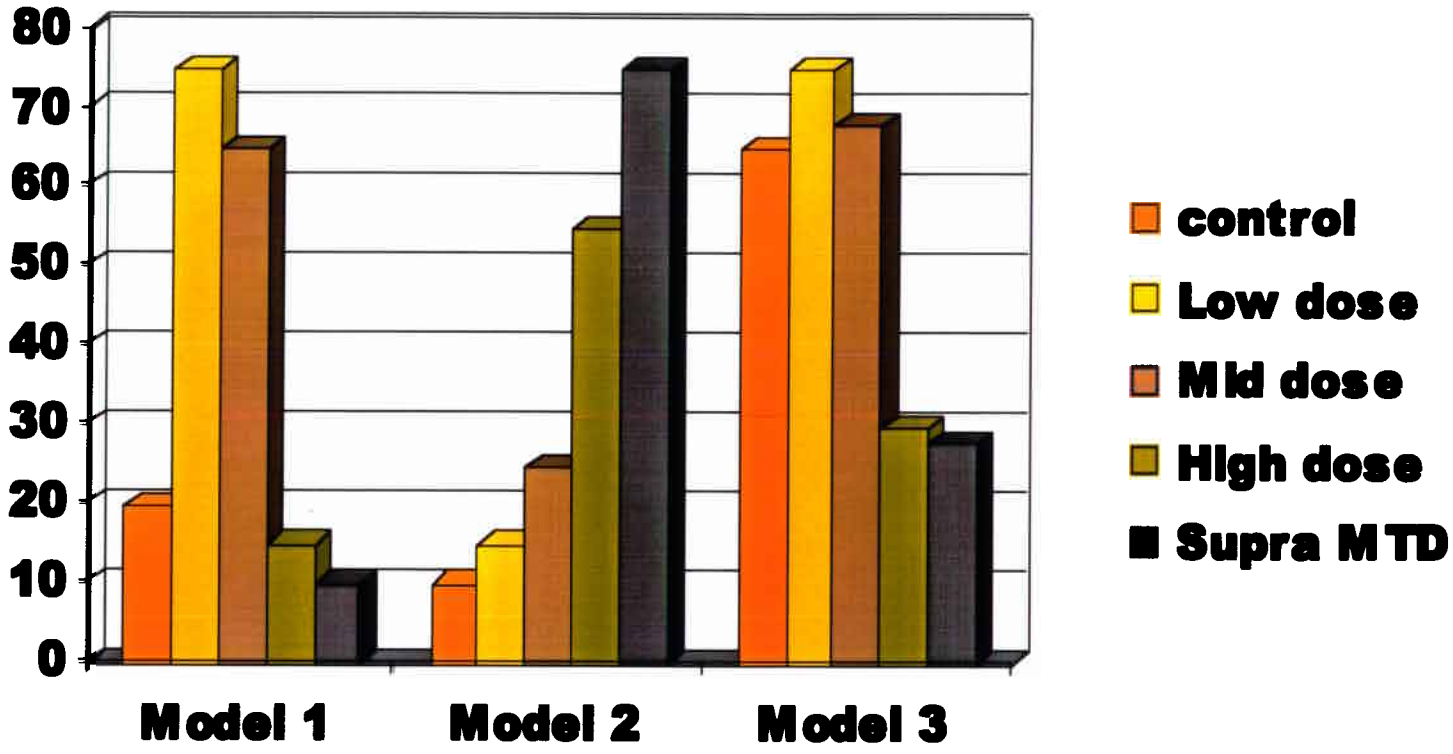
Based on the studies and compounds reviewed the subpanel has three major recommendations that cross most studies:

1) Model dose-response relationships - in multiple dose studies, modeling of dose-response relationships should be done in addition to pair-wise comparisons;
2) Replicate and validate studies - low dose effects are difficult to ascertain, hence the studies must be replicated and validated across laboratories;

3) Examine and elucidate the physiological and toxicological consequences of SDN-POA and anti-CD3 changes - determine the biological significance of the volume changes in the SDN-POA in male rodents, and the significance and relationship of the anti-CD3 changes and estrogen action.

Figure 1:

Observed Dose Response Curves



Simulated

Chapter 3:
Report of the Androgens and Antiandrogens Subpanel

Chair
Shuk-Mei Ho, University of Massachusetts Medical School

Rapporteur
Terry Brown, Johns Hopkins University School of Public Health

Facilitator
Eisuke Murono, National Institute for Occupational Safety and Health

Panelists
George Daston, The Procter & Gamble Company
Mitch Eddy, National Institute of Environmental Health Sciences
Lorenz Rhomberg, Gradient Corporation
Elizabeth Wilson, University of North Carolina at Chapel Hill

Statistics and Dose-Response Modeling Subpanel Representative
Ralph Kodell, National Center for Toxicological Research (statistics)

Introduction

This article highlights major issues discussed in the Androgens and Antiandrogens Subpanel at the 2000 Endocrine Disrupters-Low Doses Peer Review Workshop. The goal of the Workshop was to review the scientific evidence related to potential low-dose effects of endocrine active chemicals (EACs) on human health. The Subpanel was charged to examine data from a selected study on vinclozolin (V), a fungicide with antiandrogenic activity [1], and other supporting information [2-21] in order to determine whether the body of evidence demonstrated low-dose androgenic/antiandrogenic effects or the lack of such. Specifically, it was asked to evaluate the consistency and strength of the scientific evidence presented in the study, including parameters such as timing of exposure, sensitivity of the endpoints, sample sizes, sampling methods, appropriateness of the controls, strain differences, species sensitivity, number of doses in the low-dose range, and the vigor of the statistical analyses. Other main topics of discussion included the shape of the dose-response curve for V, and other relevant mechanistic data that might strengthen or weaken the plausibility of its low-dose effects. During the Subpanel's deliberation, opinions on whether the weight of evidence provided sufficient grounds to change the traditional dose-setting paradigm, particularly in the low-dose region, for mammalian toxicology studies on V and related compounds, were reflected. Finally, knowledge gaps regarding environmentally active androgenic/antiandrogenic agents were identified.

Defining "low-dose effects"

In order to assess whether low-dose effects exist, it is imperative to first define what constitutes low dose for each EAC under investigation. The organizers of the conference, NTP and NIEHS, had asked the Subpanels to consider "low dose effect as biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in EPA's standard toxicity testing paradigm". With regard to V, information on environmentally relevant exposure levels in human populations is currently unavailable. However, the no-observable-adverse-effect level (NOAEL) and low-observable-adverse-effect level (LOAEL) for acute dietary exposure to V has been established to be 6 mg/kg/day and 11.5 mg/kg/day, respectively, while the chronic dietary NOAEL and LOAEL are set at 1.2 mg/kg/day and 2.3 mg/kg/day, respectively (revised Human Health Risk Assessment 5-12-00, Office of Pesticide Program). These values have been established from rat studies, using developmental and ventral prostate (VP) weight changes as endpoints in acute exposure studies and histopathological lesions of lungs, liver, ovary, and eye as adverse effects in chronic exposure studies. Human risk assessment levels have been derived from these values after adjustment for human factors. In order to facilitate further discussion the Subpanel had arbitrarily defined "low-dose range" as one below the currently recognized NOAEL/LOAEL. However, the Subpanel was aware of the fact that no studies had been conducted in dose-ranges below NOAEL/LOAEL for V. The issue of whether studies conducted at NOAEL/LOAEL should be considered as low-dose studies or only those carried out at dose ranges substantially lower than NOAEL/LOAEL be counted was debated. At present, no information is available on environmentally relevant exposure such as exposure levels from crop residue. This data gap has hampered current and future hazard evaluation, risk assessment, toxicity testing and risk management.

Vinclozolin is an antiandrogen

Vinclozolin (V, 3-(3,5-dichlorophenyl)-5-methyl-5-vinyl-oxazolidine-2, 4-dione) is a dicarboximide fungicide widely used to control fungal growth on several fruits, vegetables, and turfgrass [22]. Structurally, it resembles hydroxyflutamide and exerts antiandrogenic action [20,23]. Prenatal exposure to V causes reduced anogenital distance (AGD), hypospadias, ectopic testes, viginal pouch formation, agenesis of the ventral prostate, and nipple retention in male offspring. Exposure of pregnant rats to low dosages of V, 12 mg/kg/d or lower, results in female-like AGD, retained nipples, and permanently reduced VP weights in some male offspring, while a high dose of 100 mg/kg/d causes hypospadias, deformity, and infertility in all male offspring. In contrast, fertility and reproductive functions are unaffected when male rats are exposed as adults to high doses of V in a chronic manner [24]. Mechanistically, the antiandrogenic action of V is mediated via metabolic conversion to two open-ring metabolites, M1 and M2, which have been shown to induce AR nuclear import, compete with androgen for AR binding, and block AR-DNA interaction [20]. They do not bind estrogen receptor (ER)-α nor inhibit 5α-reductase activity, yet do possess weak affinity for progesterone receptors. Biologically, M2 is more potent than M1, and has a *Ki* around 1 μM, as compared to hydroxyflutamide, which has a *Ki* of 0.1 nM, for the AR. In contrast, V, the parent chemical, is a poor AR antagonist [20]. Dosimetry studies reveal that when M1 and M2 in maternal serum concentrations approach their respective *Ki* values for AR binding hypospadias and more severe infertility are noted in the exposed offspring [25].

The Vinclozolin data set

The study by Gray and associates [1] was the only article selected for peer-review. The principal investigator had submitted raw data to the Statistics and Dose-Response Modeling Subpanel for re-evaluation and has responded to the 23 questions listed in the "Issues Relative to the Evaluation of Low-Dose Studies". In this study, pregnant rats were dosed (po) with V at 0, 3.125, 6.25, 12.5, 25, 50, or 100 mg/kg/day from gestational day (GD) 14 to postnatal day (PND) 3, and postnatal reproductive developmental abnormalities, including female-like AGD, retained nipples, cleft phallus with hypospadias, suprainguinal ectopic scrota/testes, a vaginal pouch, epididymal granulomas, and small to absent VP, in the male offspring were used as endpoints. The investigators reported that the AGD was significantly reduced in newborn male offspring and the incidence of areolas was increased. VP weight in one-year-old males was reduced and permanent nipples were observed in male offspring born to mothers treated with all tested doses. Males born to mothers exposed to the two higher doses (50 and 100 mg/kg/day) exhibited reproductive tract malformations and reduced ejaculated sperm numbers. Different endpoints displayed varied dose-response curves. For example, AGD, areolas, and VP weight displayed a continuous response to V treatment with no apparent threshold while hypospadias and ectopic testes exhibited threshold-responses.

Statistical re-evaluation revealed that statistical methods applied in the study were generally appropriate. The Statistic Subpanel agreed with the investigators' decision to use the litter as the basic experimental unit, given the significant "litter effects" that were present in the data. It also confirmed that at postnatal day2, AGD in male offspring was significantly reduced in the group exposed to the lowest test dose (3.125 mg/kg), but it did not find AGD reduction to be significant for the next lowest dose (6.25 mg/kg/day) group. Additionally, it agreed with the investigators that the 50 and 100 mg/kg/day V doses significantly reduced VP weight, the 100 mg/kg/day dose diminished seminal vesicle weight, and none of the tested doses affected body weights, testis weight, paired epididymis weight, or testicular and epididymal sperm counts.

However, the Subpanel's analyses disagreed with the investigators' finding that an effect of 6.25 mg/kg/day V on VP weight was significant.

Based on peer-review of the investigators' data and the Statistic Subpanel's re-evaluation, our Subpanel noted the following: 1) the study was well-designed and demonstrated V exerted antiandrogenic action developmentally, 2) the test doses were in the NOAEL/LOAEL range, 3) alteration in AGD in male offspring was the most sensitive endpoint and was significantly different at the lowest test dose (3.125 mg/kg), 4) the areolar/nipple persistence in males was a highly informative endpoint but was not subjected to statistical re-analysis, and 5) for some end-points tested, the dose-response curves displayed no threshold and appeared linear to the lowest test dose, which simultaneously approached the limit of detection.

Other antiandrogenic EACs and their mechanisms of action

Other EACs with antiandrogenic activities have been identified[10]. According to their mechanisms of action, they could be broadly classified into AR antagonists, 5α-reductase inhibitors, and inhibitors of steroidogenesis.

In addition to V, several EACs are known to exert antiandrogenic activities through their action as AR antagonists. The p,p'-DDE, a persistent metabolite of the insecticide p,p'-DDT, binds AR with moderate affinity and inhibits androgen-induced transcription with a potency similar to that reported for the antiandrogen hydroxyflutamide [16]. Prenatal exposure (gestational day 14-18) to 100/kg/day of p,p'-DDE induced AGD reduction and nipple retention in male offspring [10]. Other metabolites of p,p'-DDT have also been reported to exhibit antiandrogenenic activities in a cell culture system [26]. Similarly, 2,2-bis-p-hydroxyphenol-1, 1, 1-trichloroethane, the primary o-demethylated metabolite of the pesticide methoxychlor, is an effective AR antagonist [26]. It exerts potent antiandrogenic action *in vivo*. The fungicide procymidone also acts as an AR antagonist [14,18]. At a dose of 100 mg/kg/day, it produces *in vivo* antiandrogenic activities similar to those induced by V and p,p'-DDE [10,14,18]. Linuron, a herbicide, is a weak competitive inhibitor for the AR (*Ki* of 100 mM) although it is highly effective in altering sexual differentiation *in vivo* in an antiandrogenic manner [10]. Short-term treatment of castrated adult rats with linuron reduces testosterone- and DHT-dependent tissue weights in the Hershberger assay [13] at an oral dose of 100 mg/kg/day for 7 days. It has been suggested that the principal antiandrogenic acton for linuron is mediated via inhibition of steroidogenesis [27]. Likewise, the fungicide ketoconazole has been shown to exert antiandrogenic activity *in vivo* by lowering serum testosterone levels, and altering both gonadal synthesis and hepatic inactivation of testosterone [28]. Lastly, the phthalates, common plasticizers, are now believed to exert their antiandrogenic action via interference with the synthesis or metabolism of androgens [29].

Pharmaceuticals (e.g. flutamide and Casodex) developed to treat prostate cancers and benign prostatic hyperplasias (BPH) are potent antiandrogens. Flutamide, following *in vivo* hydroxylation to hydroxyflutamide, acts as an AR antagonist. It binds to AR with high affinity but fails to initiate transcription [30]. Casodex (ICI 176,334) binds to the AR with good affinity, but fails to induce receptor accessory protein dissociation, DNA binding, and transcriptional activation [31,32]. In a 5-day Hershberger assay [13], flutamide at 0.15, 0.6, 2.5 and 10 mg/kg/day effectively produces antiandrogenic responses [21]. In contrast, finasteride, developed to treat BPH and hair loss, exerts its antiandrogenic action primarily by acting as a 5α-reductase inhibitor [33,34]. Gestational exposure to finasteride produces transient AGD reduction and nipple retention, and increases hypospadias incidence in male rat offspring [6,7]. It has been noted that the decrease in AGD apparently shows a linear response over the tested dose range. Yet, the hypospadias

response does exhibit a threshold around 0.1 mg/kg/day and a 100% effect level at 100 mg/kg/day (with dosing through Day 20 of gestation) [6]. Interestingly, unlike AR antogonists, finasteride does not completely block prostate differentiation or feminize the external genitalia despite high-dose exposure [15]. Finally, the non-steroidal antiandrogen, nilutamide, has a weak binding affinity for the AR. Nonetheless, it has a long biological half-life and potent antiandrogenic activity *in vivo*. Its action is likely due to its inhibitory action on androgen synthesis [35-37].

When the effects of V are compared to those induced by the aforementioned EACs it becomes apparent that these compounds all behave as antiandrogens. Thus far, there is little evidence for environmental chemicals that act as androgen mimics. Although examples of potential androgenic effects in wildlife have been reported [38] presently no evidence exists for effects in humans related to environmental exposures. If androgenic mimics were to be present in the environment, they most likely will be detected by their effects on female development and reproductive functions. In this regard, it has been reported that exposure of pregnant rats to testosterone adversely affects pregnancy and masculinizes female offspring [39].

Mechanistic models for screening environmental antiandrogens

In order to safeguard detection of low-dose effects caused by environmental antiandrogens, it is important to employ appropriate mechanism-based models with the most sensitive endpoints for detection. Existing multigenerational tests sample only a small number of pups for necropsy and have missed malformations and low-dose effects of EACs. The problem is more serious when the endpoints are low frequency events. For example, multigenerational studies with DDT have failed to detect the androgenic effects of p,p'-DDE in rats, mice and beagle dogs [40]. With regard to detection of environmental androgen/antiandrogen three mechanism-based *in vivo* screening models, if utilized routinely and complementarily, should help to improve detectability. The gestational/postnatal exposure rat model system is based the ability of an EAC to interfere with low levels of endogenous androgen required for normal male development. This is recognized as a highly sensitive *in vivo* assay. The developmental endpoints such as AGD reduction, areolar/nipple persistence, preputial separation, hypospadias, testicular descent are highly sensitive in detecting antiandrogenic activities [29]. Mechanism-wise, the *in vivo* antiandrogenic activity is, in general, supported by AR binding and transcriptional activity assays. Furthermore, the EAC's *Ki* for AR binding usually agrees with the relative potency of its antiandrogenic action [20]. The Hershberger assay [13], a short-term *in vivo* assay, also has the sensitivity and specificity for detecting androgen/antiandrogen. Castrated mature rats are exposed to the chemical in the absence of presence of an androgen and accessory sex organ weights are used as detection endpoints. Androgenic activity is detected on the basis of stimulation of accessory sex organ weight gain in the absence of an androgen and antiandrogenic activity is measured as competitive inhibitory action on androgen-stimulated growth of these organs. Lastly, the peripubertal exposure model is another *in vivo* system that permits detection of androgenic/antiandrogenic activity based on preputial separation. In addition to *in vivo* assays, numerous *in vitro* or cell-free androgen/antiandrogen-screening assays have been developed. These assays are based on hormone specific mechanisms of action such as cell-free and whole-cell AR binding, androgen-dependent cell proliferation, and transcriptional activation of androgen-specific reporters or genes. Utilization of these assays as first-tier tests or as mechanism-finding assays should complement findings from *in vivo* assays [29,41].

Comments and recommendation on current and future study design

Based upon existing knowledge of mechanisms, the current study designs are found to have the sensitivity and selectivity for detecting environmental antiandrogen, particularly if these compounds behave as competitive AR antagonists. However, protocol modifications may be required for the detection of EACs that act(s) as inhibitors of steroid biosynthesis and/or metabolism modulators. The Subpanel has found little additional benefit to extend the current standard toxicology tests to dose range below the currently set NOAEL/LOAEL. This conclusion is reached partially based on the low incidence of malformation detected by the current assays at NOAEL/LOAEL, suggesting that the detection limits of these assays may approximate these levels. Furthermore, based on our present understanding of the mechanisms of action for these detection assays we recognized that current data are obtained based upon environmental chemicals acting as antiandrogens, i.e. as agents that interfere with endogenous testosterone action. In intact adult animals, testosterone is present at high levels and therefore the effect of an antiandrogenic agent is only observed if it is present at high concentrations. Similarly, in castrated adult rats, detection of an antiandrogenic compound is based on its efficacy to block the action of an exogenously administered, potent androgen (testosterone or dihydrotestosterone). However, in gestational/prenatal exposure assays, antiandrogenic activity is detected as the efficacy of the EAC to antagonize the action of low levels of endogenous testosterone. Therefore, these assays are deemed to have higher sensitivities than the non-developmental assays. Yet, they are still dependent on the *Ki*s of the EACs, which likely approach the exposure levels induced by the NOAEL/LOAEL.

Recognizing that knowledge gaps exist in this area of studies, the following recommendations have been made. First, since only <u>one</u> study has been conducted in the range of the NOAEL/LOAEL, verification and substantiation of data from this study with independent (unrelated and multiple) investigations should be a top priority for immediate future research. Furthermore, since no studies had been conducted at dose-ranges below the NOAEL/LOAEL for vinclozolin and other environmental antiandrogens a need exists to test the hypothesis that the dose-response for antiandrogens is linear to the NOAEL/LOAEL. Secondly, although so far no EACs have been identified as androgen mimics for the human, it is necessary to develop mechanism-based assays for their detection since this class of compounds apparently affects wildlife. Thirdly, research can be focused on further advancement of the basic knowledge on the mechanisms of androgenic/antiandrogenic action. These efforts should benefit future development of new detection methodologies. Likewise, as new molecular markers of tissue response are identified (by genomics or proteomics discovery platforms), it will be useful to include such molecular/biochemical biomarkers as endpoints since they may be more sensitive or specific than current biological endpoints. With the advent of bioinformatics and large-scale molecular modeling it becomes attractive and cost-effective to utilize these new approaches to analyze currently available and future biological data in order to formulate novel hypotheses to be tested under new experimental paradigms. Along this line of argument, since the AR ligand-binding domain (LBD) has been crystallized it offers new opportunities for structural modeling of the AR-LBD-ligand crystals to predict androgenic/antiandrogenic activities of various chemicals. One important issue that future research has to address is the dose-response relationships for androgenic/antiandrogenic EACs in different species and in multiple strains. This issue is critical to our understanding of sensitive populations, encompassing such factors as genetic predisposition, age, gender, past and current exposure history, dietary influences and multiple chemical sensitivity. Modeling in animal studies may prove to be fruitful endeavors

preluding large-scale epidemiology studies. Lastly, it may be useful to develop credible dosimetry/mechanistic models for exposures occurring during in utero and early neonatal development since these are the most sensitive timepoints of detection.

References

1. Gray, L. E., Jr., Ostby, J., Monosson, E., and Kelce, W. R. Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat. Toxicol.Ind.Health, *15*: 48-64, 1999.

2. Health effects test guildelines OPPTS 870.3800 Reproduction and fertility effects. US EPA:Prevention, Pesticides and Toxics Substances (7101). EPA 712-C-98-208. 1998. Ref Type: Report

3. Anderson, C. A. and Clark, R. L. External genitalia of the rat: normal development and the histogenesis of 5 alpha-reductase inhibitor-induced abnormalities. Teratology, *42*: 483-496, 1990.

4. Ashby, J. and Lefevre, P. A. The peripubertal male rat assay as an alternative to the Hershberger castrated male rat assay for the detection of antiandrogens, oestrogens and metabolic modulators. J.Appl.Toxicol., *20*: 35-47, 2000.

5. Calabrese, E. J. and Baldwin, L. A. A general classification of U-shaped dose-response relationships in toxicology and their mechanistic foundations. Hum.Exp.Toxicol., *17*: 353-364, 1998.

6. Clark, R. L., Antonello, J. M., Grossman, S. J., Wise, L. D., Anderson, C., Bagdon, W. J., Prahalada, S., MacDonald, J. S., and Robertson, R. T. External genitalia abnormalities in male rats exposed in utero to finasteride, a 5 alpha-reductase inhibitor. Teratology, *42*: 91-100, 1990.

7. Clark, R. L., Anderson, C. A., Prahalada, S., Robertson, R. T., Lochry, E. A., Leonard, Y. M., Stevens, J. L., and Hoberman, A. M. Critical developmental periods for effects on male rat genitalia induced by finasteride, a 5 alpha-reductase inhibitor. Toxicol.Appl.Pharmacol., *119*: 34-40, 1993.

8. Ewing, L. L., Desjardins, C., Irby, D. C., and Robaire, B. Synergistic interaction of testosterone and oestradiol inhibits spermatogenesis in rats. Nature, *269*: 409-411, 1977.

9. Gray, L. E., Jr., Ostby, J. S., and Kelce, W. R. Developmental effects of an environmental antiandrogen: the fungicide vinclozolin alters sex differentiation of the male rat. Toxicol.Appl.Pharmacol., *129*: 46-52, 1994.

10. Gray, L. E., Jr., Wolf, C., Lambright, C., Mann, P., Price, M., Cooper, R. L., and Ostby, J. Administration of potentially antiandrogenic pesticides (procymidone, linuron, iprodione, chlozolinate, p,p'-DDE, and ketoconazole) and toxic substances (dib. Toxicol.Ind.Health, *15*: 94-118, 1999.

11. Greco, T. L. and Payne, A. H. Ontogeny of expression of the genes for steroidogenic enzymes P450 side- chain cleavage, 3 beta-hydroxysteroid dehydrogenase, P450 17 alpha-

hydroxylase/C17-20 lyase, and P450 aromatase in fetal mouse gonads. Endocrinology, *135*: 262-268, 1994.

12. Greene R.R and Burrill M.W., e. a. Experimental intersexuality. The effect of antenatal androgens on sexual development of female rats. The American Journal of Anatomy, *65(3)*: 415-469, 1938.

13. Hershberger LG and Shipley EG, e. a. Myotrophic activity of 19-Nortestosterone and other steroids determined by modified levator ani muscle method. Proceedings of the Society for Experimental Biology and Medicine, *83*: 175-180, 1953.

14. Hosokawa, S., Murakami, M., Ineyama, M., Yamada, T., Yoshitake, A., Yamada, H., and Miyamoto, J. The affinity of procymidone to androgen receptor in rats and mice. J.Toxicol.Sci., *18*: 83-93, 1993.

15. Imperato-McGinley, J., Sanchez, R. S., Spencer, J. R., Yee, B., and Vaughan, E. D. Comparison of the effects of the 5 alpha-reductase inhibitor finasteride and the antiandrogen flutamide on prostate and genital differentiation: dose-response studies. Endocrinology, *131*: 1149-1156, 1992.

16. Kelce, W. R., Stone, C. R., Laws, S. C., Gray, L. E., Kemppainen, J. A., and Wilson, E. M. Persistent DDT metabolite p,p'-DDE is a potent androgen receptor antagonist [see comments]. Nature, *375*: 581-585, 1995.

17. Monosson, E., Kelce, W. R., Lambright, C., Ostby, J., and Gray, L. E., Jr. Peripubertal exposure to the antiandrogenic fungicide, vinclozolin, delays puberty, inhibits the development of androgen-dependent tissues, and alters androgen receptor function in the male rat. Toxicol.Ind.Health, *15*: 65-79, 1999.

18. Ostby, J., Kelce, W. R., Lambright, C., Wolf, C. J., Mann, P., and Gray, L. E., Jr. The fungicide procymidone alters sexual differentiation in the male rat by acting as an androgen-receptor antagonist in vivo and in vitro. Toxicol.Ind.Health, *15*: 80-93, 1999.

19. Robaire, B., Ewing, L. L., Irby, D. C., and Desjardins, C. Interactions of testosterone and estradiol-17 beta on the reproductive tract of the male rat. Biol.Reprod., *21*: 455-463, 1979.

20. Wong, C., Kelce, W. R., Sar, M., and Wilson, E. M. Androgen receptor antagonist versus agonist activities of the fungicide vinclozolin relative to hydroxyflutamide. J.Biol.Chem., *270*: 19998-20003, 1995.

21. Yamada, T., Kunimatsu, T., Sako, H., Yabushita, S., Sukata, T., Okuno, Y., and Matsuo, M. Comparative evaluation of a 5-day Hershberger assay utilizing mature male rats and a pubertal male assay for detection of flutamide's antiandrogenic activity. Toxicol.Sci., *53*: 289-296, 2000.

22. Sannino, A., Bandini, M., and Bolzoni, L. Multiresidue determination of 19 fungicides in processed fruits and vegetables by capillary gas chromatography after gel permeation chromatography. J.AOAC Int., *82*: 1229-1238, 1999.

23. Kelce, W. R., Lambright, C. R., Gray, L. E., Jr., and Roberts, K. P. Vinclozolin and p,p'-DDE alter androgen-dependent gene expression: in vivo confirmation of an androgen receptor-mediated mechanism. Toxicol.Appl.Pharmacol., *142*: 192-200, 1997.

24. Fail, P. A., Pearce, S. W., Anderson, S. A., Tyl, R. W., and Gray, L. E. Endocrine and reproductive toxicity of vinclozolin (vin) in male Long-Evans Hooded rats. Toxicologist, *15*:293, 1995.

25. Kelce, W. R., Monosson, E., Gamcsik, M. P., Laws, S. C., and Gray, L. E., Jr. Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. Toxicol.Appl.Pharmacol., *126*: 276-285, 1994.

26. Maness, S. C., McDonnell, D. P., and Gaido, K. W. Inhibition of androgen receptor-dependent transcriptional activity by DDT isomers and methoxychlor in HepG2 human hepatoma cells. Toxicol.Appl.Pharmacol., *151*: 135-142, 1998.

27. Lambright, C., Ostby, J., Bobseine, K., Wilson, V., Hotchkiss, A. K., Mann, P. C., and Gray, L. E., Jr. Cellular and molecular mechanisms of action of linuron: an antiandrogenic herbicide that produces reproductive malformations in male rats. Toxicol.Sci., *56*: 389-399, 2000.

28. Wilson, V. S. and LeBlanc, G. A. The contribution of hepatic inactivation of testosterone to the lowering of serum testosterone levels by ketoconazole. Toxicol.Sci., *54*: 128-137, 2000.

29. Gray, L. E., Jr. Xenoendocrine disrupters: laboratory studies on male reproductive effects. Toxicol.Lett., *102-103*: 331-335, 1998.

30. Kemppainen, J. A., Lane, M. V., Sar, M., and Wilson, E. M. Androgen receptor phosphorylation, turnover, nuclear transport, and transcriptional activation. Specificity for steroids and antihormones. J.Biol.Chem., *267*: 968-974, 1992.

31. Wakeling, A. E. Steroid antagonists as nuclear receptor blockers. Cancer Surv., *14*: 71-85, 1992.

32. Freeman, S. N., Mainwaring, W. I., and Furr, B. J. A possible explanation for the peripheral selectivity of a novel non- steroidal pure antiandrogen, Casodex (ICI 176,334). Br.J.Cancer, *60*: 664-668, 1989.

33. Wilson, J. D., Griffin, J. E., and Russell, D. W. Steroid 5 alpha-reductase 2 deficiency. Endocr.Rev., *14*: 577-593, 1993.

34. Li, X., Chen, C., Singh, S. M., Labrie, F., and Labire, F. The enzyme and inhibitors of 4-ene-3-oxosteroid 5 alpha-oxidoreductase [published erratum appears in Steroids 1995 Aug;60(8):540]. Steroids, *60*: 430-441, 1995.

35. Gaillard-Moguilewsky, M. Pharmacology of antiandrogens and value of combining androgen suppression with antiandrogen therapy. Urology, *37*: 5-12, 1991.

36. Harris, M. G., Coleman, S. G., Faulds, D., and Chrisp, P. Nilutamide. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in prostate cancer. Drugs Aging, *3*: 9-25, 1993.

37. Dole, E. J. and Holdsworth, M. T. Nilutamide: an antiandrogen for the treatment of prostate cancer [see comments]. Ann.Pharmacother., *31*: 65-75, 1997.

38. Tyler C.R., Jobling S, and Sumpter J.P. Endocrine disruption in wildlife: a critical review of the evidence. Crit.Rev.Toxicol. 28, 319-361. 1998.
Ref Type: Abstract

39. Wolf, C., Ostby, J., Hotchkiss, A. K., and Gray, L. E. Effects of prenatal testosterone propionate on the development of male and female rats: A dose-response study. The 33rd Annunal Meeting of the Society for the Study of Reproduction. Biology of Reproduction 62, 247. 2000.
Ref Type: Abstract

40. Ware, G. W. Effects of DDT on reproduction in higher animals. Residue.Rev., *59*: 119-140, 1975.

41. Gray, L. E., Jr., Kelce, W. R., Wiese, T., Tyl, R., Gaido, K., Cook, J., Klinefelter, G., Desaulniers, D., Wilson, E., Zacharewski, T., Waller, C., Foster, P., Laskey, J., Reel, J., Giesy, J., Laws, S., McLachlan, J., Breslin, W., Cooper, R., Di Giulio, R., Johnson, R., Purdy, R., Mihaich, E., Safe, S., and Colborn, T. Endocrine Screening Methods Workshop report: detection of estrogenic and androgenic hormonal and antihormonal activity for chemicals that act via receptor or steroidogenic enzyme mechanisms. Reprod.Toxicol., *11*: 719-750, 1997.

Chapter 4:
Report of the Biological Factors and Study Design Subpanel

Chair
John Moore, Sciences International, Inc.

Rapporteur
Julian Leakey, National Center for Toxicological Research

Facilitator
William Allaben, National Center for Toxicological Research

Panelists
Sue Barlow, Consultant
Paul Foster, Chemical Industry Institute of Toxicology
Robert Luebke, U.S. Environmental Protection Agency
Robert Maronpot, National Institute of Environmental Health Sciences
Cory Teuscher, University of Illinois at Urbana-Champaign

Statistics and Dose-Response Modeling Subpanel Representatives
Kenneth Portier, University of Florida (statistics)
Rory Conolly, Chemical Industry Institute of Toxicology (modeling)

This Subpanel was not charged to review any particular chemical or group of chemicals. Therefore, it chose to selectively focus on topics within its area of expertise that it believes have generic application to design of studies and evaluation of data for evidence of possible endocrine disruption.

General conclusion. The Subpanel was of the consensus view that there is an adequate scientific basis to recommend development of a scheme that identifies and characterizes chemicals that may modulate or perturb endocrine systems. However, restricting scientific attention to schema that identify effects at low dose is, from a public health perspective, myopic. It is believed that within the cadre of chemicals found to have endocrine modulating characteristics within a traditional range of toxicological doses, will be found the majority of those chemicals that may cause effects at low doses in select circumstances. In our use of the modifier *low* in the previous sentence it has meaningful communicative value because it has a biological context. Observing that a chemical may produce effects at *low* doses is not revealing a new scientific concept; it is a broadened awareness that chemicals with endocrine modulating characteristics can exert effects in situations that had not been previously considered. For example, in the ontogeny of development there are temporal windows of sensitivity where exposure to a dose that is usually without effect can have undesired and lasting consequences. DES may serve as a classic example of estrogenic effects. There are other chemicals that appear to exert effects through anti-androgen influences such as linuron and dibutyl phthalate.

Factors that may account for discrepant results. The subpanel noted that some reports identified effects of exposure to bisphenol A (BPA) and diethylstibesterol (DES) during late gestation in CF-1 mice on subsequent prostate weight in the male offspring[1,2] and on the rate of sexual maturation in female offspring[3]. A mechanism by which such effects may occur has been proposed[4]. However, the reported low dose effects of BPA and DES could not be replicated by other laboratories[5,6] and the panel therefore assessed different parameters which might account for the lack or reproducibility of the low dose effects across laboratories. These were as follows:

Intrauterine Position (IUP):
The IUP phenomenon was first observed 20 years ago in mice and subsequently has been reported also to occur in rats, gerbils and guinea pigs[7]. It describes whereby a male fetus located in the uterus between two female fetuses (0M) is exposed to higher blood concentrations of estradiol and lower blood concentrations of testosterone than is a male fetus located between two males (2M). Similarly, an OM female fetus has higher estradiol and lower testosterone levels than a 2M female fetus. The effect is due to the passage of endogenous steroids between fetuses, transported across the placental membranes by amniotic fluid. Not only can raised levels of estradiol and lower levels of testosterone be detected in 0M fetuses relative to 2M fetuses prior to birth, but also alterations in endocrine-related endpoints have been observed after birth, which can be plausibly related to the early differences in the endocrine milieu[7]. These endpoints include variations in anogenital distance, prostate size and weight, prostate androgen receptor numbers and androgen-dependent behaviors in males, and timing of first estrus, estrous cycle length and duration of reproductive life in females.

Some of the above endpoints have also been reported to be influenced by low-dose exposures to certain endocrine-active compounds and one laboratory has reported interactions between IUP

and exposure to BPA or to 17ß-estradiol[1-3,8]. The Subpanel therefore considered whether tracking IUP might be a critical factor in revealing possible low-dose effects. It concluded, from the evidence available to date, that although IUP influenced the magnitude of the response to an exogenously administered compound, the low-dose effects reported were detectable irrespective of whether IUP was taken into account. The Subpanel therefore recommends the use of study designs that keep track of IUP, but that control of IUP is not essential for the observation of potential low-dose effects.

Strain/substrain Differences:
The CF-1 mice used in the negative studies[5,6] were obtained from a commercial breeding colony (Charles River), whereas those used in the positive studies[1,2] had been maintained as a colony in the university institute's animal facility for several years and had not been subjected to commercial selection pressures to maintain high rates of fecundity and body weight growth. When the 6-month necropsy data were compared, the mice used by Zeneca study[5] were observed to be heavier with greater individual weight variation than those used by Dr. vom Saal's group[1, 2]. The panel also concluded that genetic differences could also occur in mice of the same strain obtained form the same commercial breeder because most breeders maintain separate colonies at different facilities. Outbred rodent strains can exhibit genetic polymorphism in enzymes that metabolize androgens and estrogens. For example, Wistar rats have been shown to express a recessive mutation of the UGT2B2 gene which codes for the UDP-glucuronosyltransferase isoform that conjugates androgen metabolites such as androsterone[9,10].

Differences in Diet:
Different diets have been used by different groups and little or no attempt has been made to assay phytoestrogen content of individual batches of diets being used. However, studies performed with the synthetic diet AIN 76, which contains no soy products, was no less uterotrophic in prepubertal rats than open formula diets containing soy proteins[11]. The uterotrophic activity of AIN 76 could be inhibited by estrogen antagonists. The total caloric content of a diet can also influence reproductive endpoints. Caloric restriction in immature rats and mice can delay puberty and disrupt secretion of LH, FSH and growth hormone[12-14]. This would suggest that the caloric content of diets used in different studies might influence reproductive parameters and that anorexic effects of dosed feed due to problems with either palatability or toxicity might produce artifactual changes in reproductive endpoints. These effects would be more likely to occur at high doses than at low doses, and could therefore influence the shape of dose response curves.

Caging Considerations:
Different studies used different types of caging and bedding. For example some used stainless steel cages[6] whereas other studies used polypropylene cages[5]. The Subpanel was informed (F.S. vom Saal personal communication) that significant amounts of BPA could leach out of polycarbonate cages, particularly when the cages became old and had damaged or worn surfaces. Phytoestrogens could also be present in certain bedding materials. Differences in whether rodents are group-housed or individually housed could also influence study outcome. The housing protocol used in the studies that showed low dose effects on prostate weight used group housing of the male mice (3 per cage) from weaning until they were 5 months old. The mice were then individually housed for 1 month prior to evaluation[1]. Dr. vom Saal stated to the

Subpanel that during the group housing stage a single male in each cage would acquire dominance over the other two males and the dominant male will exhibit larger accessory reproductive organs (including the prostate) than the subordinate males. Individually housed males all had similar sized accessory reproductive organs as group-housed dominant males. Thus, under the experimental paradigm that showed low dose effects subordinate mice would be expected to exhibit more rapid catch-up growth of the prostate gland during the month prior to evaluation. Only a single mouse from each original cage was evaluated for low dose effects, but it was not known whether either dominant or subordinate mice were selected or whether the mice were randomly selected. This situation differed in the studies that failed to show low dose effects with BPA. These either individually housed the mice for the entire period between weaning and evaluation[15], or used the vom Saal procedure but evaluated all three mice from each cage[5].

Dosing Considerations:
Although all three groups utilized the same doses and similar dosing techniques the extremely low doses used precluded the possibility of detailed pharmacokinetic studies to confirm that similar amounts of the test chemicals were reaching the fetal target tissues.

Temporal/Seasonal Differences:
Many biological parameters such as immune function assays in both humans and rodents vary in magnitude with the season of the year[16]. Seasonal variation is apparent in rodents kept under constant temperature and light cycle, suggesting that true circannual rhythms do occur[17]. Furthermore, there appears to be a genetic component to peak season of responsiveness, since different strains of mice make maximum responses during different seasons[16]. A study which utilized over 9,000 mice maintained for over several years under a constant light/dark cycle and temperature reported that the ratio of males to female offspring varied significantly between spring and fall with more females being born in the springtime[18]. The study also reported that the differences in sex ratio also resulted in a significant seasonal variation in proportions of 2M to 0M male and female pups. The authors suggested that there might be an evolutionary advantage to this phenomena because 2M progeny have been reported to be more aggressive and less nurturing than 0M progeny[19] and would be more likely to survive through the harsh conditions of winter.

Comments on the sensitivity of "Tier 2" multigeneration test to detect effects.

The traditional multigeneration reproduction study protocol has in the last few years been utilized to characterize the potential hazards of endocrine active chemicals. Adopting this protocol in large part is due to the study design that incorporates exposure of animals through all the critical windows of sexual differentiation in an F_1 generation and assessment of an F_2 generation through postnatal day 21. The multigeneration reproduction study has served toxicology well over the last twenty or more years and has essentially maintained its original design. The study provides substantial information of the effects of agents on reproduction (e.g. fertility, fecundity, pregnancy, gametes etc) with more limited information on (postnatal) development (e.g. pup survival, growth, developmental landmarks etc).

When used to test endocrine active chemicals the major purpose of the "Tier 2" test is to provide "definitive" information on hazard characterization of endocrine active chemicals. The test is

therefore expected to: (1) confirm or refute observations noted in Tier 1screens/ assays; (2) identify activity with special regard to end points for which concern has been raised in humans (e.g. decreased sperm count, cryptorchidism, reproductive tract malformations); (3) identify other endocrine activity of interest; and (4) provide the critical dose-response information on endocrine active chemicals to be used in risk assessment.

In contrast to the prenatal toxicology study in which every pup is examined in at least 20 litters, the multigeneration protocol goes through a series of reductions in number of animals for evaluation. Thus although all animals are exposed through gestation, frequently litter size is reduced on post-natal day 4 (usually to 4 males and 4 females) to "standardize" the litter although there are arguments both for and against this reduction in size [22, 23]. Litter size is then further reduced at weaning (usually pnd 21) at which time only one animal/sex/litter is selected to carry on until adulthood. Even at this stage a further reduction occurs in that only 10 animals per group are scheduled for pathological investigation.

There are several severe limitations to the "Tier 2" test, particularly with regard to the detection of low incidence phenomena (e.g. reproductive tract malformations). A major shortfall is that only one pup/sex/litter of F_1 is examined at adulthood. Examining but one pup per sex provides inadequate power to detect reproductive tract malformations, i.e., the possibility exists to produce false negatives. It also can lead to false positives if the control population is variable (e.g. prostate weight[24]). Additional deficiencies are that a number of sensitive end points (e.g. retention of thoracic nipples in male offspring, measurement of anogenital distance) are not evaluated or are only triggered in F_2 pups that are terminated and necropsied at weaning. Gross necropsy at weaning is unlikely to detect subtle malformations (e.g. epispadias), and will not detect effects on organ systems not yet developed (e.g. sperm production, prostate). A number of agents have now been shown to have endocrine activity even though well conducted multigeneration and prenatal studies by competent laboratories were negative in the standard study design (e.g. linuron[25]; DINP[26]). This causes concern for the ability of these guideline studies to fulfill the needs outlined above and provide the necessary information with regard to hazard characterization and use in risk assessment.

Other Design Factors Warranting Consideration

Animal model selection:
The Subpanel asserts that the selection of species or strain for future studies should be the product of a more deliberate thought process. In a review of the literature provided to the Workshop it is apparent that test animal selections were driven by availability, convenience and familiarity. It should therefore, come as no surprise that the reviewed data showed a wide range of responses. There would be value in developing a core of data across a selected array of mouse and rat genotypes with a modest set of chemicals known to possess endocrine active properties of interest. The experimental design for such a study should reflect a spectrum of inbred and random-bred genetic backgrounds and incorporate parameters that reasonably characterize endpoints of interest. The results of such an effort could provide a basis for reasoned selection of test animals. It may also provide a basis for modifying and realigning current protocols with respect to dose groups, group size and parameter and endpoint selection.

Species/strain Selection:
While the abundance of historical control data makes the CD-1 mouse and Sprague Dawley rat attractive animal models for reproductive toxicity testing, inbred strains such as the $B6C3F_1$ mouse may produce less variable endpoints. For example, inbred mouse strains do not appear to exhibit differences in sexual behavior and other endpoints that are related to intrauterine position (F.S. vom Saal personal communication). Furthermore the apparent advantage of historical control data has been compromised for several rodent strains due to selective breeding and/or genetic drift[27-29].

Pharmacokinetic Data:
When establishing the reproducibility of low dose effects it is important to establish what concentration of the test chemical or its active metabolites reaches the target tissue. Because of the extremely low doses of the test chemicals being used in these studies, more sensitive analytical methods, such as radioimmunoassay, need to be developed.

Control of Experimental Variation:
Food consumption, body weight and stress can all influence reproductive endpoints and influence the outcome of toxicity and carcinogenicity bioassays[30]. Controlled feeding can reduce variation in body weight and in relative organ weights in $B6C3F_1$ mice[31] but its effects on reproductive endpoints have not been investigated. Culling litters to set numbers of pups will help standardize weight gain and rate of development in suckling rodents, but the number of fetuses *in utero* also influences weight gain and energy expenditure during postnatal life[32]. Group housing of male mice results in differences in prostate size in dominant and subordinate mice. While individual housing may eliminate this variability the low dose effects of BPA were only detected in group housed mice. Establishment of reproducible low dose effects on prostate weight might require establishing the subordinate or dominant status of test animals.

Establishment of Multiple Parameters for an Endpoint:
To gain broader acceptance that endpoints such as prostate enlargement or uterine maturation are of biological or toxicological consequence the reporting that an end point is affected needs to be supported by multiple parameters. Further, parameters that assess function are of greater value. In particular, behavior or more mechanistic endpoints such as serum hormone levels, tissue hormone receptor expression or enzyme activity should be utilized. Emerging techniques of genomics and proteomics[33,34] could be extremely useful in establishing potential biomarkers for endocrine disruption.

Associate Endocrine Disruption with Pathological effects or Toxicity:
Experimental models need to be designed that will determine whether alterations in endocrine function such as prostate enlargement or accelerated uterine development are associated with clinically relevant pathological conditions such as prostrate or uterine/ovarian cancers. Transgenic mice with increased susceptibility to these pathologies may prove useful here as would susceptible strains such as the Lobund/Wistar rat[35].

Establishment of Windows of Susceptibility:
There is a need to map windows of susceptibility for endocrine disrupting chemicals by using knowledge of the developmental profiles of hormone receptors in target tissues and of plasma

binding proteins, steroid metabolizing and xenobiotic detoxicating enzymes in the developing organism. Such windows should then be tested empirically using the endpoints listed above. In certain cases F_2 generations may have to be evaluated until adulthood.

Immune System Endpoints:
To determine whether low dose immuno-toxicological effects occur with endocrine active chemicals requires the use of a comprehensive screen. General indices of immune system development and function such as lymphoid organ weights and cellularity, bone marrow function can provide clues of loss or shifts in cell type in lymphoid tissues. Hints that there are defects in bone marrow function can be gleaned from erythrocyte counts plus total and differential leukocyte counts or phenotypic analysis of immune system cells. However, functional assays that evaluate whole body immuno-reactivity, e.g., antibody production and delayed-type hypersensitivity responses to injected antigens, or other assays that reflect a specific response to antigen, should be conducted if immunosuppression is suspected. Burns et al. present and discuss methods to assess immunocompetence[36].

References

1. Nagel SC, Vom Saal FS, Thayer KA, et al: Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ Hlth Persp 105:70-76, 1997.

2. vom Saal FS, Timms BG, Montano MM, et al: Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc Natl Acad Sci USA 94:2056-2061, 1997.

3. Howdeshell KL, Hotchkiss AK, Thayer KA, et al: Exposure to bisphenol A advances puberty. Nature 401:763-764, 1999.

4. Sheehan DM: Activity of environmentally relevant low doses of endocrine disruptors and the bisphenol A controversy: initial results confirmed. Proc Soc Exp Biol Med 224:57-60, 2000.

5. Ashby J, Tinwell H, Haseman J: Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF1 mice exposed *in utero*. Regul Toxicol Pharmacol 30:156-166, 1999.

6. Cagen SZ, Waechter JM,Jr., Dimond SS, et al: Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A. Toxicol Sci 50:36-44, 1999.

7. vom Saal F, Finch CE and Nelson JD. Natural history and mechanisms of reproductive aging in humans, laboratory rodents and other selected vertebrates. Chapter 61 In: Physiology of Reproduction, 2nd Ed, Vol 2, eds E Knobil and JD Neill, Raven Press Ltd, New York, pp1213-1314, 1994.

8. Howdeshell KL and vom Saal F Developmental exposure to bisphenol A: interaction with

endogenous estradiol during pregnancy in mice. American Zoologist 40:429-437, 2000.

9. Homma H, Kawai H, Kubota M, et al: Large deletion of androsterone UDP-glucuronosyltransferase gene in the inherited deficient strain of Wistar rats. Biochim Biophys Acta 1138:34-40, 1992.

10. Satoh H, Nagai F, Homma H, et al: Regional assignment of rat androsterone UDP-glucuronosyltransferase gene (UGT2B2) to chromosome 14p21.2-p22. Cytogenet Cell Genet 62:49-51, 1993.

11. Ashby J, Tinwell H, Odum J: Uterotrophic activity of a "phytoestrogen-free" rat diet. Environ Hlth Persp 108:A12-A13, 2000.

12. Holehan AM, Merry BJ: Lifetime breeding studies in fully fed and dietary restricted female CFY Sprague-Dawley rats. 1. Effect of age, housing conditions and diet on fecundity. Mech Ageing Devel 33:19-28, 1985.

13. Hamilton GD, Bronson FH: Food restriction and reproductive development: male and female mice and male rats. Am J Physiol 250:R370-R376, 1986.

14. Sisk CL, Bronson FH: Effects of food restriction and restoration on gonadotropin and growth hormone secretion in immature male rats. Biol Reprod 35:554-561, 1986.

15. Celius T, Haugen TB, Grotmol T, et al: A sensitive zonagenetic assay for rapid *in vitro* assessment of estrogenic potency of xenobiotics and mycotoxins. Environ Hlth Persp 107:63-68, 1999.

16. Ratajczak, HV, Thomas, PT, Sothern, RB, Vollmuth, T and Heck, JD. Evidence for genetic basis of seasonal differences in antibody formation between two mouse strains. Chronobiol Internat 10, 383-394.

17. Brock, MA. 1983. Seaspma; rhythmicity in lymphocyte blastogenic responses of mice persist in a constant environment. J Immunol 130, 2586-2588 1993.

18. Novikov SN, Churakov GA: Seasonal variability of the intrauterine position of the fetus in laboratory mice as a potential microevolution vector. Doklady Akademii Nauk 368:717-720, 1999.

19. Clark MM, Vonk JM, Galef BG,Jr.: Intrauterine position, parenting, and nest-site attachment in male Mongolian gerbils. Devel Psychobiol 32:177-181, 1998.

20. Newbold RR, Hanson RB, Jefferson WN, et al: Increased tumors but uncompromised fertility in the female descendants of mice exposed developmentally to diethylstilbestrol. Carcinogenesis 19:1655-1663, 1998.

21. Newbold RR, Hanson RB, Jefferson WN, et al: Proliferative lesions and reproductive tract

tumors in male descendants of mice exposed developmentally to diethylstilbestrol. Carcinogenesis 21:1355-1363, 2001.

22. Agnish, N. D. & K. A. Keller: The rationale for culling of rodent litters. Fundam Appl Toxicol 38:2-6,1997.

23. Palmer, A. K. & B. C. Ulbrich: The cult of culling. Fundam Appl Toxicol 38:7-22,1997.

24. Elswick, B. A., F. Welsch & D. B. Janszen: Effect of different sampling designs on outcome of endocrine disruptor studies. Reprod Toxicol 14:359-67, 2000.

25. McIntyre, B. S., N. J. Barlow, D. G. Wallace, S. C. Maness, K. W. Gaido & P. M. D. Foster: Effects of in utero exposure to linuron on androgen-dependent reproductive development in the male Crl:CD(SD)BR rat. Toxicol Appl Pharmacol 167:87-99, 2000.

26. Gray, L., J. Ostby, J. Furr, M. Price, D. Veeramachanemi & L. Parks: Perinatal exposure to the phthalates DEHP, BBP, and DINP but notDEP,DMP or SOTP alters sexual differentiation of the male rat. Toxicol Sci 58:350-365, 2000.

27. Roe FJC: What does carcinogenicity mean and how should we test for it? Food Chem Toxicol 31:225-229, 1993.

28. Seilkop SK: The effect of body weight on tumor incidence and carcinogenicity testing in B6C3F_1 mice and F344 rats. Fundam Appl Toxicol 24:247-259, 1995.

29. Bronson FH: Energy allocation and reproductive development in wild and domestic house mice. Biol Reprod 31:83-88, 1984.

30. Leakey JEA, Seng JE, Barnas CR, et al: A mechanistic basis for the beneficial effects of dietary restriction on longevity and disease. Consequences for the interpretation of rodent toxicity studies. Intnl J Toxicol 17 Suppl.2:5-57, 1998.

31. Seng JE, Allaben WT, Nichols ML, et al: Reduction of experimental variability within animal assays: A method for controlling body weight using restricted feeding. Lab Animal 27:35-38, 1998.

32. Haigh GR, Bronson FH: Variation in litter size encountered in utero influences the bioenergetic characteristics of adult female mice. Physiol Behav 43:831-833, 1988.

33. Rockett JC, Dix DJ: Application of DNA arrays to toxicology. Environ Hlth Persp 107:681-685, 1999.

34. Brewis IA: Proteomics in reproductive research: the potential importance of proteomics to research in reproduction. Human Reprod 14:2927-2929, 1999.

35. Pollard M: Prevention of prostate-related cancers in Lobund-Wistar rats. Prostate 39:305-309, 1999.

36. Burns LA, Meade BJ, Munson AE: Toxic responses of the immune system in *Casarett and Doull's Toxicology, The Basic Science of Poisons*; CD Klaassen ed. 368-373,1996

Chapter 5:
Report of the Statistics Subpanel

Chair:
Joseph Haseman, National Institute of Environmental Health Sciences

Panelists:
John Bailer, Miami University of Ohio
Ralph Kodell, National Center for Toxicological Research
Richard Morris, Analytical Sciences, Inc.
Kenneth Portier, University of Florida

BACKGROUND

The Endocrine Disruptor Meeting's Peer Review Panel and associated expert Subpanels are charged with the responsibility of reviewing the available endocrine disruptor data and reaching conclusions regarding the possible presence of low dose effects. To help facilitate this effort, the Organizing Committee identified and requested raw data from 58 selected studies, involving 15 different sets of investigators, that the Committee felt were relevant to this evaluation. These data were primarily from published studies, but included some unpublished datasets for which a manuscript (or Abstract) had been prepared. The Statistics Subpanel was then asked to examine these data, to re-evaluate the authors' experimental design, data analysis, and interpretation of experimental results, and to provide a written report, which would be distributed to the various Subpanels to aid in their deliberations. The study investigators were also asked to submit responses to 23 specific questions which were designed to help the Subpanels better understand important features of each study. The investigators' responses to the 23 questions are summarized in Appendix B.

Before we began our data evaluation, the Statistics Subpanel requested the assistance of experts in the field to help prioritize the studies and variables for statistical analysis. The Organizing Committee responded by identifying the specific variables that they wished for the Statistics Subpanel to evaluate from each study, which often were simply all the data summarized in the publication. The Organizing Committee then prioritized the studies in terms of the desired order of data evaluation by the Statistics Subpanel. Our Subpanel is grateful to the various investigators for their willingness to submit their raw data to us for re-evaluation.

The Chairman of the Statistics Subpanel (Joseph Haseman) assigned a primary statistical evaluator for each study. To avoid the appearance of a conflict of interest, statisticians were generally not assigned primary responsibility for studies in which they may have directly or indirectly been involved. Thus, Haseman was not the primary evaluator of the data from Chapin or Newbold (NIEHS scientists), nor from most of the NCTR Delclos studies (which had some involvement with NIEHS scientists). Similarly, he did not have primary responsibility for two studies - one from Ashby; the other from vom Saal - in which he was a coauthor of a manuscript reporting study results. Ralph Kodell (NCTR) was not the primary statistical evaluator of the Delclos studies. Richard Morris (through his association with Analytical Sciences, Inc.) had indirect involvement with the Chapin studies. Neither of the other two statisticians had previous direct or indirect involvement with any of the 58 studies.

There were several important limitations associated with our data evaluation. These included

(i) large number of data sets - The final total of 58 studies for which data was requested was approximately double the original estimate. The large number of studies (and the limited time frame) made it impossible for us to evaluate all of the submitted data by the time of the October meeting. We completed statistical analyses of 38 studies from twelve different investigators, and these individual writeups are summarized in Appendix A. Importantly, the studies with

submitted data that we were unable to analyze (most notably the datasets of Newbold and Chapin) should not be viewed as being flawed in any way. The primary reason for the lack of statistical analysis was simply the lack of time, as is discussed below.

(ii) limited time frame - Only two investigators submitted data by the original requested submission date of August 4. Most datasets were received in late August or early September, with one high priority data set being received as late as September 19. Since the Organizing Committee also requested that the draft report of the Statistics Subpanel be distributed one week prior to the October 10-12 meeting, that left only approximately 4-6 weeks for the Subpanel to evaluate the available data and to prepare our report. The Statistics Subpanel felt that this was insufficient time to evaluate the data as thoroughly as we would have liked.

(iii) incomplete documentation/submission of data - Our task was made more difficult in those instances in which the only reference document was an Abstract rather than a full length publication, The raw data submitted to us were not always clearly defined (especially in those instances in which the data submission was in a non-English language), and we found it very helpful to have full length publications with data summary tables to serve as a clear reference point for our statistical analyses. In several studies, comparison of the raw data with the published summary tables identified important errors in the raw data provided to us, which required corrective measures in cooperation with study investigators. Moreover, certain key investigators did not provide the Statistics Subpanel with all the data that were requested.

(iv) The focus of our analysis was on individual studies - The Statistics Subpanel did not have the time to systematically compare and contrast results across different studies/investigators and to speculate as to why similar or differing results were observed. While this is certainly an important issue, our focus was primarily on the experimental design, data analysis, and interpretation of experimental results of each individual study within the context of its own experimental conditions. Consideration of the various statistical issues discussed in this report and a detailed analysis of the responses to the 23 questions would be of value in any follow up study that deals with the broader issue of comparing study outcomes of different investigators.

(v) statistics alone cannot resolve the low dose issue -No statistical analysis or set of statistical analyses can totally resolve the basic scientific issue of whether or not biologically important low dose effects truly exist. For example, chemical-related low dose effects may or may not be "statistically significant," but the scientific value of the statistical analyses that we provide is a function of the quality of the data given to us, which is beyond our control. It was not possible for the Statistics Subpanel to assess the validity or reliability of the data we received.

The purpose of this statistical reevaluation was to provide an independent assessment of the experimental design and data analysis used in each of the studies and, perhaps even more importantly, to identify and discuss key statistical issues relevant to all studies. Although the assessment of individual studies is given in Appendix A, the main focus of this report will be on important experimental design and data analysis issues that affect the evaluation and interpretation of all endocrine disruptor studies. Such issues will be illustrated using examples from the various studies that we evaluated.

STATISTICAL APPROACH

The 38 studies that we evaluated used a variety of statistical methods. Recognizing that each study had its own objectives, it was nevertheless decided that our evaluation would use a uniform statistical approach that would be applied to all studies (see discussion of the choice of statistical methodologies given below). Since the primary objective of most of these studies was to determine if significant effects were present in selected dosed groups relative to controls, pairwise comparisons were made by Dunnett's test. Analysis of Variance (ANOVA) was used to account for specific design effects (e.g., replicate effects) in addition to dose effects. Linear mixed-effects models were often employed using litter as a random effect and allowing for responses from littermates to be correlated. In addition to accounting for litter effects and replicate effects, we also used Analysis of Covariance (ANCOVA) in the evaluation of organ weights to adjust for body weight differences among groups. Since low dose effects were of interest, regression models (linear and quadratic) were also used to study dose-response trends.

When appropriate, a logarithmic transformation was used to eliminate heterogeneity of variances across treatment groups. In those few instances in which heterogeneity could not be removed by a log transformation, nonparametric techniques were used. The most common situation for which we used nonparametric procedures was highly skewed data such as mammary gland differentiation in which >90% of the responses were zeros. It would not have been appropriate to use parametric methods in such instances.

Although a uniform statistical approach was used, we retained the flexibility of carrying out any statistical analyses of the data that any individual statistical evaluator deemed to be appropriate. More specific details on the statistical methods used in our analyses of each study is given in Appendix A.

Each evaluator prepared a written report summarizing the results of his re-analysis of each study for which he had primary responsibility. The Chairman of the Statistics Subpanel then merged these writeups into a single report, which is included here as Appendix A. The primary statistical evaluator for each study is also identified in Appendix A.

IMPORTANT STATISTICAL ISSUES

As noted earlier, this report will focus on important statistical issues that arise both in the areas of experimental design and data analysis. These matters are discussed in more detail below.

I. Experimental Design Issues

A. Study sensitivity (power) - One important experimental design consideration is a study's power, which is defined as the probability of detecting a treatment effect if it is present in the data. Study sensitivity or power is influenced by a number of factors: (i) sample size; (ii) the underlying variability of the data; (iii) the magnitude of the treatment effect that is present; and

(iv) the method of statistical analysis and the associated level of significance chosen.

Obviously, a larger study will generally have more power for detecting chemical-related effects than a smaller study. Moreover, the interpretation that a study is "negative" should be given more weight when relatively large sample sizes are used. The number of animals per group ranged from 3 to 179 in the studies that were re-evaluated, and this is a factor that must be considered when comparing and interpreting study results.

Importantly, the effective sample size of a study is the number of independent sampling units. Thus, if littermates are used and litter effects are present in the data, the effective sample size becomes the number of litters, not the number of individual pups. This matter is discussed in more detail below.

Although we would in general anticipate that larger group sizes would lead to greater statistical power, this may not be realized if larger group sizes are obtained at the cost of introducing new uncontrolled sources of variability. Reducing variability should be an important study objective, and this can be achieved in a variety of ways. One is to identify and control those factors most likely to produce variability in response. Selected sources of variability are discussed in more detail below as general experimental design issues.

B. Replication - Reproducibility of experimental results is an important and necessary feature of any scientific finding before it can be generally accepted as valid. There are several types of replication, which are discussed below.

First, there is replication within an individual experiment. If multiple replications are used within a study, then each experimental group should be represented in each replicate. In one experiment we evaluated, three replicates were used, but the mid and high dose groups (which had only three animals per group) were represented only once, and in different replicates. Additionally, there were significant differences among the control groups in the three replicates, although the study authors pooled these groups in their statistical analysis. This is not an ideal experimental design or data analysis.

In another study we investigated, control and three dosed groups were each evaluated in separate time frames, extending over a period of one year. The Statistics Subpanel felt that the lack of concurrent controls was a serious deficiency of the experimental design that greatly limited the general inferences that could be drawn from this study.

Another type of replication is the reproducibility of results among separate experiments within a given laboratory. In one publication we evaluated, the investigator carried out eight similar experiments with the same chemical, although these technically were not replicates, because different dose levels of the test compound were used in some experiments. This investigator found statistically positive effects on uterus weight in four experiments and no effect in the other four experiments. The author concluded that his investigation had shown that even the same investigator may be unable to repeat experimental findings, and we agree with this conclusion.

Perhaps the most important type of replication is reproducibility among different laboratories

trying to confirm the findings of another laboratory. Among the data sets we evaluated, there were several studies that attempted to duplicate the studies of other investigators. Some confirmed the original results, but many did not. It is difficult to achieve exact reproducibility of all aspects of an experimental design, and when conflicting results are obtained by different investigators, one should try to identify study differences that could account for the contradictory results. That is one reason that we requested answers to the 23 questions noted previously, which we hoped would help identify the sources of variability that were most likely responsible for certain findings not being replicated from study to study. Although the limited time available to us did not permit a comprehensive analysis of this information and its impact (if any) on the reproducibility of experimental outcomes from study to study, this is certainly an important matter worthy of future study.

C. Litter effects - Using data from littermates is neither an inherently good or bad experimental strategy, but if littermate data are to be used, it is essential that this source of variability be taken into account in the statistical analysis. Some of the studies we evaluated used littermates; others did not. However, in the studies that used littermates, there was generally a significant "litter effect," indicating that the pups within a litter were responding more "alike" than pups from different litters. Failure to adjust for litter effects (e.g., to regard littermates as independent observations and thus the individual pup as the basic experimental unit) can greatly exaggerate the statistical significance of experimental findings.

One of the studies we evaluated carried out a simulation study in which they concluded that one pup per litter experimental designs "should not be used when assessing effects on highly variable organ weights and other reproductive endpoints," since such a design results in "a substantial percentage of incorrect conclusions about the presence or absence of treatment effects." However, their simulation study and its conclusions were flawed for several reasons (see Appendix A for more details).

For a fixed total number of litters, increasing the number of pups per litter will increase power (will reduce the false negative rate), but it has no impact on the false positive rate. The false positive rate associated with a particular statistical methodology is fixed by the selection of alpha (typically 0.05). If the null hypothesis is true and there is no difference among the experimental groups, then the p value (i.e., the actual false positive rate) should be essentially equal to alpha.

None of the authors' simulations indicated that the sampling strategy used (one, two or three pups per litter) had any impact on what they regarded to be the false positive rate. Nevertheless, the authors concluded that "the sampling of only one or two pups per litter" may have been a "contributing factor" to the positive low dose effects observed by some investigators, effects that were not confirmed by others who used more than two pups per litter. However, nothing in the authors' paper supports their speculation that the significant low dose effects were merely false positive outcomes resulting from the use of only one or two pups per litter.

Moreover, none of the authors' simulations actually assessed false positive rates in any case. A simulation study should be based on comparing samples selected at random from underlying populations. If the underlying populations are identical, then the simulations assess false positive rates; if they are different, then the simulations assess false negative rates (power).

The authors' simulations were based on comparisons of subsamples selected without replacement from two (or more) finite samples with different observed mean responses. In those instances in which the samples themselves were not statistically different, the authors considered the samples to be identical, and then used them as populations in their simulation study. However, since the samples (now regarded as populations) were in fact different, all of the authors' reported "false positive rates" (based on the comparisons of subsamples selected from the nonidentical samples) were in fact power calculations. Moreover, the use of small finite "populations" greatly limits the possible p values that could result from a multiple pup per litter sampling strategy.

Finally, the authors' recommendation to increase the number of pups per litter may be misleading. If strong litter effects are present, then the gain in power (for a fixed total number of pups) is best achieved by increasing the number of litters, not by increasing the number of pups per litter. Why is this so? For data showing significant litter effects, the within group variation is dominated by variation among dams resulting in high correlation among pups within dams. Since the individual pups do not respond independently, the appropriate experimental unit is the litter, not the individual pup.

For example, when significant litter effects are present, a study with dosed groups comprised of 20 pups will have more power if each of the 20 pups is from a separate litter rather than having four pups from each of five litters. The false positive rates will be identical in both cases. Thus, the authors' emphasis on increasing the number of pups per litter rather than increasing the number of litters is misguided.

In an ANOVA-based statistical analysis, if littermates are used only within a single treatment group, then litter is a nested factor. However, in some experimental designs, one pup from the same litter may be assigned to each experimental group prior to treatment. In such instances, litter is a crossed, not a nested factor. This distinction is important, since regarding crossed factors as nested factors or vice versa in an ANOVA can result in a very misleading test.

Another potentially complicating "litter effect" is the location of the pup within the uterus (e.g., whether it is located between two males, between two females, or between one male and one female). At least one investigator has data indicating that intra-uterine position can influence certain biological responses. Thus, intra-uterine position is another potentially important source of variability that must be considered by study investigators.

D. Potential investigator bias - To avoid the possibility of subtle bias, post-experimental measurements on pups should be made without prior knowledge of whether they are from dosed or control groups. In other contexts (e.g., histopathology evaluation), it has been argued that control animals need to be examined in an unblinded fashion to identify what is "normal variability" and only then can experimental groups be evaluated. The Statistics Subpanel does not accept this argument (which is debatable even in other contexts), since the primary variables of interest (organ weight; anogenital distance; sperm counts, etc.) are objective measures that do not require prior knowledge of control values to be accurately assessed.

A closely related issue is the order of experimental evaluation. Here again, some have argued

that the controls must be examined first to ascertain what is "normal variability." The Statistics Subpanel rejects this argument in the current context, and feels that dosed and control groups should be examined "together" in a blinded fashion. While it may (or may not) be impractical to use a completely randomized order of evaluation, at a minimum the experimental design should ensure that there is no bias associated with a systematic ordering of the data evaluation.

Responses to the 23 questions by one of the laboratories revealed a potential problem in this area (see Appendix B). This laboratory reported that they evaluated all the controls (unblinded) on one day, followed by all the low dose animals on the next day, etc. In response to another question, this lab indicated that within a day, a single technician looked at all the pups, but different technicians might be used on different days. The potential bias in such a strategy should be obvious.

Incidentally, each laboratory should avoid situations in which one technician or scientist is responsible for generating data for the control group, while other technicians/scientists are responsible for the dosed groups. Even if the technicians/scientists have been uniformly trained, it is difficult to avoid differences among them, and there is no reason to have this potential source of variability present in a well designed study.

E. Differing types of control groups - Some studies used both an untreated and a vehicle control group to evaluate the possible effect of the test vehicle. In most studies it was not expected that the vehicle would have an effect (an expectation that was generally confirmed), and the control groups were subsequently pooled. We agree that the pooling of vehicle and untreated control groups is reasonable if there is no evidence of a difference between them. If there is any evidence of a possible vehicle effect, then the two control groups should not be pooled, and the primary comparisons of interest should be relative to the vehicle control group.

F. Quality Control - A study's experimental design should include procedures to ensure the accuracy of data recording/transcription. Although subsequent tests for outliers can identify questionable data points, by that time it may be too late to know for certain the accuracy of such values. In some studies we evaluated, there were organ weights that were ten fold or even 100 fold greater than the mean of the other values in the group, and yet these points were included in the authors' data analysis.

There are also quality control issues with respect to the raw data provided to us for statistical analysis. In two studies, the raw data provided to us had significant errors. In one case (involving two different studies from the same investigator) multiple pups were mis-assigned to litters, and multiple litters were mis-assigned to dosed groups. Another investigator inadvertently omitted in his raw data submission entire blocks of data dealing with eight high dose animals. These errors were detected (and corrected) only because we had access to summary data in published papers for comparative purposes. For raw data submissions in which the only reference information is an Abstract, such errors would likely have gone undetected.

II. Data Analysis Issues

A. Choice of statistical methodology - There are many different statistical procedures that may

be used in a given experimental setting. However, it must be recognized that these procedures may have different objectives, make different underlying assumptions, and have different degrees of protection against false positive and false negative outcomes. In this context, one procedure is more "conservative" than another if it tends to have a lower false positive rate and a correspondingly higher false negative rate. Balancing false positive and false negative rates in the selection of statistical methodology is to some extent a matter of scientific judgement.

To take a specific example, for normally-distributed data for which the desired comparisons are limited to dosed groups vs. controls, Dunnett's test is a widely used and appropriate test for this purpose. This is the method of statistical analysis we elected to use, as noted above. Dunnett's test controls the experiment-wide error rate by taking into account the multiple comparisons being made, and thus is more appropriate than, for example, multiple applications of Student's t test, which could result in an unacceptably high false positive rate. Williams has proposed a modification to Dunnett's test that is appropriate if it is reasonable to assume a monotonic dose-response, but because of the potential for "low-dose effects" not seen at higher doses (which would invalidate the monotonicity assumption), we decided not to use the Williams procedure.

In our re-analysis, we found that even if two investigators choose the same test procedure, they may apply it differently. For example, Dunnett's test is a "stand-alone" test that does not require the statistical significance of an overall ANOVA to be valid. However, many investigators who used Dunnett's test required statistical significance of an overall ANOVA before making pairwise comparisons, a linkage that our Subpanel does not feel is necessary or appropriate. Since the critical values for Dunnett's test were derived without consideration of an overall ANOVA, requiring this additional significance may result in a slightly conservative test. Specifically, there were a few instances in which our reanalysis found significant pairwise differences by Dunnett's test that were not reported as significant by the study investigators who themselves also used Dunnett's test. Such differences were apparently due to the extra requirement of a significant overall ANOVA imposed on Dunnett's test by the study investigator.

If all possible pairwise comparisons are desired, then there are many multiple comparison tests that could be used. For a comparison of certain of these methods, see Carmer and Swanson (Journal of the American Statistical Association 68: 66-74, 1973) and Hochberg and Tamhane (Multiple Comparison Procedures, John Wiley, New York, 1987). One such method is the widely-used Fisher's (protected) Least Significant Difference (LSD) test, which, unlike Dunnett's test, does require the significance of an overall ANOVA before pairwise comparisons can be made, in order to control the experiment-wide error rate. The conditional pairwise comparisons are then made by a statistic similar in form to Student's t test, but one that uses all the data to estimate the underlying variability.

While the protected LSD test is a valid test that can be used for endocrine disruptor data (depending upon study objectives), the overall ANOVA should not include the positive control group or any group known a priori to produce a positive response. Otherwise, overall significance would be virtually guaranteed, and the benefits of a preliminary ANOVA would be lost. The result would likely be a test with an unacceptably high false positive rate. As a practical matter, even if done correctly, the protected LSD is generally "more liberal" (i.e., more prone to false positives, as discussed above) than certain alternative multiple comparisons

procedures.

Three other examples of multiple comparison procedures for making all possible pairwise comparisons are Duncan's Multiple Range Test, the Student-Neuwman-Keuls test, and Tukey's HSD test. These procedures are widely used and in our judgement are acceptable tests for the evaluation of endocrine disruptor data in those instances in which all possible pairwise comparisons are of interest.

There was also some possible confusion regarding the use of Jonckheere's test, a very useful nonparametric trend test. One investigator implied that it was a test for linear trend, but since it is a nonparametric test, it assumes no specific shape of the dose-response curve, and is simply a test for a monotonic (i.e., non-increasing or non-decreasing) trend. Another investigator used Jonckheere's test as the sole method of statistical analysis. The disadvantage of this approach is that non-monotonic dose-response trends (e.g., U-shaped or inverted U-shaped dose response curves with significant low dose effects) would probably not be detected by Jonckheere's test. The data that we evaluated confirmed this.

Although there are many advantages to tests that assume underlying normality (ANOVA, Dunnett's test etc.), an investigator must be aware of situations in which the data are extremely skewed, which would invalidate a normal theory-based approach. For example, one investigator assessed mammary gland differentiation (number of structures per squared mm of mammary gland) and reported a highly significant ($p<0.01$) effect of DES by Student' t test. What this investigator apparently failed to realize was that 39 of the 40 animals in the various dosed and control groups had a zero response. The one single positive response in the DES group was solely responsible for the apparent statistical significance. For such highly skewed data we prefer a nonparametric approach, which clearly (and correctly in our view) would find no statistical significance associated with a single positive response.

Some investigators used a Bonferonni correction to the p values when making pairwise comparisons, which in essence divides the nominal significance level alpha by the number of groups being compared. While there is nothing inherently "wrong" with such an approach, it is a rather conservative procedure, that will have a relatively high false negative rate. Such an approach should be unnecessary if an investigator uses one of the multiple comparison procedures discussed above or when the significance of an overall test is required to ensure that the proper experiment-wide error rate is maintained.

It is not our purpose to specify a methodology that must be used for the statistical analysis of endocrine disruptor data. Our main points are that if pairwise comparisons are of interest, then (i) a valid multiple comparisons test should be used; (ii) the choice of a specific multiple comparisons test should depend upon study objectives; (iii) if a parametric test is to be used, the investigator should evaluate whether the data are consistent with the underlying assumption of normality; and (iv) among the valid multiple comparison procedures, some are inherently more conservative than others. An investigator should maintain an awareness of these matters when choosing a method of statistical analysis.

B. Heterogeneity of the data - Virtually all of the procedures that assume an underlying normal

distribution (ANOVA, Dunnett's test, etc.) also assume homogeneity of variance, that is, within-group variability is constant across all groups. Occasionally, a simple transformation (e.g., a logarithmic transformation) can eliminate heterogeneity. Although the use of ANOVA-based procedures is not invalidated by modest variance heterogeneity, if the heterogeneity is extreme, even after a data transformation, then alternative nonparametric methods should be considered. Significant heterogeneity may also indicate the presence of an outlier in one or more of the test groups that must be dealt with.

Apparent failure to recognize heterogeneity was a common feature in many of the data sets we evaluated, and this occasionally led to unusual results. To take an extreme example, one investigator reported that by Dunn's test (a valid nonparametric alternative to Dunnett's test), the ovary/body weight ratio was significantly increased in a dosed group relative to controls. The author based this interpretation on the fact that the mean dosed group response exceeded the mean control response and that Dunn's test indicated statistical significance.

What this investigator failed to recognize was that the dosed group contained a single ovary weight that was approximately ten times the value of the group mean (we strongly suspect a decimal point error in this value, but the investigator was unable to confirm this). As a result, although the mean response was indeed slightly elevated in the dosed group relative to controls, the preponderance of the individual animal data showed the opposite trend, and Dunn's test (which was apparently carried out correctly) actually identified a statistically significant DECREASE, not a significant increase, in the organ/body weight ratio in the dosed group. Awareness of the heterogeneity problem may have prompted a more detailed examination of the individual animal data and avoided this problem.

As noted earlier in this report, good quality control of the data will often help eliminate heterogeneity. In the studies we evaluated, some investigators appeared to carry out a more careful data inspection than others. One of the data outliers not identified by the study investigators was noted above. As another example, in one study an observed seminal vesicle weight was reported to be 1.0195 (more than four times the testis weight of that animal), while the seminal vesicle weights of the other 79 animals in that particular dosed group ranged from 0.0063 to 0.05331. We strongly suspect that the 1.0195 value is a typographical error, with the correct value being 100-fold less: 0.010195. An appeal to the original lab book may be required to resolve such unusual observations. Even if the value is "real" then as a minimum it should be identified as a statistical outlier. The presence of such outliers in the data can have a dramatic impact on study results.

While many of the datasets we evaluated had significant variance heterogeneity, a simple log transformation was generally successful in eliminating this heterogeneity. Ignoring heterogeneity and carrying out ANOVA-based tests and pairwise comparisons can produce false negative outcomes, with the group showing the excessive variability unduly inflating the error term. However, false positive outcomes can also occur if the cause of the heterogeneity is improper randomization. Thus, we recommend that an investigator take appropriate measures to ensure that heterogeneity of variance is not a problem in his or her study. Possible measures include (i) a preliminary test for heterogeneity (e.g., Levene's test); (ii) a simple data transformation (e.g., a logarithmic transformation) to eliminate significant heterogeneity or

alternatively, a nonparametric analysis of the data; and (iii) a statistical procedure to identify potential outliers in the data.

C. Adjusting organ weight for body weight - There is one statistical issue that is especially relevant for endocrine disruptor data: How to take body weight differences into account when assessing possible changes in organ weight. The possible strategies include

(i) no adjustment - This may be satisfactory if organ weight is independent of body weight, but for many organ weights, this is not the case.

(ii) organ/body weight ratio - While this method of adjustment is commonly used, and may be reasonable in many instances, it makes a specific assumption about the relationship between organ and body weight, namely, that it is directly proportional. In other words, an animal twice the size of another should have organ weights twice the size as well. While this may seem logical, our experience indicates that this relationship does not hold for many organs, and that the organ/body weight ratio tends to "over-adjust" organ weight for the impact of body weight.

(iii) using body weight as a covariable - This is the approach that we prefer and have adopted in our re-analysis. This approach lets the data determine the linear relationship between organ weight and body weight and make the appropriate adjustment. We found that in some studies, organ and body weights were essentially independent, and thus no adjustment was needed. Rarely, the relationship between body and organ weight appeared to be directly proportional. Most commonly, there was a significant positive association, but the "adjustment" was less than that imposed by the organ to body weight ratio. In one study (unique in our experience), uterus weight actually showed a significant ($p<0.01$) negative correlation with body weight, even within the experimental groups. The biological significance of this negative correlation is unclear.

Using body weight as a covariable is not without potential difficulties, especially when the test chemical affects both organ and body weight. Technically, in an ANCOVA, the covariable should be independent of treatment. If the test chemical affects both body weight and organ weight, then it may be difficult to disentangle the effects of body weight and the effects of treatment on organ weight, as is discussed in more detail below.

Moreover, ANCOVA analyses generally assume that the effect of the covariate is the same for all dose levels. This is usually assessed with a test for the possible interaction between the covariate (body weight in our analysis) and the treatment/factor of interest - chemical dose. Significant interactions between body weight and treatment were observed in certain of our analyses. This implies that the effect of body weight was not the same for all dose levels of the test chemical. This clearly complicates the evaluation of experimental results.

To take a specific example of the possible problems associated with a study in which the test chemical affects both organ and body weight, consider Figures 1 and 2, taken from one of the datasets we evaluated. In this study all three animals in the top dose group had very low body weights and extremely low testis weights relative to the other animals. A regression line through the origin (consistent with the use of the organ/body weight ratio) resulted in the testis weights of all three high dose animals falling below the regression line, implying a reduced testis/body

weight ratio in this group. That was indeed the conclusion of the original investigator.

However, when the regression line is not forced through the origin, the best fitting regression line closely fits the high dose data (Figure 2). This suggests that body weight alone (without consideration of treatment) could have accounted for the reduced testis weights observed. That is, the high dose animals may have had testis weights that were normal for control animals of comparable size.

This example, and others like it, illustrate that when the test chemical reduces both organ weight and body weight, the relative impact of the test chemical and reduced body weight on organ weight may become confounded. If a chemical reduces organ weight by simply making an animal smaller, then this is inherently different from a study in which a chemical has a direct effect on organ weight. However, often these two outcomes are indistinguishable. An investigator must maintain an awareness of this when interpreting organ weight changes using an ANCOVA (or any other) approach.

Figure 3 illustrates another potential difficulty using ANCOVA. In this instance (ignoring the test chemical), there appears to be a negative association between uterus weight and body weight, with the heavier DES animals having a reduced uterus weight relative to controls. However, a closer examination of the data reveals that within the DES and control groups, there is no significant association between uterus weight and body weight. This suggests that the effects of DES are to increase body weight, and, independently, to decrease uterus weight. Thus, despite the apparent negative overall association between body weight and uterus weight in Figure 3, these two variables appear to be independent within the DES and control groups, so no adjustment for body weight may be needed.

When body weight is used as a covariable, an investigator must be aware of the influence of body weight on organ weight both before and after adjusting for the effect of the test chemical. This association may not be as straightforward as would initially appear. This example, like the previous one, illustrates the possible interpretational difficulties associated with a chemical that affects both body and organ weight.

D. Regression vs. ANOVA - Regression models relate responses to some function of dose, while ANOVA models essentially view the dose levels categorically. Thus, if there is an underlying pattern of dose-response, the regression models will be more sensitive to detecting these trends than a more omnibus ANOVA. For this reason, our analyses frequently examined the data for linear and quadratic trends. A quadratic pattern becomes especially important if non-monotonic dose-response patterns are possible when considering endocrine disrupting compounds.

E. Biological Interpretation - Not all endpoints exhibiting statistically significant treatment effects provide evidence of biologically important responses. When the treatment effect is real but the magnitude of response is small or the nature of change is not interpretable as an "adverse" response, biological interpretation may overrule statistical significance. However, for these particular studies, the Organizing Committee indicated to us that the various Subpanels were interested only in evaluating biological change, without passing judgement on whether or not

these responses are "adverse."

Some isolated treatment effects that cannot be replicated under identical conditions may be false positives, with their rate depending upon the chosen level of significance. As with true positives, interpretation of false positives is a matter for scientific judgement of the investigator and lies outside the realm of statistical analysis.

On the other hand, endpoints not exhibiting statistically significant treatment effects may nevertheless be affected by treatment in potentially important ways. When endpoints responding to treatment are mistakenly overlooked, it may compromise the value of a study. The extent to which this can occur is embodied in the concept of statistical power, which is discussed earlier in this report.

F. Data Selectivity - There are several valid reasons for discarding data in a given experimental setting. For example, there may have been a technical problem in the execution of the study and/or the measurement of the variables of interest that rendered the data of little or no scientific value. Alternatively, there may have been a statistical outlier in a given group that was discarded. Another example may be a study in which the positive control does not produce the expected positive response. The prudent course of action in such cases may be to declare the study inadequate and repeat it, regardless of the experimental outcome in the test groups.

However, data should not be discarded simply because the test groups did not produce the expected (or desired) response. Similarly, if several replicates are used, it is not appropriate to report only the one(s) producing the strongest (or weakest) response. The scientific principal is that the data evaluation and reporting of study outcomes should be evenhanded, and that there should be no selective reporting of experimental results. One of the 23 questions dealt with this matter. No investigators reported data selectivity that would have materially affected their conclusions.

III. Investigator responses to the 23 Questions (Appendix B)

Answers to 23 questions regarding study design and study conditions were requested for 58 studies. Responses were received for 42 studies. Since many studies were done in the same laboratories, these responses cannot be viewed as a representative sample of some larger population of laboratories. Nevertheless, a summary of some features of these studies does provide a useful indication of the variety of methods, materials, and procedures used to generate the data under review.

Although studies of both rats and mice are represented, more studies used rats than mice and some studies used both species. Of the 42 studies reporting, 24 used Sprague Dawley (CD) rats, 5 used Alderley Park rats, and the remainder used Fischer 344, Noble, Han-Wistar, or special in-house strains. Among 18 studies using mice, 8 used CD-1 mice, 5 used CF-1 mice, and the remainder used a variety of different strains.

Individual housing for breeding pairs and dams with litters was common to all studies. Housing for weanlings, however, was split, with 28 studies employing multiple housing, usually with

littermates separated by sex caged together, and 14 studies reporting individual housing.

The variety of experimental objectives is reflected in a mixture of protocols.
Some studies culled litters while others did not, some studies used littermates while others randomly selected a single pup per litter, and some studies included more than one replicate of dose groups, but most did not. Although nearly all studies used a 12 hour light-dark regime, bedding and caging varied among laboratories in no easily characterized way. One aspect of protocol that differed among studies was whether actual dose formulations employed were tested for concentration of experimental substance. Of the 42 respondents, 17 included some measurement of dose formulations and 25 did not. Nearly all studies had concurrent control groups and many employed some form of blinding with respect to treatment during necropsy or other data collection.

Although the limited time frame did not permit an in depth assessment of the 23 questions, the Statistics Subpanel believes that this useful information (given in Appendix B) should be given careful consideration by each of the Subpanels.

FINAL COMMENTS

This report intentionally did not have any "bottom line conclusion" regarding the presence or absence of low dose endocrine disruptor effects. As noted earlier, that was not our objective. We are hopeful that the various Subpanels can examine this report, the data evaluations in Appendix A, the answers to the 23 questions for each study given in Appendix B, and other relevant biological information and reach their own conclusions on whether or not endocrine disruptors are causing low dose effects. For the interested readers, the data used in these analyses are being made available separately from this report.

ACKNOWLEDGEMENT: The Statistics Subpanel is grateful to Kevin McGowan of Analytical Sciences Inc. for the long hours he spent preparing the data sets we received in a uniform format that made our statistical analysis much easier. As noted previously, we are also grateful to the various investigators for their willingness to submit their raw data to us for re-evaluation.

Figure 1 - Relationship between testis weight and body weight for a selected dataset (ignoring treatment group). Line is the best fitting regression line, restricted to passing through the origin, which is consistent with the use of testis/body weight ratios. Since the three high dose data points fall below the line, this is the consistent with the authors' conclusion that the testis/body weight ratio is significantly reduced in the high dose group relative to controls.

Figure 2 - Relationship between testis weight and body weight for the same selected dataset (ignoring treatment group). Line is the best fitting unrestricted regression line, which fits the high dose data points very well. This analysis suggests that the lower testis weights in the three high dose animals could very easily be the weights expected in control animals of equivalent size.

Figure 3 - Relationship between uterus weight and body weight for a selected dataset. Ignoring dose, there appears to be a negative correlation between uterus weight and body weight. However within both the DES and control groups, there is no significant correlation.

Figure 1
Linear Regression Through Origin

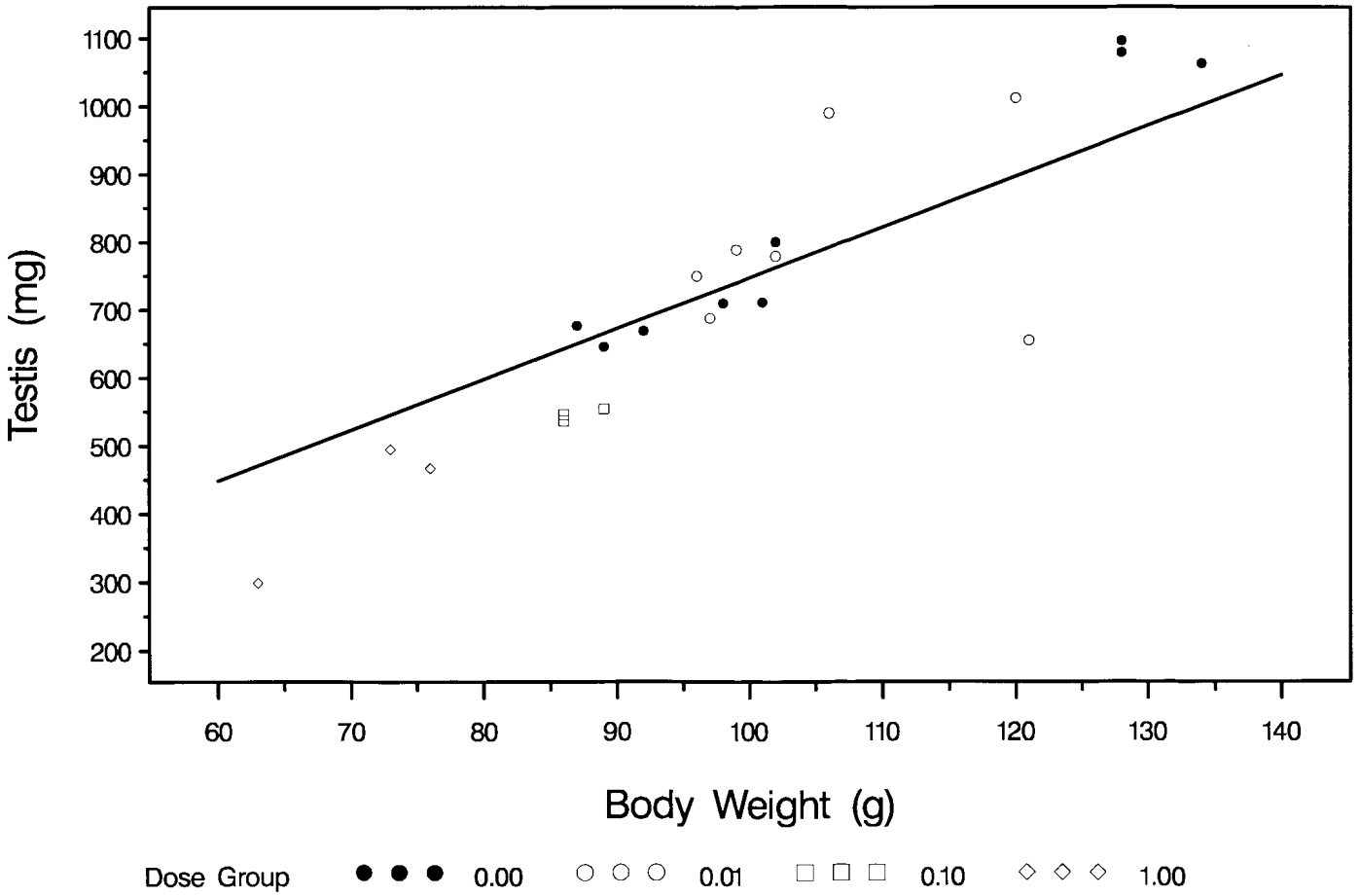


Figure 2
Linear Regression

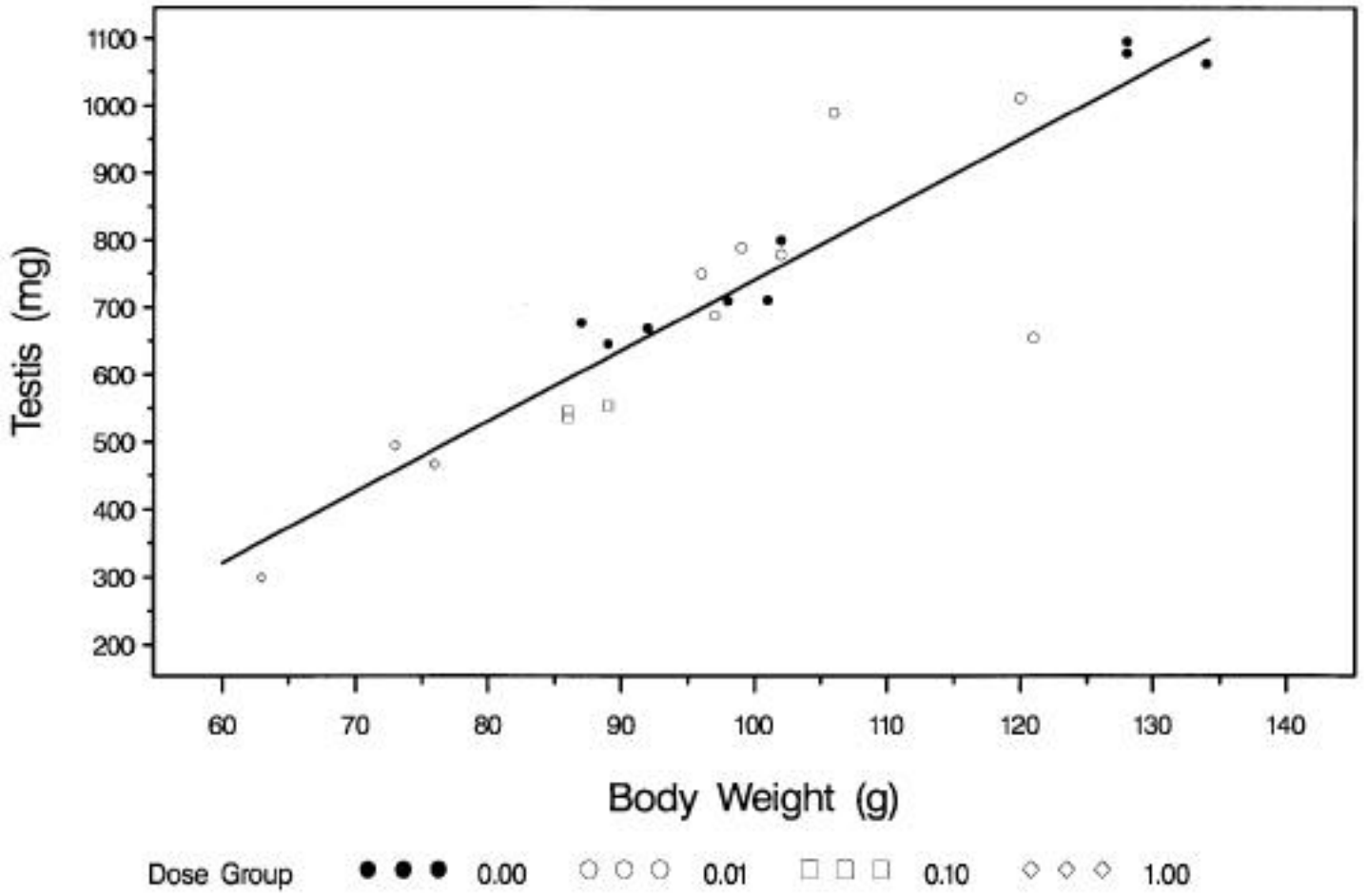
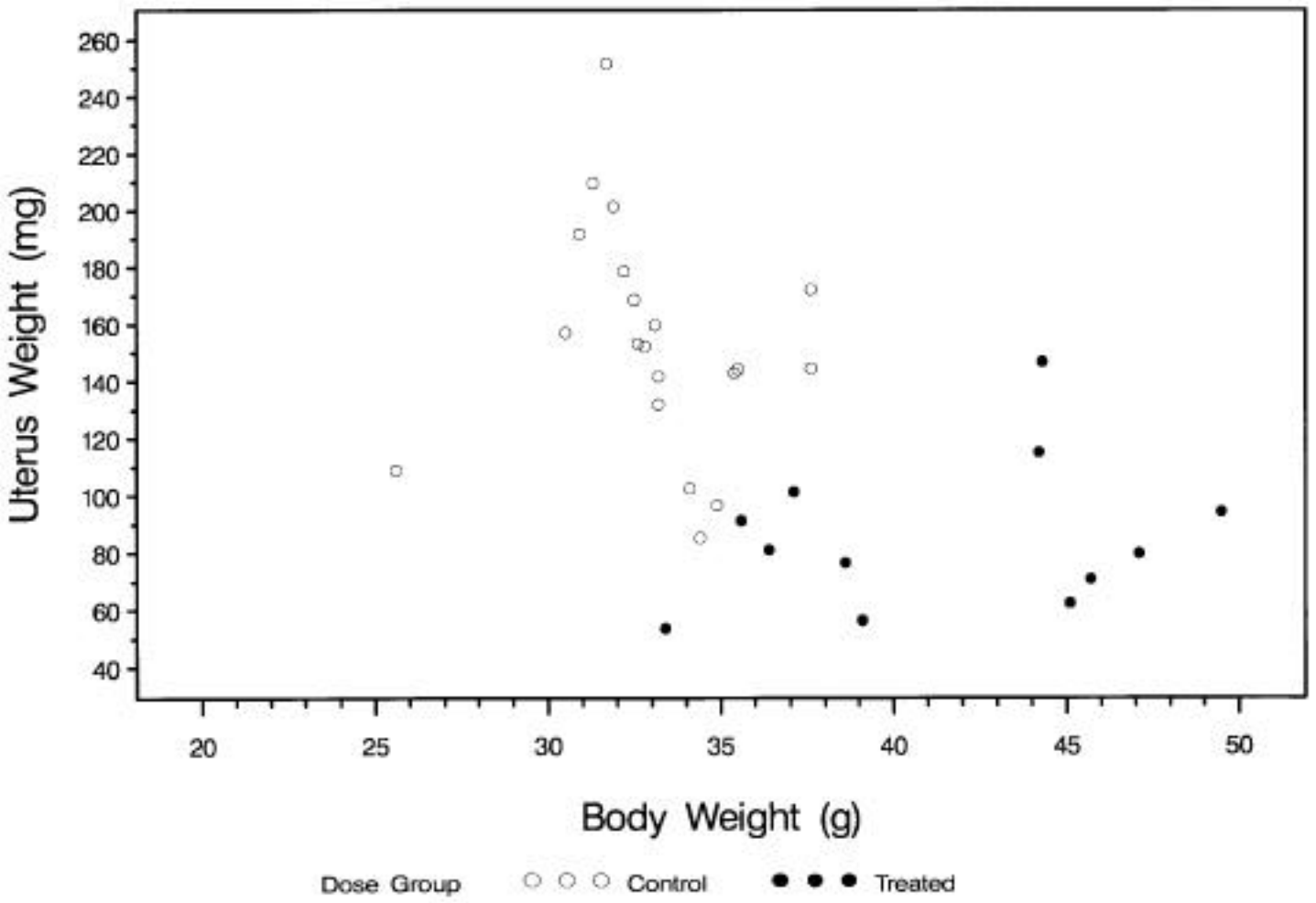


Figure 3


Uterus Weight (mg)
260
240
220
200
180
160
140
120
100
80
60
40
20
25
30
35
40
45
50
Body Weight (g)
Dose Group
Control
Treated

Chapter 6:
Report of the Dose-Response Modeling Subpanel

Chair:
Michael Kohn, National Institute of Environmental Health Sciences

Panelists:
Hugh Barton, U.S. Environmental Protection Agency
Jim Cogliano, U.S. Environmental Protection Agency
Rory Conolly, Chemical Industry Institute of Toxicology
Robert Delongchamp, National Center for Toxicological Research

Dose–response profiles can exhibit a variety of curve shapes, including rectangular hyperbolic, sigmoidal (*e.g.* a Hill exponent greater than 1), and U-shaped (either concave upwards or convex upwards). A U-shaped dose–response curve can reflect either stimulatory or inhibitory effects on the biological end point, depending on the mechanistic role of the plotted variable. Models of different biological systems support linear or non-linear responses at low doses for receptor-mediated processes, depending on how the model represents regulation of those processes. The shape at low doses similar to those expected to be achieved from environmental exposures can have important consequences for estimates of the risk of adverse health effects due to such exposure.

This sub-panel explored a number of mechanisms to identify factors that influence the curve shape. Although this work does not examine all possible mechanisms, it does demonstrate that the choice of mechanism and the numerical values of the constants (parameters) in the resulting equations have a major effect on the predicted dose–response These insights can guide the design of experiments to address unresolved issues and can help validate assumptions made in the risk assessment process.

Michael Kohn, NIEHS

A theoretical model of competition between a weakly estrogenic xenobiotic compound and endogenous estrogen was constructed in the SCoP language. The model presumes that the compound achieves a steady state in the target tissue, binds to the estrogen receptor, and that the effect is proportional to the fraction of DNA binding sites occupied by the liganded receptor. The variables and initial conditions in this model are: R = estrogen receptor, R(0) = 1; E_2 = estradiol, $E_2(0) = 0.1$; X = xenobiotic agent; ERE = estrogen responsive element, ERE(0) = 1.

$$R + E_2 \rightleftharpoons R{\bullet}E_2 \qquad K_E = 0.03$$

$$ERE + R{\bullet}E_2 \rightleftharpoons ERE{\bullet}R{\bullet}E_2 \qquad K_{RE} = 0.02$$

$$R + X \rightleftharpoons R{\bullet}X \qquad K_X = 10$$

$$ERE + R{\bullet}X \rightleftharpoons ERE{\bullet}R{\bullet}X \qquad K_{RX} = 5$$

The above chemical equations were converted into differential equations using the law of mass action, and the differential equations were integrated from *time* = 0 to *time* = 10 to evaluate the steady state corresponding to the imposed constant concentration of X. The effect ($ERE{\bullet}R{\bullet}E_2$ + $ERE{\bullet}R{\bullet}X$) was calculated for various concentrations of the xenobiotic agent (Fig. 1).

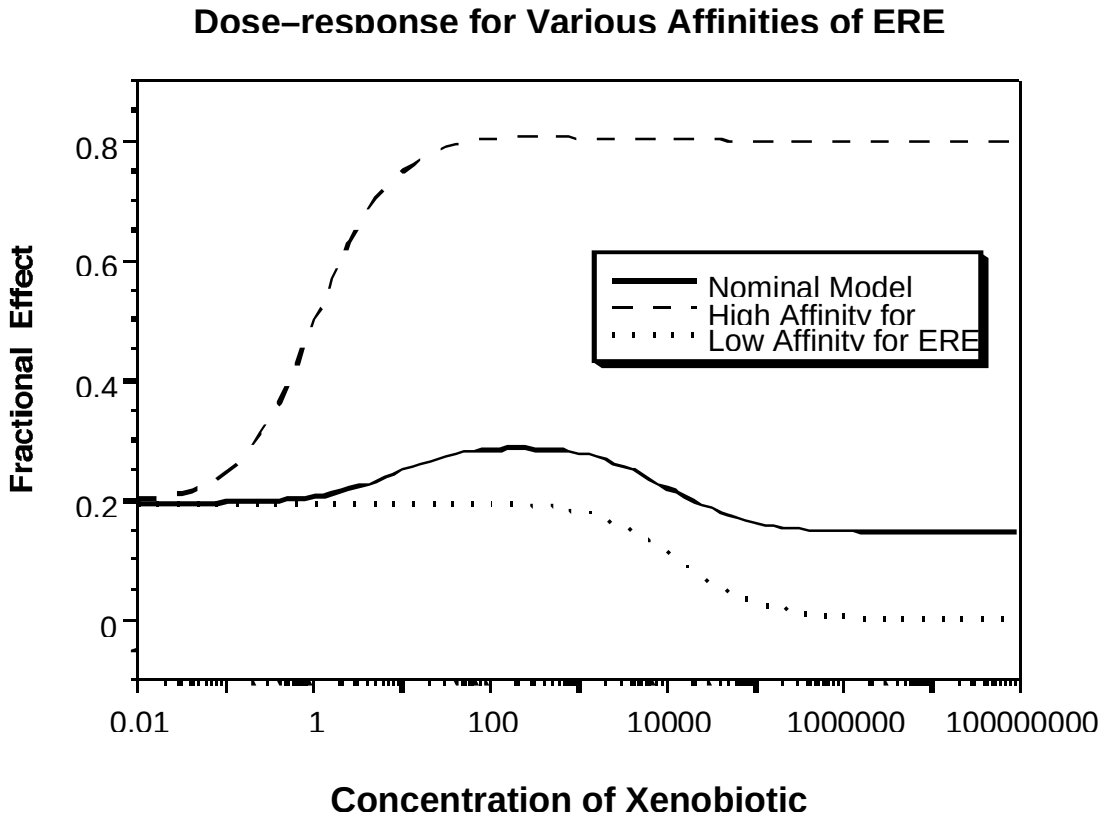


Fig. 1. Dose–response profiles for estrogen receptor-mediated effects.

The model produced an inverted U-shaped dose–response curve. This behavior results from the assumption that there is an excess of receptors so that the endogenous ligand E_2 is not saturating. At modest concentrations of X more receptors are liganded and the effect is increased. At high concentrations of X the xenobiotic agent out-competes E_2 for R and, because $K_{RX} >> K_{RE}$, the effect is reduced. Raising the affinity of the ERE for R•X ten-fold ($K_{RX} = 0.5$) caused X to behave like a pure agonist. Lowering the affinity ten-fold ($K_{RX} = 50$) resulted in pure antagonistic behavior. Similar changes to K_X did not alter the qualitative behavior of the model.

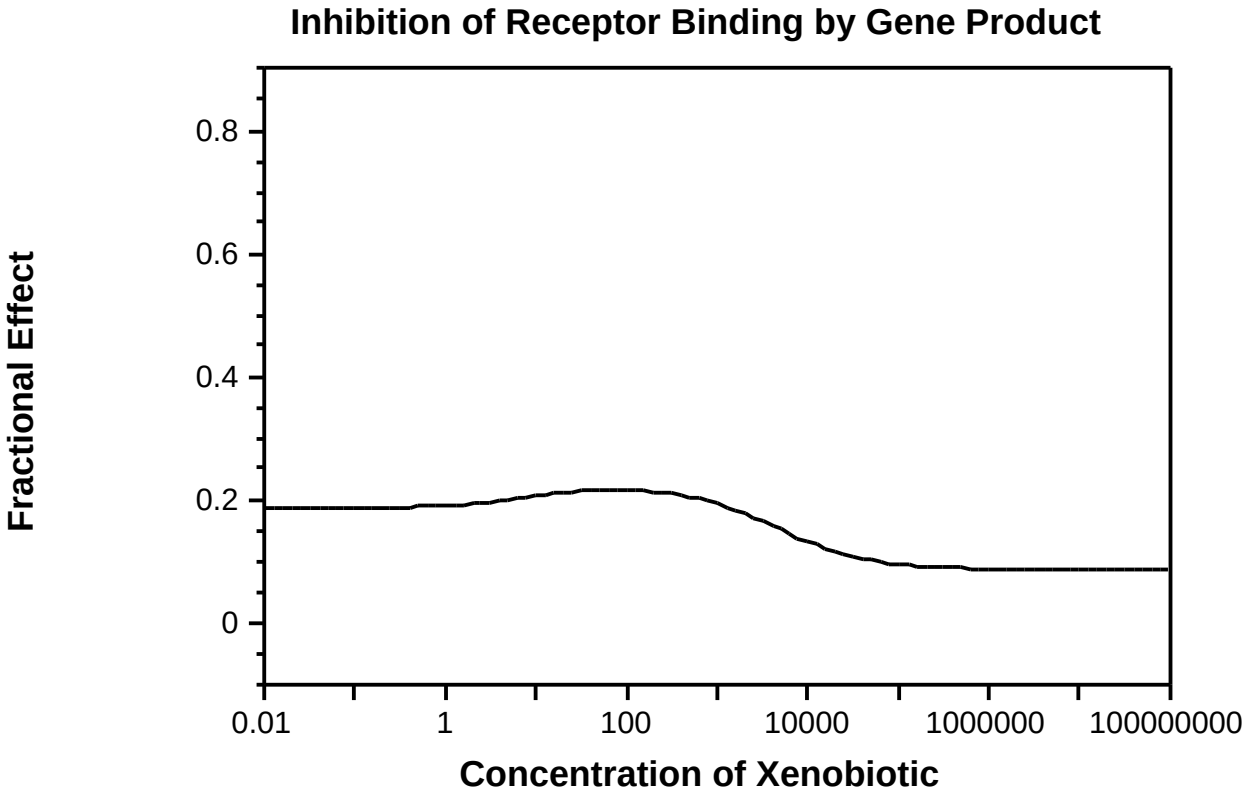


Fig. 2. Damping of the dose–response by feedback inhibition of receptor binding.

Equations for the estrogen receptor-mediated production of a gene product (P) were added to the model of an endocrine active compound. Both E_2 and X were assumed to be effective inducers of P. Linear kinetics were assumed for synthesis and degradation of the gene product (equation 1).

$$dP/dt = k_{synthesis} \times (ERE \bullet R \bullet E_2 + ERE \bullet R \bullet X) - k_{degrade} \times P \quad (1)$$

$$k_{synthesis} = 5$$

$$k_{degrade} = 1$$

Also, the product P was assumed to interfere with binding of the liganded receptor ($R \bullet E_2$ or $R \bullet X$) to response elements. The effective binding constant for the liganded receptor is

$$K_{app} = K_R \times (1 + P / K_i) \quad (2)$$

where K_R is either K_{RE} or K_{RX} and K_i is the dissociation constant of the protein-DNA complex. With $K_i = 0.5$ and all other parameters set to their nominal values as given above, inhibition of

binding of liganded receptor according to equation 2 significantly damps the effect of the xenobiotic agent (Fig. 2).

To demonstrate the invariance of the qualitative features of this model, different values for the initial conditions were used. For example, the concentration of the ERE was reduced from 1.0 to 0.2 while maintaining a sub-saturating level E_2. Parameter values could still be found that predicted the xenobiotic agent to be either a pure agonist, a pure antagonist, or producing an inverted U-shaped dose–response, indicating that this result depends solely on the existence of spare receptors. The robustness of this prediction illustrates the plausibility of non-linear response profiles, in particular, the inverted U shape, arising from receptor-mediated processes.

It is likely that only some of the EREs mediate the specific responses induced by the xenobiotic compound. Another variant of this model was constructed; there were two classes of ERE which differed in their affinities for liganded receptor. It was assumed that there is four times as much of the lower affinity response elements as the higher affinity response elements. Regardless of which class of ERE (or both) was treated as mediating the response the same behavior was predicted as for the nominal model. That is, it was possible to find parameter values for which the xenobiotic acts as a pure agonist, a pure antagonist, or a “mixed agonist” that produces an inverted U-shaped response. These results indicate that quantitative aspects of the mechanistic description are crucial to reliably predict responses to exposures to a toxicant.

Rory Conolly, CIIT

The above SCoP model was reproduced in MATLAB and gave the same results (Fig. 3). The inverted U-shaped dose–response curve arises when the binding capacity of the ERE is sufficiently large that the endogenous ligand (E_2), when complexed with its receptor ($R•E_2$), can’t fill up the ERE (*i.e.*, the amount of ERE is greater than the amount of $R•E_2$). The inverted U behavior in Figure 3 arises when ERE and R have values of 1 and E_2 has a value of 0.2. When E_2 is set to 1.0 the inverted U behavior disappears.

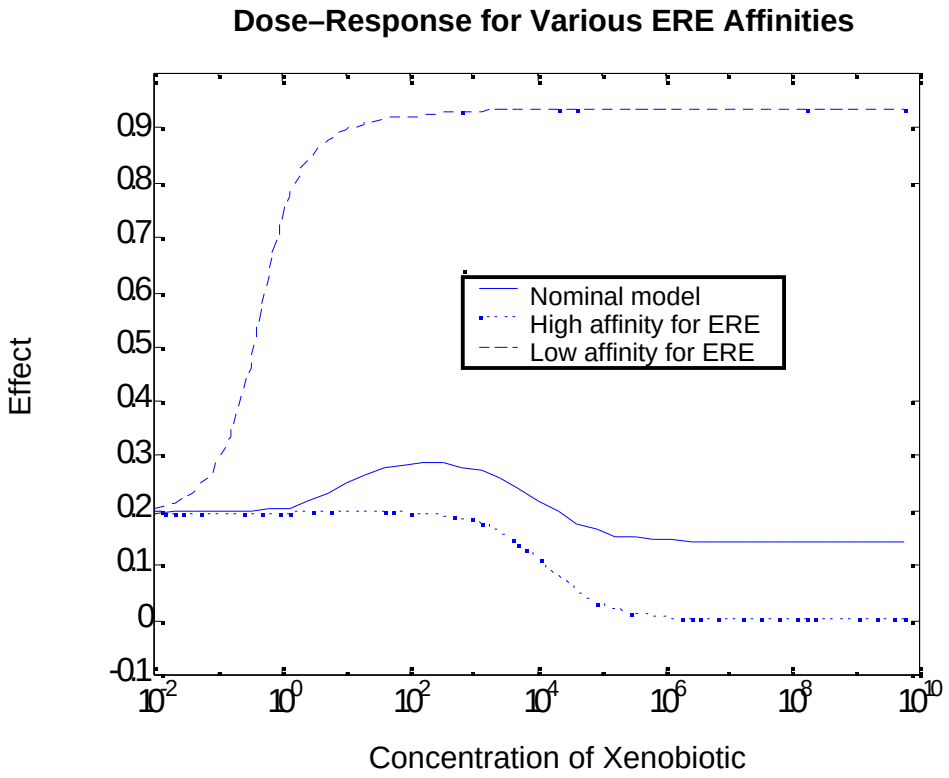


Fig. 3 . MATLAB reproduction of Kohn's SCoP model.

The nominal model assumes that the xenobiotic is supplied at a rate sufficient to maintain a steady state in the target tissue. Saturable metabolism of the xenobiotic was added to the model such that a bolus dose of X is gradually depleted over the 10-unit time interval. The most dramatic effect (Fig. 4) occurs for the case where the xenobiotic has high affinity for the ERE, so that it acts like a full agonist although the response predicted by the nominal model is also steeper. Comparison of the high affinity plots in Figs. 3 and 4 shows that including metabolic clearance can greatly increase the sigmoidicity of the dose–response curve and demonstrates the importance of pharmacokinetics to the response produced.

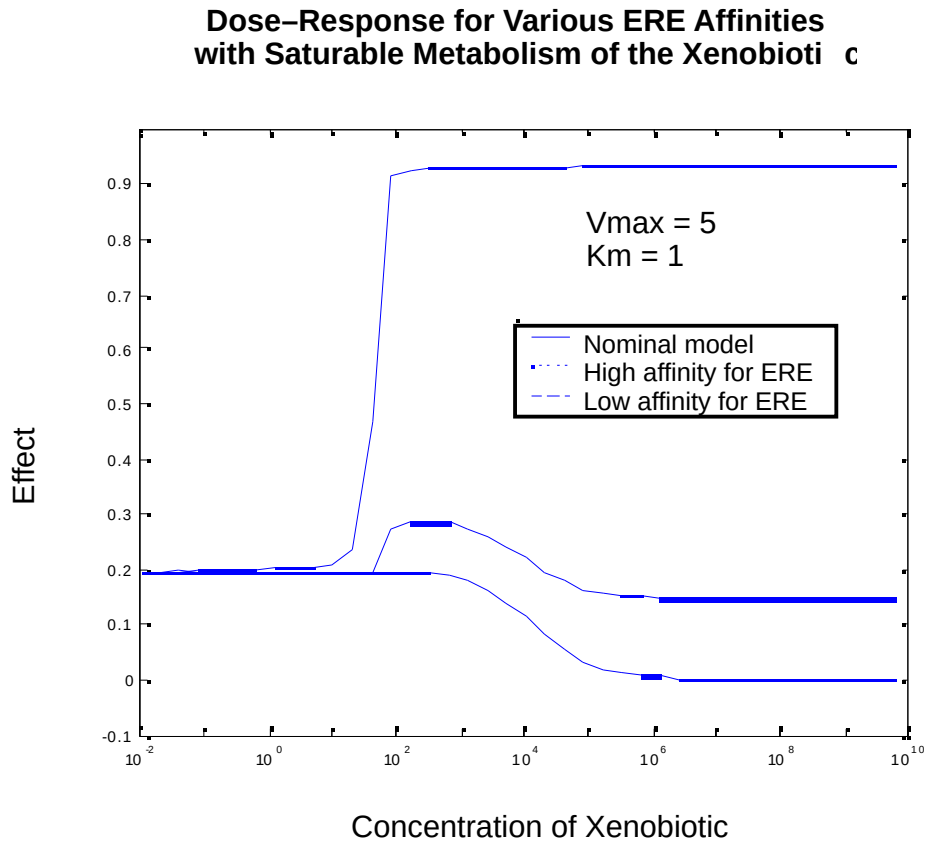


Fig. 4. Dose–response behaviors with saturable metabolism of the xenobiotic.

Hugh Barton, US EPA

Models have been published that illustrate how biological regulatory processes can create different dose–response behaviors. Two endocrine-related ones are briefly summarized here. In addition, a recent review paper on modeling of gene networks is briefly described. This paper suggests that other biological literature outside toxicology must be considered for evaluating dose–response of endocrine active compounds.

1. Male Central Axis Modeling

A simplified model of the male central axis was constructed with production of testosterone in the testes, the unusual blood flow shunt in the testicular blood supply, negative feedback of testosterone on lutein (LH) production, positive feedback of LH on testicular testosterone production, and clearance of testosterone and LH from the peripheral blood (Barton and Andersen, 1998). The model was parameterized for a 70-day-old intact male rat (300 g). It

simulates reasonable physiological steady state values of testosterone in the interstitial fluid, the spermatic cord venous blood, and the peripheral blood and of LH in peripheral blood.

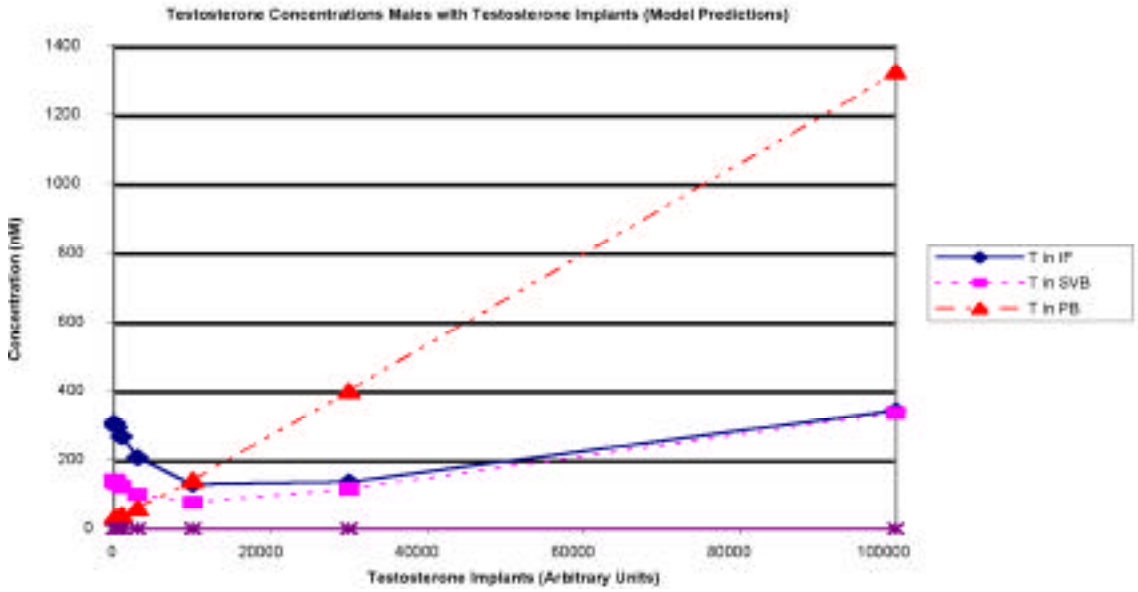


Fig. 5: Simulated testosterone concentrations in interstitial fluid (IF), spermatic cord venous blood (SVB), and peripheralblood (PB) of intact rats.

Silastic implants with testosterone are often used experimentally to dose continually for an extended period. Implants were modeled as dosing continuously directly into the peripheral blood. A dose–response for implants was run for nine doses between 10 and 100,000 (arbitrary units). The resulting concentrations of testosterone and LH are plotted in Figures 5 and 6. A U-shaped dose–response occurs for interstitial fluid (IF) and, to a lesser degree, for spermatic cord venous blood (SVB). Though it is not observable in the plot, peripheral blood (PB) concentrations increase very slightly for 10–300 implants due to the homeostatic negative feedback.

As the implants raise testosterone in peripheral blood and shut down LH production (Fig. 6),

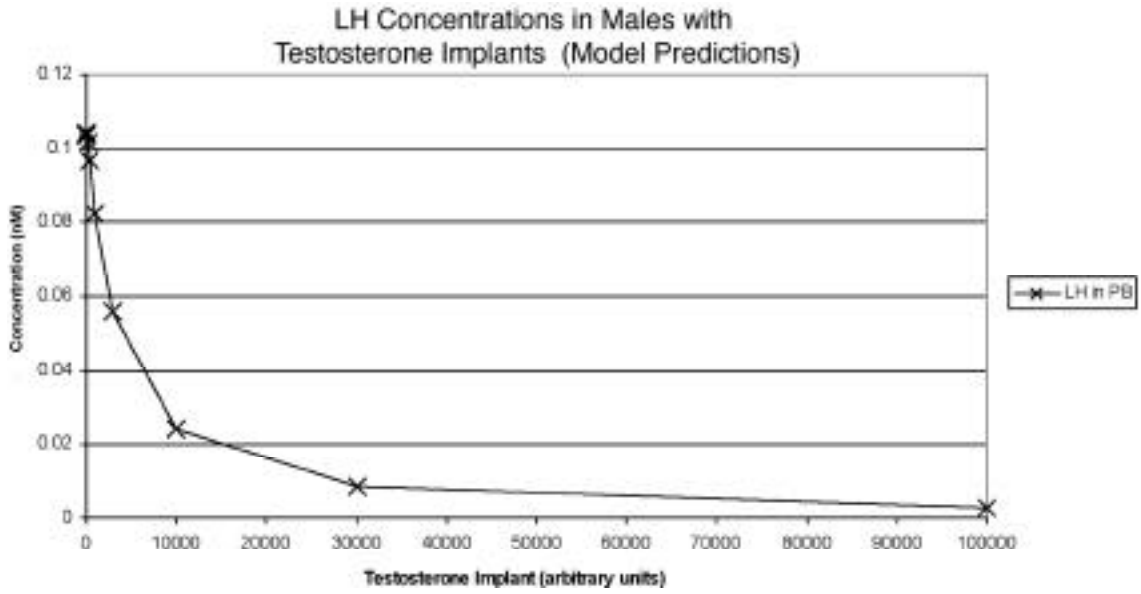


Fig. 6: Simulated LH Concentration in Peripheral Blood (PB) of Intact Rats

peripheral testosterone concentrations increase proportionately with the implants. At high concentrations this can restore testicular testosterone to normal levels, but now the peripheral blood concentrations are extremely high compared to normal. The large gradient in testosterone concentrations is created by rapid synthesis in the testes and the unique shunt blood flow to that organ.

This process is mimicked by decreasing and then increasing spermatogenesis observed in studies in Ewing's laboratory at John's Hopkins. The system creates this behavior because the brain/pituitary production of LH responds to peripheral testosterone concentrations (approximately 2 ng/ml or 8 nM), while the testes respond to interstitial fluid concentrations of approximately 90 ng/ml or 307 nM. Note that the androgen receptor K_d is roughly 1 nM, so testicular receptors are fully occupied and receptor concentration would be limiting whereas testosterone concentration would be limiting in the brain.

Absent the feedback relationship in the brain, the slight increase in circulating testosterone from silastic implants would increase testicular concentrations too slightly to have much effect on testicular functions such as spermatogenesis. Given that there is feedback, adding low

exogenous testosterone decreased LH and consequently testicular testosterone synthesis, so the direct effects in the brain are the critical factors. Thus, increasing peripheral blood concentrations alters brain responses at levels that would have no direct impact on testicular function.

2. Positive-Feedback Regulation of Receptor and Ligand Synthesizing Enzymes

Theoretical models that capture the essential behavior of biological regulatory systems in a wide range of species were developed to explore the role of positive feedback on levels of receptor and on synthesis of high-affinity ligands for those receptors (Andersen and Barton, 1998). These models display rectangular hyperbolic behavior absent feedback and increasingly non-linear monotonic behaviors due to the operation of the positive feedback, thus essentially switching the system from one state to another.

These models can also show hysteresis, *i.e*. the dose–response curves can differ depending upon the prior exposures. For example, if ligand exposures increase, inducing the receptor, when ligand levels subsequently drop to their original values, the increased receptor levels (the receptor is only slowly degraded) will result in higher ligand–receptor complex levels and thus greater response. The hysteresis will be observable in areas of the dose response curve where receptor concentration is limiting. A similar effect would be anticipated if you had down-regulation of the receptor (*e.g*. starting with a high dose that down-regulated a receptor, greatly reducing the dose, and then returning to higher doses could give different dose–responses on the way down and up). Thus, the response with dropping ligand concentrations can be shifted compared to response with increasing concentrations.

3. Modeling Gene Networks

Owing to the recent availability of genomic information and new experimental and analytical genomic methodologies, there is currently much modeling of genetic networks. Modeling efforts have included Boolean logic, continuous differential equations, and hybrid descriptions (Smolen et al. 1999). These modeling efforts have reported multiple stable states, oscillating systems, and other behaviors. The convergence of these efforts and the interest in chemically induced perturbations of biological regulatory systems could be very useful to the field of toxicology and

dose–response analysis. Efforts to bring these research areas together could be very beneficial.

References

Andersen ME and Barton HA (1999) Biological regulation of receptor-hormone complex concentrations in relation to dose-response assessments for endocrine-active compounds. Toxicol Sci 48, 38-50.

Barton HA and Andersen ME (1998) A model for pharmacokinetics and physiological feedback among hormones of the testicular-pituitary axis in adult male rats: a framework for evaluating effects of endocrine active compounds. Toxicol Sci 45, 174-187.

Smolen P, Baxter DA, and Byrne JH (1999) Modeling transcriptional control in gene networks – methods, recent results, and future directions. Bulletin of Mathematical Biology 62, 247-292.

Robert DeLongchamps, NCTR

NCTR measured the volume of the sexually dimorphic nucleus (SDN, a small area in the brain) as part of the NTP studies of endocrine disrupters. Studies of nonylphenol, genistein, ethinylestradiol and vinclozolin have been completed. The first three are estrogenic, and vinclozolin is an anti-androgen. The estrogenic compounds affected the SDN volume of male rats.

Suppose that several chemicals induce a response through a common physiological pathway such that the only difference in the respective dose–responses is attributable to the relative potency of the chemicals, ρ_c . That is, the dose–response for this set of chemicals, C , can be written as $R(d \mid c) = R(\rho_c d)$ for any $c \in C$. This type of relationship appears to describe the dose response of SDN volume in male rats exposed to three estrogenic chemicals, ethinylestradiol, genistein, and nonylphenol.

The model, which was fit to the data (Fig. 6), is composed of two Michaelis–Menten-like relationships. Suppose that the response for a reference chemical follows the Michaelis–Menten relationship, then

$$R(d) = \frac{V\ d}{k + d} \tag{3}$$

Likewise, if a response decreased in proportion to the binding implied by equation (3), then the

response would follow the relationship,

$$R(d) \quad V - \frac{V\ d}{k + d} = \frac{kV}{k + d} \tag{4}$$

The scatter plot of the observations suggest that two components describe the data, *i.e.*,

$$R(d|c) = \frac{k_1 V_1}{k_1 + \rho_c} + \frac{V_2 \rho_c d}{k_2 + \rho_c d} \tag{5}$$

The optimal parameter values and statistics of the fit to data obtained using equation 5 are:

$$k_1 = 23.95, V_1 = 78.4$$
$$k_2 = 630.8, V_2 = 86.7$$

SOURCE	DF	SS	MS	F	P
Model	5	27746.1	5549.22	8.04	0.00001
lof	15	14975.5	998.37	1.45	0.15527
error	61	42096.3	690.10		

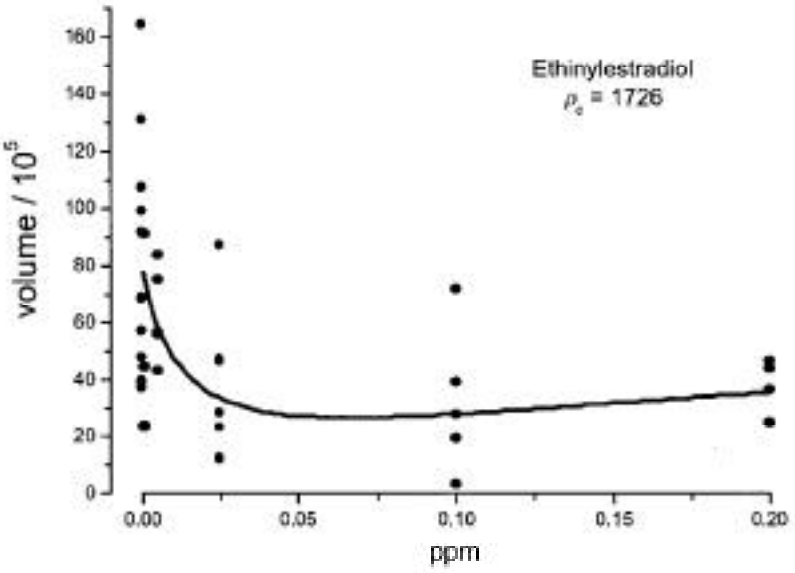


Fig. 6A. Effect of ethynlestradiol on SDN volume.

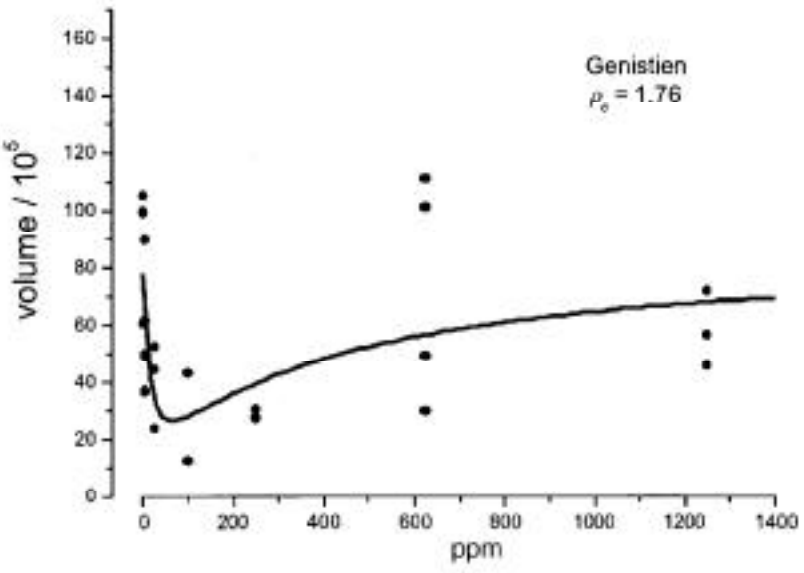


Fig. 6B. Effect of genistein on SDN volume.

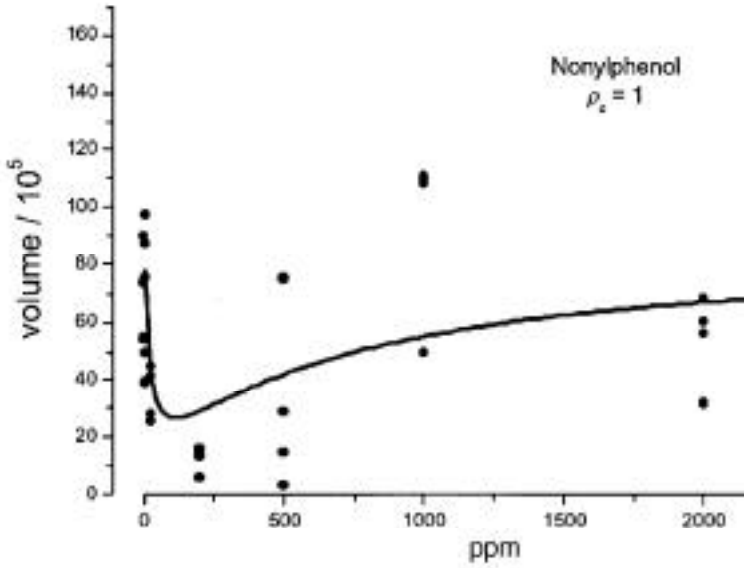


Fig. 6C. Effect of nonylphenol on SDN volume.

Figures 6A–C illustrate the fit of equation 5 to the dose–response data for three estrogenic compounds. Nonylphenol was used as the reference chemical. A dose of 1 ppm of ethinylestradiol has a relative potency that is 1726 times greater than 1 ppm of nonylphenol, and

ppm of genistein has a relative potency that is 1.76 times greater than 1 ppm of nonylphenol. The model gives a significant reduction in the residual ($p = 0.00001$) relative to a constant mean, and it does not have a significant lack of fit ($p = 0.15$) relative to the pooled within-dose chemical variation.

James Cogliano, US EPA

Mechanism-based models hold the promise to improve the credibility of a risk assessment's predictions at environmental doses. The use of mechanistic data in risk assessment is the major focus of EPA's proposed cancer guidelines *(EPA 1996, 1999)*, and these guidelines can be viewed as a framework for harmonizing future assessments for cancer and other effects. These guidelines discuss how understanding of mechanism follows from identification of *key events*, which are empirically observable precursor steps that are either a necessary part of the mechanism or a marker for one.

The most credible mechanism-based models will, therefore, have parameters that reflect the key events of the mechanism, and these parameters will be directly estimated from experimental data. (This is possible because key events are *observable* steps in the mechanism.) Without direct estimation of parameters from experimental data, the model would be only another curve-fitted model whose parameters have a possible mechanistic explanation, but there would be little confidence that the model was appropriate.

Confidence in the model is a key requirement in models that are used to support public health decisions, as public health officials ask for risk assessments that help them assure protection of susceptible populations. This illustrates how the application of a model can influence the level of confidence that is required. For example, models with little experimental support can be appropriate to describe a hypothesized mechanism and provide a structure for collecting data and testing the hypothesis. Such models, however, would be premature for use in risk assessment, as public health decisions require more assurance that the model would not understate health risks to susceptible populations.

While much has been discussed about the role of mechanistic data in risk assessment to

assess the relevance of laboratory animal results to human environmental exposures, mechanistic data can also be used in models to:

- Provide insight into the likely shape of the dose–response curve at low doses,
- Quantify the relative sensitivity of laboratory animals and human populations,
- Estimate differential risks to sensitive populations.

This last point leads to the importance of human variation as it affects the shape of the dose–response curve. Human variation is a determinant of the shape of the dose–response curve for a population, and it can differ from the dose–response curve for an individual. In a human population, genetic and lifestyle factors contribute to variation in sensitivity that spreads the dose–response curve over a wider range *(Lutz 1990)*. Consequently, data on the variation of key parameters across the human population is needed for confidence that a mechanism-based model reflects the potential for risk to different populations.

References

Lutz WK (1990) Dose-response relationship and low dose extrapolation in chemical carcinogenesis. Carcinogenesis 11(8): 1243–1247.

U.S. Environmental Protection Agency (1996) Proposed guidelines for carcinogen risk assessment; notice. Federal Register 61(79): 17960–18011.

U.S. Environmental Protection Agency (1999) Guidelines for carcinogen risk assessment. Washington: U.S. EPA, NCEA-F-0644, review draft.

Appendix A:
Detailed Evaluations of Individual Studies
Index

ASHBY DATASETS

DELCLOS DATASETS

A-25]

3. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep (Unpublished Abstract). **[JB] [PAGE A-27]**

4. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of genistein in male and female Sprague Dawley rats." (Unpublished Final Report). **[JH] [PAGE A-28]**

5. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of methoxychlor in male and female Sprague Dawley rats." (Unpublished Final Report). **[JH] [PAGE A-30]**

6. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of nonylphenol in male and female Sprague Dawley rats." (Unpublished Final Report). **[JH] [PAGE A-31]**

7. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats." In prep (Unpublished Abstract). **[JB] [PAGE A-33]**

8. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression." In prep (Unpublished Abstract). **[JB] [PAGE A-34]**

9. Meredith, J. M., C. Bennett, et al. (2000). "Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats." Society for Neuroscience Abstracts (in press). 10. Scallet, A. C., C. Bennett, et al. (1999). "Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure." Society for Neuroscience Abstracts 25: 227. **[KP] [PAGE A-36]**

O'CONNOR DATASETS

1. Biegel, L. B., J. C. Cook, et al. (1998). "Effects of 17β-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats." Toxicological Sciences 44: 143-154. **[JH] [PAGE A-37]**

2. Biegel, L. B., J. A. Flaws, et al. (1998). "90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol." Toxicological Sciences 44: 116-142. **[JH] [PAGE A-40]**

3. Cook, J. C., L. Johnson, et al. (1998). "Effects of dietary 17β-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats." Toxicological Sciences 44: 155-168. **[JH] [PAGE A-43]**

4. O'Connor, J. C., S. R. Frame, et al. (1998). "Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17β-estradiol." Toxicological Sciences 44: 169-184. **[JH] [PAGE A-44]**

VOM SAAL DATASETS

1. Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October. **[JH] [PAGE A-47]**

4. Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives 105(1): 70-76. **[JH] [PAGE A-50]**

6. Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." (in press). **[RK] [PAGE A-51]**

9. vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health 14 (1/2): 239-260. **[JH] [PAGE A-53]**

10. vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November. **[JH] [PAGE A-55]**

11. Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health 15: 12-25. **[JH] [PAGE A-56]**

Vom Saal datasets requested but not provided [PAGE A-57]

2. Howdeshell, K. L., A. K. Hotchkiss, et al. (1999). "Exposure to bisphenol A advances puberty." Nature 401: 763-764.

3. Howdeshell, K. L. and F. S. vom Saal (2000). "Developmental exposure to bisphenol A: interaction with endogenous estradiol during pregnancy in mice." American Zoologist 40(3). (in press).

5. Palanza, P., S. Parmigiani, et al. (1999). "Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and o,p'-DDT alters aggressive behavior of male and female house mice." Pharmacology Biochemistry and Behavior 64(4): 665-672.

7. Timms, B. G., S. L. Petersen, et al. (1999). "Prostate gland growth during development is stimulated in both male and female rat fetuses by intrauterine proximity to female fetuses." Journal of Urology 161: 1694-1701.

8. vom Saal, F. S., B. G. Timms, et al. (1997). "Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses."

Proceedings of the National Academy of Sciences 94: 2056-2061.

CHAHOUD STUDY

1. Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." (3 Abstracts). **[KP, JH] [PAGE A-57]**

EMA STUDY

1. Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." (Unpublished Study Report). **[JH] [PAGE A-59]**

GRAY STUDY

1. Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health 15: 48-64. **[JH] [PAGE A-62]**

LEE STUDY

1. Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine 9(1): 105-111. **[JH] [PAGE A-67]**

SPEAROW DATASETS

1. Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science 285: 1259-1261. 2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August. **[RK] [PAGE A-68]**

2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August. **[RK] [PAGE A-71]**

TYL DATASETS

1. Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology 30: 81-95. **[JH] [PAGE A-73]**

2. Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 (Draft Final Report). **[JH] [PAGE A-77]**

WAECHTER DATASETS

1. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences 50: 36-44. **[JH] [PAGE A-82]**

2. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology 30: 130-139. **[JH] [PAGE A-84]**

WELSCH STUDY

1. Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology **14: 359-367**. 2. Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A. **[RK, KP, JH] [PAGE A-85]**

FINAL COMMENTS [PAGE A-91]

Key to Statistical Subpanel's Primary Statistical Evaluator for Each Study

JH: Joe Haseman
RK: Ralph Kodell
JB: John Bailer
KP: Ken Portier

Note: Fifth Subpanel member Richard Morris was responsible for SAS-formatting the raw data as well as evaluating the "23 questions."

Appendix A:
Detailed Evaluations of Individual Studies

ASHBY DATASETS

1. Ashby, J., H. Tinwell, et al. (1999). "Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF1 mice exposed in utero." Regulatory Toxicology and Pharmacology 30: 156-166.

Agents: BPA (bisphenol A), 2 and 20 ug/kg/day
DES (diethylstilbestrol). 0.2 ug/kg/day
TSCO (tocopherol-stripped corn oil control),
NAIVE (not dosed control)

Procedure:

Each chemical was applied at various doses to pregnant females on a per body weight basis between pnd 11-17. Females were terminated at ~44 weeks. Males were kept in litter to pnd 112 then a subset randomly chosen and kept singly for an additional 71 days. All males terminated between pnd 183-187.

Parameters of Interest:

Table 3 - Females - Age at vaginal opening, onset and completion,
body wt.

Table 4 - Males - Prostate wt, Left and Right Testis wt, Epididymis wt,
Seminal vesicle wt.

Table 5 - Males - Total sperm/testis, Sperm wt/g testis, Sperm/testis
DSP, Sperm/g testis/d efficiency.

Covariate - body weight (BW).

Statistical Methods:

Manuscript indicated use of multi-factor ANOVA or ANCOVA using body weight as the covariate to assess significance among group means and to study dam effects for the various responses. Specific contrasts were examined comparing Naive control to TSCO controls, with pooling if no significance found. Where control differences were found, treatments were compared to controls separately. Dam effects were apparently treated as fixed effects in the ANOVA and because of this, it is not clear whether the variability associated with Dam effects is accounted for in the comparison of treatments.

Reanalysis used one way ANOVA and ANCOVA approaches with the Dam factor assumed to

be a random effect and SAS Proc Mixed used to perform computations for statistical tests. Dunnett and Hsu's MCP in mixed effects models, were computed to examine treatment effect differences with the two controls.

Results:

Table 3

Sexual Maturation in Females: Authors identified strong dam (fixed) effects for age at onset of vaginal opening, vaginal opening completion age, and body weight. Marginally significant differences between control groups were observed ($p < 0.1$). DES group showed statistically delayed onset of vaginal opening compared to TSCO controls ($p=0.014$). Restudy replicated these findings with treatment effects significantly different ($df=4$, $p=0.0425$). Restudy also found a significant ($p<0.001$) linear relationship of body weight with vaginal opening completion age, but no additionally significant treatment differences were observed.

Table 4

(1) Male Pup Body Weight: Authors pooled controls and found highly significant dam effects not explained by litter size. Both low and high dose BPA tended to be heavier than controls with low dose BPA being statistically significant ($p<0.05$). Authors also compared isolated housed pups to group housed pups for the control groups and found isolated group on average significantly heavier.

The restudy did not incorporate any of the group housed pups. DAM was incorporated as a random effect in a Mixed model for the analysis. Dam was highly significant ($p=0.0075$) as were treatment effects ($p=0.018$). Controls were not significantly different but were not pooled for this analysis. Low dose BPA was found to result in significantly higher BW than either controls using Dunnett's test.

(2) Prostate Weight: Authors found a significant association with BW and hence used an ANCOVA model for further analysis. The slightly elevated prostate weights in the BPA group was attributed to BW differences since no treatments were significant in the ANCOVA analysis other than DAM effects.

The restudy found strongly significant BW association with marginally significant DAM effects ($p=0.046$). No other effects were found significant.

(3) Seminal Vesicle Weight: Authors found differences between the two controls with TSCO lower than NAIVE which was the reverse pattern found with BW. A significant association between SVW and BW was observed and subsequent analysis used ANCOVA. DAM effect was significant but no treatment effects were observed.

The restudy found significant BW association, DAM effects and differences in controls. Treatments were not found significant.
(4) Testicular Weight: Controls were pooled and a significant ($p<0.001$) association with both

left and right TW and BW observed, hence ANCOVA used. Right TW was significantly larger than left TW across all treatment groups. A significant DAM effect was observed with BPA treatments significantly higher than pooled controls.

The restudy found for both left and right testicle significant DAM effects and BW association with high dose BPA significantly higher than controls. No low dose BPA effects were observed.

(5) Epididymal Weight: Controls were pooled and BW covariate used in the analysis. Left and right EW were similar. DAM effect was significant, and after adjustments, high dose BPA was marginally ($p\sim=0.05$) higher than pooled controls for both left and right EW.

The restudy found for right EW significant DAM effects, no BW association and the high dose BPA significantly greater than TSCO control. For left EW, significant DAM effects and BW associations were observed and again a significant high dose BPA to TSCO controls comparison.

Table 5

Daily Sperm Production and Efficiency: No control differences, BW associations or DAM effects were observed so pooled controls were used and analysis performed with adjusting for DAM or BW effects. DSP increased significantly for both doses of BPA, efficiency was unaffected by treatments.

The restudy found a significant difference ($p=0.028$) between high dose of BPA and TSCO controls for Sperm/testis/d (x 106). Everything else was not significant.

Final Comments:

The results reported in Ashby, Tinwell, and Haseman were essentially reproduced. The restudy analysis did not pool controls and dam effects were considered random. Dam and body weight effects in the restudy seemed more statistically significant than originally reported by Ashby but this may be due to the different analysis models used. Body weight covariate slopes when significant were usually had very small p-values. DAM variability was estimated via REML and ranged from 25% to 65% the magnitude of the residual variability, with typical values around 55%.

As noted in the text, the Statistics Subpanel did not have time to compare and contrast results across different studies and to speculate as to why similar or differing results were observed. However, because of the importance of this particular study in its attempt to replicate the results of Nagel et al. (see discussion of the Nagel study below in the section on vom Saal datasets), we decided to compare directly the body weights and prostate weights from these two studies, and this information is summarized below (SD: standard deviation; CV: coefficient of variation).

Ashby et al.

	No. of pups	Body weight Mean	SD	CV	Prostate weight Mean	SD	CV
Naive	22	42.0	3.1	.074	50.0	7.0	.140
TSCO	32	43.5	5.0	.115	48.4	8.2	.169
BPA-2	37	47.9	6.1	.127	53.2	8.2	.154
BPA-20	29	45.9	5.1	.111	50.3	9.1	.181

Nagel et. al

Controls	11	37.9	2.8	.074	40.8	3.3	.081
BPA-2	7	34.6	3.0	.087	52.8	6.1	.116
BPA-20	7	36.7	3.0	.082	54.9	16.7	.304

Notable points:

(i) the Ashby study was somewhat larger than the Nagel study;

(ii) the animals in the Ashby study were much heavier than the animals in the Nagel study;

(iii) the low dose of BPA significantly increased body weight in the Ashby study and significantly decreased body weight in the Nagel study;

(iv) the prostate weights in six of the seven groups were comparable, the exception being the lower prostate weights in the Nagel controls;

(v) there was more variation in the within group variability in prostate weight in the Nagel study than in the Ashby study; and

(vi) the Nagel study found significantly ($p<0.05$) elevated prostate weights in both BPA groups; the Ashby study did not.

2. Ashby, J., H. Tinwell, et al. (2000). "Current issues in Mutation Research. DNA adducts, estrogenicity and rodent diets." Mutation Research (in press).
3. Ashby, J., H. Tinwell, et al. (2000). "Uterotrophic activity of a "phytoestrogen-free" rat diet." Environmental Health Perspectives 108(1): A12-A13.

Raw data provided: uterus weights from rats fed one of three rodent diets: RM1, AIN-76A and Purina 5001. Three separate studies were carried out.

Comment on Statistical Methodology

These brief communications provided little or no description of the statistical methods used in the data analysis.

Results

Study 1

(1) We agree with the authors that the first study (summarized in Figure 1 of the short Mutation Research paper) showed a significant ($p<0.01$) increase in uterine weight in both the AIN-76A and Purina 5001 diets relative to the RM1 diet (the standard rat maintenance diet used by the study authors).

(2) These increases were accompanied by a corresponding significant ($p<0.01$) increase in body weight in animals receiving the AIN-76A and Purina 5001 diets. However, after adjusting for body weight (ANCOVA), the increased uterine weights remained significant ($p<0.01$) for both diets. The increases in body weight (13-16%) in the two diets were much less than the corresponding increases in uterine weight (wet weights: 61-79%; dry weights: 44-52%).

(3) Body weight was significantly ($p<0.05$) correlated with uterine weight in these studies.

Study 2

(1) In a second study (summarized in the EHP Letter to the Editor), we agree that the increase in uterine weight was confirmed for rats receiving the AIN-76A diet relative to those on the RM1 control diet.

(2) The authors report that this increase in uterine weight "was abolished by concomitant treatment with the antiestrogen Faslodex." While this statement is technically true, it is noteworthy that the group receiving the AIN-76A diet + Faslodex actually had significantly ($p<0.01$) lighter uteri than the animals receiving the RM1 control diet, after adjusting for body weight differences (the animals on the AIN-76 and AIN-76 + Faslodex diets were significantly ($p<0.05$) heavier than animals receiving the RM1 diet).

(3) Body weight was significantly ($p<0.01$) correlated with uterine weight in this study.

Study 3

(1) In a third study (summarized in the EHP Letter to the Editor, the authors confirmed in a larger study (30 animals per group compared with 10-12 for the other two studies) that the AIN-76A diet resulted in elevated uterus weights (and elevated body weights) relative to rats receiving the RM1 control diet. We agree with this interpretation of the data.

(2) Body weight and uterus weight were significantly ($p<0.01$) correlated in this study.

Comment

The consistent and highly significant ($p<0.01$) elevated uterine weights in rats receiving the AIN-76 diet compared to those receiving the RM1 control diet provide very convincing evidence that this is a real biological effect.

4. Ashby, J., H. Tinwell. (2000). "Activity of bisphenol A in pregnant SD and Alpk rats: preliminary data." (Unpublished Abstract).

Raw data provided: body weight and anogenital distance data for male and female rats exposed to bisphenol A (BPA) at doses of 0.02, 0.1, or 50, or mg/kg/day. Ethinyl estradiol (EE) acted as the positive control. This is a study in progress, and these were the only data we received in time to analyze prior to the date of the Endocrine Disruptor meeting.

Results

The anogenital distance (AGD) and body weight data received are summarized below. This study used littermates (generally 4-8 pups per litter), and there were highly significant ($p<0.01$) litter effects for both body weight and AGD (both strains). Thus, the summary statistics reported below are based on litter averages, and the statistical analysis also used the litter as the experimental unit.

I. ALPK rats

Chemical	Body weight						AGD					
	Males			Females			Males			Females		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Controls	7	6.78	0.73	7	6.23	0.66	7	4.81	0.52	7	2.72	0.23
BPA - 0.02	7	6.91	0.79	7	6.49	0.73	7	4.93	0.49	7	2.89	0.41
BPA - 0.1	6	6.74	0.79	6	6.35	0.72	6	4.75	0.32	6	2.70	0.19
BPA - 50	7	5.90	0.46	7	5.70	0.60	7	4.51	0.24	7	2.66	0.21
EE	3	6.19	0.38	3	5.98	0.68	3	4.38	0.12	3	2.67	0.37

The slight reductions in body weight in the 50 mg/kg/day BPA groups were not statistically significant by Dunnett's test, although the 13% reduction in the top dose male group was suggestive ($p<0.10$). We further agree with the authors that there was no significant BPA (or EE) effects on AGD, with or without adjusting for body weight. AGD was also significantly ($p<0.01$) correlated with body weight in both males and females.

II. SD rats

Chemical	Body weight						AGD					
	Males			Females			Males			Females		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Controls	7	7.04	0.53	7	6.49	0.54	7	4.45	0.09	7	2.55	0.13
BPA - 0.02	7	6.66	0.47	7	6.35	0.50	7	4.48	0.12	7	2.50	0.14
BPA - 0.1	7	6.91	0.33	7	6.73	0.22	7	4.60	0.20	7	2.59	0.12
BPA - 50	7	7.44	0.65	7	6.86	0.67	7	4.61	0.24	7	2.58	0.15

We agree with the authors that there was no significant BPA effects on body weight or on AGD, with or without adjusting for body weight. AGD was also correlated with body weight in both

males (p=0.07) and females (p<0.01).

Final comment

Additional raw data were provided by the authors after the Low Dose Endocrine Disruptor meeting for preputial separation and vaginal opening. Statistical analyses of these data revealed that the only statistically significant effect (other than significant litter effects) was a slight, but statistically significant increase in day of vaginal opening in the top dose BPA group and a slight decrease in the EE group for ALPK rats (with or without body weight adjustment), as is shown below.

Chemical	Day of vaginal opening			Weight at vaginal opening		
	N	Mean	SD	N	Mean	SD
Controls	7	33.81	0.83	7	109.1	8.0
BPA - 0.02	7	34.02	0.71	7	116.4	12.3
BPA - 0.1	6	33.61	0.80	6	109.2	4.9
BPA - 50	7	35.40*	0.65	7	117.5	11.0
EE	3	32.13*	1.52	3	105.1	7.2

* p<0.05 vs.controls (Dunnett's test)

5. Odum, J., P. A. Lefevre, et al. (1997). "The rodent uterotrophic assay: critical protocol features, studies with nonylphenols and comparison with a yeast estrogenicity assay." Regulatory Toxicology and Pharmacology 25: 176-188.

Raw data provided:

(i) uterus wet weight and vaginal opening data for 17beta-estradiol given either by sc injection (0.5, 1, 2, 5, 10, 20, 40, 200 or 400 ug/kg) or by oral gavage (10, 20, 40, 100, 200, or 400 ug/kg). Data were summarized in Table 1 and in Figure 2.

(ii) uterus wet weight and vaginal opening data for 17beta-estradiol benzoate given either by sc injection (0.1, 0.5, 1, 2, 5, 10, 20, 25 or 125 ug/kg) or by oral gavage (10, 20, 40, 100, 200, or 400 ug/kg). Data were summarized Figure 3.

(iii) uterus wet weight and vaginal opening data for ethinyl estradiol given either by sc injection (0.02, 0.2, 2, 20, 200, or 400 ug/kg) or by oral gavage (0.02, 0.2, 2, 20, 200, or 400 ug/kg). Data were summarized Figure 3.

(iv) uterus wet weight and vaginal opening data for branch-chain nonyl phenol (NP): 47.5, 95, 190, and 285 mg/kg, 285 mg/kg NP + 10 mg/kg ICI 182,780, positive control (400 ug/kg estradiol) and negative control (arachis oil). Data were summarized in Figure 11.

(v) uterus wet weight and vaginal opening data for four replicates, each involving 285 mg/kg

branched nonyl phenol (NP), 285 mg/kg n-nonyl phenol (nNP), positive control (400 ug/kg estradiol) and negative control (arachis oil). Data were summarized in Figure 12.

Comments on Statistical Methodology

No statistics were used in this paper.

Results

Table 1/Figure 2

(1) There is no significant 17beta-estadiol effect on body weight, even at the highest doses. In contrast, there is a significant ($p<0.01$) effect of 17-beta-estradiol on uterine weight. Uterus weights are significantly elevated at doses of 5 or greater in the sc injection animals and at 100 or greater in the oral gavage animals.

(2) The vaginal opening data follow the same pattern of response as the uterine weight data.

(3) We agree with the authors' conclusions that 17beta-estradiol is more active in the uterotrophic assay when injected subcutaneously than when given by oral gavage.

Figure 3

(1) There is no significant 17beta-estadiol benzoate effect on body weight, even at the highest doses. In contrast, there is a significant ($p<0.01$) effect of 17-beta-estradiol benzoate on uterine weight. Uterus weights are significantly elevated at doses of 0.5 or greater in the sc injection animals and at 40 or greater in the oral gavage animals.

(2) The vaginal opening data follow the same pattern of response as the uterine weight data.

(3) We agree with the authors' conclusions that 17beta-estradiol benzoate is more active in the uterotrophic assay when injected subcutaneously than when given by oral gavage.

Figure 4

(1) There is no significant ethinyl estradiol effect on body weight, except for a 10% reduced body weight ($p<0.05$) in the 400 ug/kg group in the sc injection portion of the study. In contrast, there is a significant ($p<0.01$) effect of ethinyl estradiol on uterine weight. Uterus weights are significantly elevated at doses of 0.2 or greater in the sc injection animals and at 2 or greater in the oral gavage animals.

(2) The vaginal opening data follow the same pattern of response as the uterine weight data.

(3) We agree with the authors' conclusions that ethinyl estradiol is more active in the uterotrophic assay when injected subcutaneously than when given by oral gavage, although this difference is less pronounced than that observed for 17beta-estradiol or for 17-beta estradiol

benzoate.

Figure 11

(1) For the Figure 11 data, we agree with the authors that NP produced a dose-related increase in uterine weight (significant ($p<0.001$) at the top two dosed groups), that was abolished by coadministration of the estrogen receptor antagonist ICI 182,780.

(2) No cases of vaginal opening were observed in any of the NP groups.

(3) Body weight was significantly ($p<0.05$) reduced in the top dose NP group, but not in any other experimental group.

(4) The correlation of uterus weight with body weight was significant ($p<0.01$) if treatment groups are ignored, but not significant after adjusting for treatment effects. Using or not using body weight as a covariable had little impact on study results.

Figure 12

(1) We agree with the study authors that NP (but not nNP) produced a significant increase in uterus weight that was consistently observed across all four replicates.

(2) We also agree that the five instances of vaginal opening among the 20 NP animals (none were observed in the 20 controls or in the 20 nNP animals) provides further evidence of uterotrophic activity in the NP group.

(3) Body weight was significantly ($p<0.01$) reduced in the NP group, but not in the other two experimental groups.

(4) The correlation of uterus weight with body weight was significant ($p<0.01$) if treatment groups are ignored, but not significant after adjusting for treatment effects. Using or not using body weight as a covariable had little impact on study results.

(5) Neither the replicate effect, nor the replicate x treatment interaction were significant, which indicates that the uterine weight responses were reproducible and consistent across the four replicates.

(6) We basically agree with the authors' conclusions. We might have characterized the 32% increase in uterine weight observed in the NP group in the Figure 12 data (mean of 47.045 mg vs. 35.59 for controls) as being more than only a "weak assay response" as do the authors on page 184. However, that is a matter of scientific judgement. Certainly, the response is "weak" relative to the positive control response. The consistency of the uterine weight responses (in all groups) among four replicates in Figure 12 was an impressive finding.

Final Comment

We basically agree with the authors' interpretation of these data.

6. Odum, J. and J. Ashby (1999). "Neonatal exposure of male rats to nonylphenol has no effect on the reproductive tract." Toxicological Science (in press).

Raw data provided: testis, epididymis, seminal vesicle, and ventral prostate weights from male rats given i.p. nonylphenol (NP) doses of 8 mg/kg/day from day 1 to day 10 post-partum in arachis oil (Experiment 1). In Experiment 2, animals received 8 mg/kg/day NP, administered in DMSO. In addition to this group and a DMSO control group, a third group received 8 mg/kg/day NP and "an additional dose of DMSO". A fourth group received NP together with an i.p. dose of 0.5 mg/kg/day of the estrogen receptor antagonist Faslodex (FAS). A fifth group received FAS alone. These data (which included multiple pups per litter) are summarized in Table 1.

Comments on Statistical Methodology

The authors appear to have used a Protected Fisher's LSD analysis to make pairwise comparisons, although they describe their procedure slightly differently. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Table 1 - Experiment 1

All four organ weights (and body weight) showed highly significant ($p<0.01$) litter effects, indicating that the litter rather than the individual pup is the appropriate experimental unit. It was unclear whether the study authors used a litter-based or a pup-based statistical analysis, but the summary statistics given in Table 1 are calculated on a per-pup basis.

In any case, we agree with the authors' conclusion that there is no evidence of an NP effect on body weight or on organ weight. Organ weights were significantly ($p<0.05$) correlated with body weight in this study.

Table 1 - Experiment 2

All four organ weights (and body weight) showed highly significant ($p<0.01$) litter effects, indicating that the litter rather than the individual pup is the appropriate experimental unit. It was unclear whether the study authors used a litter-based or a pup-based statistical analysis, but the summary statistics given in Table 1 are calculated on a per-pup basis.

In any case, we agree with the authors' conclusion that there is no evidence of an NP effect on body weight or on organ weight. Neither was there any effect of FAS or FAS + NP on organ or body weight. Organ weights were significantly ($p<0.05$) correlated with body weight in this study.

7. Odum, J., I. T. G. Pyrah, et al. (1999). "Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays." Regulatory Toxicology and Pharmacology 29: 184-195.

Raw data provided:

(i) uterus weight data from Noble rats exposed to oral doses of 45, 75, 150 or 225 mg/kg/day p-nonylphenol (NP) or to 0.04 mg/kg/day DES. E2 was used as a positive control. These data are summarized in Figures 1 and 2

(ii) mammary gland differentiation data - number of structures per square millimeter of mammary gland for animals receiving 0.076 mg/kg/day DES or 0.073 or 53.2 mg/kg/day NP. These data are summarized in Table 2.

(iii) mammary gland differentiation data - area per mammary gland structure for animals receiving 0.076 mg/kg/day DES or 0.073 or 53.2 mg/kg/day NP. These data are summarized in Table 3.

(iv) the effect of NP (0.073 or 53.2 mg/kg/day) and DES (0.076 mg/kg/day) on the numbers of S-phase cells as percent total cells in Noble rats. These data are summarized in Table 5.

Comments on Statistical Methodology

The authors appear to have used a Protected Fisher's LSD analysis to make pairwise comparisons, although they describe their procedure slightly differently. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Results

Figure 1

(1) The positive control (E2) significantly ($p<0.01$) increased uterine weight, as expected.

(2) NP had no significant effect on body weight.

(3) Uterine weight was significantly ($p<0.01$) correlated with body weight in these studies.

(4) We agree with the authors that the top three doses of NP produced significantly ($p<0.01$) elevated uterine weights, after adjusting for body weight differences. However, we found that by ANCOVA and Dunnett's test, this elevation is also significant ($p<0.01$) at the lowest (45 mg/kg/day) dose NP group, whereas the study authors reported only a $p<0.05$ elevation for wet weight and no significant elevation for dry weight at this dose.

Figure 2

(1) We found some minor numerical discrepancies: we could find only 5 NP 150 mg/kg/day animals with the 11 daily oral doses - the authors report 6; conversely, we found 6 controls for the 3 daily oral doses - the authors report 5 animals in this group.

(2) When given in three oral doses, there was no DES or NP effect on body weight; when given in 11 oral doses, DES (but not NP) significantly ($p<0.05$) reduced body weight.

(3) Body weight and uterus weight were significantly ($p<0.01$) correlated for these data.

(4) We agree with the authors that both DES and 150 mg/kg/day NP significantly ($p<0.05$) increased uterus weight in both studies (3 daily oral doses or 11 daily oral doses). We also agree that the low dose NP group (53 mg/kg/day) had no significant effect on uterus weight.

(5) We note the relatively low sample sizes in these studies (3-6 animals per group).

Table 2

(1) We agree with the study authors that NP did not affect the number of mammary gland structures for terminal ducts, terminal endbuds, or Lobules type 1 or 2 for either peripheral or central regions. We were not provided raw data for Lobules type 3, but there was no response in the NP or control animals for Lobule 3.

(2) As noted in the comments on a previous study by this investigator, we do not believe that a normal-theory based test is appropriate for highly skewed data in which the majority of the data points are zero, as was the case for Lobules 2 and 3.

(3) We agree with the study authors that DES produced a significant ($p<0.01$) increase in the number of structures for Lobules 1 (peripheral area only) and Lobules 2 (both peripheral and central areas).

(4) We further agree with the study authors that DES significantly ($p<0.01$) reduced the number of structures for terminal ducts (both regions) and terminal end buds (central area only)

(5) Table 2 reports that the slightly reduced DES response for terminal end buds in the peripheral area is highly significant, whereas it is in fact not significant. The authors subsequently confirmed our suspicion that this is a typographical error. Figure 4 and the text appear to be correct in this regard.

(6) We note that sample size is missing for the NP 0.073 mg/kg/day group in Table 2

Table 3

(1) There is a minor conceptual issue that requires some discussion: is it valid to compare the areas of structures for those animals having structures with the areas for those animals having no

structures (and thus by definition, we suppose, zero areas)?

In response to our question on this issue, the authors emailed us back that "We decided that when counting numbers of structures that 0 should contribute towards the mean but that for calculating mean area per structure or proliferating cells within a structure that animals or fields which had no structure should be excluded. This was the rule for both mammary gland papers." This is a very reasonable position, but it was not followed consistently in the authors' evaluation.

For example, in Table 3, for terminal ducts the 0.076 mg/kg/day DES group is reported as having an area of zero for both peripheral and central areas, which is reported to be significantly ($p<0.01$) lower than the mean area for the control group. However, using the philosophy given by the authors above, the DES group should have been excluded altogether from an analysis of area, since they had no structures (see Table 2). Similar problems occurred for other groups in this table.

Despite the length of the discussion given above on this issue, the Statistics Subpanel viewed this as a relatively minor matter that had little or no impact on the authors' major conclusions.

(2) We agree with the study authors that NP did not affect the area of mammary gland structures for terminal ducts, terminal endbuds, or Lobules type 1 or 2 for either peripheral or central regions. We were not provided raw data for Lobules type 3, but there was no structures in the NP or control animals for Lobule 3.

(3) As noted in the comments on a previous study by this investigator, we do not believe that a normal-theory based test is appropriate for highly skewed data in which the majority of the data points are zero, as was the case for Lobules 2 and 3.

(4) We agree with the study authors that DES produced a significant ($p<0.01$) increase in the area of structures for Lobules 1 and 2 (peripheral area only), if the comparisons are limited only to those animals with structures (see point (1) above).

(5) We further agree with the study authors that DES significantly ($p<0.01$) reduced the number of structures for terminal ducts (both regions) and terminal end buds (central area only) under the philosophy that no structure = zero area (see point (1) above).

(6) We note that sample size is missing for the NP 0.073 mg/kg/day group in Table 3.

Table 5

(1) We agree with the authors' conclusions concerning the significant ($p<0.05$) increases in S-phase cells produced by DES in Lobules 1 and 2 (and in Lobule 3, except that there were no structures in the control group for comparison).

(2) We also agree with the authors that neither dose of NP significantly affected the number of S Phase cells as a percentage of total cells in any of the five regions.

8. Odum, J., I. T. G. Pyrah, et al. (1999). "Effects of p-nonylphenol and diethylstilbestrol on the alderley park rat: comparison of mammary gland and uterus sensitivity following oral gavage or implanted mini-pumps." Journal of Applied Toxicology 19: 367-378.

Raw data provided:

(i) uterus weights for Alpk rats given oral doses of 37.5, 75, 150, or 225 mg/kg/day p-Nonylphenol (NP) (Experiment 1) or in Alpk and SD rats given 250 mg/kg/day NP (Experiment 2). These data are summarized in Table 2;

(ii) uterus weights in Alpk rats given an oral dose of 100 mg/kg/day NP or 0.01 mg/kg/day DES (Experiment 3). These data are summarized in Table 3;

(iii) uterus weights in Alpk rats given s.c. minipump doses of 0.037 or 27.2 mg/kg/day NP or 0.042 mg/kg/day DES. These data are summarized in Table 3;

(iv) mammary gland differentiation (numbers and area of structures per squared mm mammary gland) for intact Alpk rats receiving via s.c.-implanted minipumps (for 11 days) 0.052 or 37.4 mg/kg/day NP or 0.055 mg/kg/day DES (Experiment 5). These data are summarized in Table 4.

(v) mammary gland differentiation (numbers of structures per squared mm mammary gland) for intact Alpk rats receiving via s.c.-implanted minipumps (for 11 days) 100 mg/kg/day NP or 0.075 mg/kg/day DES (Experiment 6). These data are summarized in Table 5.

(vi) the effect of oral NP (100 mg/kg/day) and DES (0.075 mg/kg/day) on the numbers of S-phase cells as percent total cells in Alpk rats (Experiment 6). These data are summarized in Table 6.

Comments on Statistical Methodology

The authors appear to have used a Protected Fisher's LSD analysis, although they describe their procedure slightly differently on page 369. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Results

Table 2 (Experiment 1)

(1) We agree with the authors that the top three NP doses (but not the lowest dose) and the single E2 sc dose (positive control) significantly ($p<0.01$) increased uterine weight relative to controls.

(2) Neither E2 nor NP significantly affected body weight.

(3) Uterine weight was significantly ($p<0.05$) correlated with body weight in this study.

Table 2 (Experiment 2)

(1) We agree with the authors that in both the Alpk and SD rats oral administration of 250 mg/kg/day NP significantly ($p<0.01$) increased uterine weight relative to controls.

(2) NP did not significantly affect body weight.

(3) Uterine weight was not significantly ($p<0.05$) correlated with body weight in this study.

Table 3 (Experiment 3)

(1) We agree with the authors that oral administration of 100 mg/kg/day NP and 0.01 ml/kg/day DES (positive control) significantly ($p<0.01$) increased uterine weight relative to controls.

(2) Neither DES nor NP significantly affected body weight.

(3) Uterine weight was significantly ($p<0.05$) correlated with body weight in this study.

(4) The study authors inadvertently provided us raw data from five DES animals that were not part of Experiment 3. They subsequently confirmed that our decision to delete them from the statistical analysis was correct.

(5) In Table 3, Experiment 3, the uterus dry weights have been interchanged for the DES and NP groups. Actually, the DES animals had heavier dry and wet uterus weights relative to the NP animals, not lighter dry weights as reported in Table 3. Moreover, we had a slight disagreement with the summary statistics for the wet and dry uterus weights for the NP group. Based on the data provided to us, our calculated wet weights were 93.0 +-16.1 (not 91.0 +- 20.6) and the dry weight was 19.2 +- 3.4 (not 18.8 +- 4.3).

Table 3 (Experiment 4)

(1) We agree with the authors that sc administration (mini-pump) of 0.037 or 27.2 mg/kg/day NP had no significant effect on uterus weight (or on body weight).

(2) We also agree that sc minipump administration of 0.042 mg/kg/day DES (positive control) significantly ($p<0.01$) increase uterine weight relative to the DMSO controls. This dose of DES also significantly ($p<0.01$) reduced body weight.

(3) Uterine weight was not significantly ($p<0.05$) correlated with body weight in this study.

Table 4 (Experiment 5)

(1) We agree with the study authors that neither NP nor DES affected the number of mammary gland structures for terminal ducts, terminal endbuds, or Lobules type 1 for either peripheral or central regions.

(2) The authors report a highly significant ($p<0.01$) DES effect on the number of structures for Lobules type 2 and 3 and both peripheral and central regions, but we disagree with this conclusion. The authors based their conclusion on the application of Student's t test following a general linear models analysis (see page 369). However, such an approach (which assumes an underlying normal distribution) is inappropriate when virtually all of the data are zero, as was the case for Lobules types 2 and 3.

For example, for LobuIes type 3 in the peripheral region, every animal in every group had a zero response, except for a single animal in the DES group, which showed a response of 0.30305. Based on this one animal's positive response (with the nine other animals in the DES group showing no response), the authors concluded that a highly significant ($p<0.01$) increase had been observed. We disagree with this interpretation of the data.

Similar problems occurred in the other DES groups. The only statistically significant DES effect that we could confirm (and it was $p<0.05$, not $p<0.01$), was for Lobules type 2 in the central region.

If positive responses are rare, then certain of the DES effects reported by the authors could in fact be "real" and biologically important (see discussion of this matter in the text). However, it is misleading (and incorrect) to attach a high level of statistical significance ($p<0.01$) to the response of a single animal (or even to two animals), as was the case in this study

(3) The results for area of structures closely parallels that of number of structures.

Table 5 (Experiment 6)

(1) We generally agree with the authors concerning the significant effects of DES and NP in this experiment.

(2) The exceptions include the reduced effect seen by NP in the peripheral region for terminal end buds. By Dunnett's test, this reduction was not statistically significant relative to controls.

(3) Perhaps more importantly, the increase observed in Lobules type 3 in the central region for DES (0.12 +- 0.27), which reflected positive responses in only two of the ten animals was not statistically significant relative to the zero response in the ten controls, whereas the authors reported this slight increase as highly significant ($p<0.01$). Interestingly, an identical response produced by DES for Lobules type 2 in the peripheral region (also 0.12 +-0.27 vs zero in the controls) was not reported as being statistically significant, even at the $p<0.05$ level. We agree with this interpretation and believe that it applies to the DES Lobules type 3 effect as well.

(4) The results for area of structures closely parallels that of number of structures.

Table 6 (Experiment 6)

We agree with the authors' conclusions concerning the increases in S-phase cells produced by DES and NP as reported in Table 6.

9. Tinwell, H., R. Joiner, et al. (2000). "Uterotrophic activity of bisphenol A in the immature mouse." Regulatory Toxicology and Pharmacology (in press).

Data provided: uterus weight and body weight data from nine independent immature mouse uterotrophic assays involving bisphenol A (BPA) administered by sc injection (eight studies) or by oral gavage (one study). DES was included as a positive control in each study. Doses of BPA (/kg) used in one or more of the eight sc studies were 0.02 ug, 0.2 ug, 2 ug, 20 ug, 200 ug, 500 ug, 1 mg, 2 mg, 5 mg, 10 mg, 20 mg, 50 mg, 100 mg, 200 and 300 mg, The oral gavage doses (/kg) of BPA were 500 ug, 1 mg, 5 mg, 10 mg, 50 mg, 100 mg, 200 mg, or 300 mg. Data are summarized in Table 1.

Comments on Statistical Methodology

In general, the statistical methods appear appropriate for the data. However, it is unclear how the authors used ANOVA (an overall test) to make the pairwise comparisons whose significance is reported in the various tables in this paper.

Results

(1) DES showed significant ($p<0.01$) and reasonably reproducible elevated uterine weights in all nine studies.

(2) In most of the studies, body weight and uterine weight were significantly ($p<0.05$) correlated.

(3) Body weights were unaffected by BPA in all studies.

(4) Our reanalyses used all the available data. A few animals were deleted by the study authors, apparently based on an a priori decision to exclude any animals whose body weight was 18g or more. These animals did not appear to show extreme or "outlier" uterus weights, and we included them, although exclusion of these animals would have had essentially no impact on study results.

Experiment 1

Our reanalysis confirmed the significantly ($p<0.05$) elevated uterine weight in the 200 mg/kg BPA group. However, the response in the 200 ug/kg group showed significant heterogeneity. The range of uterus weights in the eight control animals was 7-10 mg. Five of the 200 ug/kg animals had uterus weights above this range (11.3 to 16.5 mg), but the other three animals had uterine weights that were below (or nearly so) the lowest control value: 6.6, 6.9, and 8.1 mg. Because of this high variability (which prompted a nonparametric analysis for this group), we did not find the elevated uterus weights in the 200 ug/kg group to be significant, as reported by the authors.

Experiment 2

We agree with the authors that there were no significant uterus weight effects in this experiment,

in which the eight doses of BPA used ranged from 0.02 ug/kg to 200 mg/kg.

Experiments 3 and 4

We agree with the authors that there were no significant uterus weight effects in either of these experiments, in which the five doses of BPA used ranged from 0.2 ug/kg to 300 mg/kg.

Experiment 5

We agree with the authors that there were no significant uterus weight effects in this experiment, in which the four doses of BPA used ranged from 2 ug/kg to 200 mg/kg.

Experiments 6 and 7

We agree with the authors that in both of these two studies, there were significantly ($p<0.05$) elevated uterine weights in the 200 mg/kg BPA group, but not in the 20 ug/kg group. Body weight was unaffected at all doses. We also agree that there was no effect of BPA on uterine weight in the 200 ug/kg BPA group in Experiment 6. However, in Experiment 7, after adjusting for body weight, the elevated uterine weight in the 200 ug/kg group was significant ($p<0.05$), whereas the authors reported no significance for this dosed group.

Experiment 8

We agree with the study authors that there was a significantly ($p<0.05$) elevated uterine weight in the 5 mg, 50 mg, 100 mg, and 200 mg groups. Additionally, the elevated uterine weight in the 200 ug/kg/day group was suggestive ($p<0.10$). There were no significant body weight effects in this study.

Experiment 9

We agree with the authors that there were no significant uterus weight effects in this oral gavage experiment, in which the eight doses of BPA used ranged from 500 ug/kg to 300 mg/kg. Body weight was also unaffected at all doses.

Evaluating the consistency of overall results of for the 8 sc experiments

The authors summarize the results of these eight studies in Table 2, and with so many sc experiments, it is tempting to carry out an overall analysis. A visual comparison of the overall means can be misleading, since not all doses were represented in all experiments, and there was significant study-to-study variability in response.

Our reanalysis indicates that for BPA doses in the 0.02 to 20 ug/kg range, there is no evidence of an elevated uterus weight in these eight studies. The 200 ug/kg dose was used in six of the eight studies, and three of these studies produced elevations in uterus weight that were either significant ($p<0.05$) or nearly so. In contrast, the other three experiments produced little or no

evidence of an effect. An analysis of log(uterus weight) adjusting for body weight and for differences among experiments, reveals a marginally significant ($p<0.05$) effect. This effect, if real, represents (on average) only approximately an 8% increase in uterus weight. As noted earlier, a direct comparison of overall means is misleading, since the two experiments for which the 200 ug/kg dose was not used had two of the three highest control mean uterus weights.

Complicating the interpretation of this increase in uterine weight is the fact that uterus weight was not significantly elevated in the 500 ug/kg or 1 mg/kg groups (one experiment) or in the 2 mg/kg groups (three experiments). Interestingly, the single study using 5 mg/kg produced a significant ($p<0.05$) increase in uterus weight, while the same study at 10 mg/kg showed no effect. Four studies using 20 mg/kg showed no effect.

BPA doses in the 50-200 mg/kg range often showed significantly elevated uterus weights, although the 50 and 100 mg/kg doses were used in only a single experiment. The 200 mg/kg dose was used in all eight experiments, with significant uterine weight increases observed in four studies, marginally elevated uterine weights seen in a fifth, but no evidence of an effect in the other three experiments. Although significant overall ($p<0.01$), the uterine weight at this dose on average was only 12-13% elevated relative to controls. The 300 mg/kg dose produced no significant increase in uterine weight in two experiments, although the companion doses of 200 mg/kg used in these two studies also failed to produce a significant increase in uterine weight.

In summary, there is a wealth of information here, but it is difficult to draw firm overall conclusions. We agree with the authors' conclusion that "overall, we have failed to define BPA as reproducibly active in the immature mouse uterotrophic assay" and that they have shown "it is possible for individual investigators to be unable to confirm their own observations." We do not, however, agree that there was "a complete absence of a dose-relationship". In the four experiments in which a significantly increased uterine weight was observed (Experiments 1, 6, 7, and 8), this response followed a definite dose-related trend. However, in four other experiments (including two with the highest BPA doses), no significant BPA effect was apparent. And this inconsistency is what is most troubling.

DELCLOS DATASETS

1. Delclos, K. B., T. J. Bucci, et al. (2000). "Effects of dietary genistein exposure during development on male and female CD rats." In prep (Unpublished Abstract).

Chemical: Genistein

Raw data provided: In this study, female rats were dietarily exposed to genistein (0, 5, 25, 100, 250, 625, and 1250 ppm). The primary variables of interest were male reproductive tract organ weights (testes, prostate), perinatal responses (birthweight, AGD), markers of puberty (vaginal opening and preputial separation) and mammary gland pathology (not analyzed here).

Statistical Methodology

Mixed effects models were applied to all responses with litter/birthdam entering these models as

a random effect. All comparisons of the exposed groups to the control groups were conducted using Dunnett's procedure. Linear trends were evaluated by fitting regression models with random intercepts, a mixed model, to these data. Body weight was considered as a potential covariate in these analyses.

One concern with the analysis of the puberty data (preputial separation and vaginal opening day) was that at the time when the experiment was concluded (50 days), not all events had been observed. A crude imputation of value was conducted in this case where a value of 50.5 was assigned to these censored observations. Nine of the 137 observations in this data set had values imputed in this fashion. These nine values were distributed across all concentration conditions with no concentration having more than 20% censored. Vaginal opening data were not analyzed here.

Results:

1. Male organs: In mixed modeling with body weight as a covariate, no chemical effects on the dorsal prostate. For testes and ventral prostate responses, an interaction between dose and body weight was observed along with a dose effect. In other words, dose-related effects were observed for the testes and ventral prostate; however, this effect differed for animals of different body weights. Birthdam was an important source of variability in this analysis accounting for anywhere from 10% (testes) to 40% (ventral prostate) of the total variability.

2. Preputial Separation: None of the concentration groups differed from the controls. Birthdam/litter was a significant source of variability in this analysis.

Commentary:

This was similar to the analytic strategy reported by the investigators although no analysis of organ weight ratios was conducted. Not all responses were included in the data provided.

2. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary ethinyl estradiol exposure during development on male and female CD rats." In prep (Unpublished Abstract).

Chemical: Ethinyl Estradiol

Raw data provided: In this study, female rats were dietarily exposed to ethinyl estradiol (0, .1, 1, 5, 25, 100, and 200 ppb). The primary variables of interest were perinatal responses (birthweight, AGD), and markers of puberty (vaginal opening and preputial separation).

Statistical Methodology

Mixed effects models were applied to the puberty responses with litter/birthdam entering these models as a random effect. All comparisons of the exposed groups to the control groups were conducted using Dunnett's procedure. Linear trends were evaluated by fitting regression models with random intercepts, a mixed model, to these data.

Simple ANOVA models were applied to the birth weight and AGD data since these were already composite measurements defined for each litter. Dunnett's procedure were used to evaluate differences between the various concentration conditions and the control group.

One concern with the analysis of the puberty data (preputial separation and vaginal opening day) was that at the time when the experiment was concluded (50 days), not all events had been observed. A crude imputation of value was conducted in this case where a value of 50.5 was assigned to these censored observations. Vaginal data opening age required very little imputation of values (only 1 of 138). Preputial separation day was more severe. Censoring of these observations was most extreme in the 200 ppb concentration where 17 of 20 observations were imputed. The other 11 imputed values were distributed over the other 6 concentration conditions.

Results:

1. Perinatal responses: None of these responses, ranging from male AGD to pct. males in the litter exhibited a trend over the concentrations.

2. Preputial Separation: Only the 200 ppb group exhibited preputial separation day significantly different from the control group. A significant positive linear trend was also detected for this response. Birthdam/litter was a significant source of variability in this analysis (approx. 44% of total variability attributable to this component). The general pattern was for slight/no decrease before an increase in response

3. Vaginal opening day: Only the 200 ppb group exhibited vaginal opening day significantly different from the control group. A significant negative linear trend was also detected for this response. Birthdam/litter was a significant source of variability in this analysis (approx 50% of total variability attributable to this component). The general pattern was for slight/no increase before decrease in response.

4. Ovary weight - In our initial analysis, which used all the data provided to us, we felt that the ethinyl estradiol effect on ovary weight was less impressive than reported by the authors. However, in response to our concerns, the authors stated that they had plotted the data and that

"Examination of these plots indicated that one of the ovaries in the high dose, animal 74 in the 200 ppb dose group, was an outlier with a high ovary weight. The individual animal pathology report indicates that this animal had ovarian cysts; the only animal in the experiment that had this condition. This perhaps explains the high ovary weight for this animal. The results I reported in the narrative were based on the analysis of the data with this outlier excluded."

We agree that the ovary weight of this animal (0.214) was well outside the range for the other animals in this group (0.05 to 0.101), and that the authors' reasons (statistical and biological) for excluding this animal seem reasonable. With this animal excluded, we agree that the reduction in ovary weight in the 200 ppb dosed group (after adjusting for body weight and litter effects) is highly significant ($p<0.01$). It is a matter of scientific judgement if the 24% reduction in (adjusted) ovary weight at this dose is sufficiently impressive to be termed "dramatically

affected" as the authors characterize the response.

5. Testis weight - We agree with the authors that testis weight is significantly ($p<0.05$) reduced in the top dose (200 ppb) group. However, the characterization of the testis weight as "starting to decline" at 100 ppb is perhaps a bit of a stretch, since this decrease (after adjusting for body weight and litter effects) is only 1%. Thus, this appears to be basically a high dose only effect. However, we consider this to be a minor point in any case.

6. Prostate weight - Ventral prostate weight was unaffected. After adjusting for body weight and litter effects, dorsal prostate weight was marginally ($p=0.055$) increased in the 5 ppb group. The slight elevations in the other dosed groups were not statistically significant. The slight increase in the 5 ppb group may or may not be biologically important.

3. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep (Unpublished Abstract)

Chemical: Nonylphenol

Raw data provided: In this study, female rats were dietarily exposed to nonylphenol (0, 5, 25, 200, 500, 1000 and 2000 ppm). The primary variables of interest were male reproductive tract organ weights (testes, epididymis, seminal vesicle, prostate), markers of puberty (vaginal opening and preputial separation) and epididymal sperm count.

Statistical Methodology

Mixed effects models were applied to all responses with litter/birthdam entering these models as a random effect. All comparisons of the exposed groups to the control groups were conducted using Dunnett's procedure. Linear trends were evaluated by fitting regression models with random intercepts, a mixed model, to these data. Body weight was considered as a potential covariate in these analyses.

One concern with the analysis of the puberty data (preputial separation and vaginal opening day) was that at the time when the experiment was concluded (50 days), not all events had been observed. A crude imputation of value was conducted in this case where a value of 50.5 was assigned to these censored observations. This did not occur in the vaginal opening data set; however, it was quite common, especially in the 2000 ppm group, for the preputial separation response. In fact, 16 of the 20 animals in the 2000 ppm group in the preputial separation study were censored observations.

Results:

1. Male organs: In mixed modeling with body weight as a covariate, no chemical effects on the prostate or seminal vesicle was observed. For testes and epididymis responses, an interaction between dose and body weight was observed along with a dose effect. In other words, dose-related effects were observed for the testes and epididymis; however, this effect differed for

animals of different body weights. Birthdam was an important source of variability in this analysis accounting for anywhere from 8% (epididymis) to 50% (testes) of the total variability.

2. Epididymal sperm count: All of the concentration groups except the 2000 ppm group differed from the control. Eight of the 15 animals in the control group had zero counts recorded for sperm production. This dramatically reduced the mean value for the control group response which may explain the pairwise results. The general pattern was for slight/no increase before decrease in response (figure not included).

3. Preputial Separation: Only the 2000 ppm group exhibited preputial separation day significantly different from the control group. A significant positive linear trend was also detected for this response. Birthdam/litter was a significant source of variability in this analysis. The general pattern was for slight/no decrease before an increase in response (figure not included).

4. Vaginal opening day: Only the 2000 ppm group exhibited vaginal opening day significantly different from the control group. A significant negative linear trend was also detected for this response. Birthdam/litter was a significant source of variability in this analysis. The general pattern was for slight/no increase before decrease in response (figure not included).

Commentary:

This was similar to the analytic strategy reported by the investigators
although no analysis of organ weight ratios was conducted in our re-evaluation.

4. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of genistein in male and female Sprague Dawley rats." (Unpublished Final Report).

We were provided with the Executive Summary. More detailed written reports for this study were later received from the authors.

Raw data provided: Bone marrow cell number and colony forming unit (CFU) data and anti-CD3 data for male and female Sprague Dawley rats receiving dietary exposure of 0.0, 25, 250 or 1250 ppm genistein for 77 days.

Comments on statistical methodology:

(1) We agree with the statistical methods used, except that Dunnett's test does not require an overall ANOVA to be significant.

(2) The Executive Summary tables list only the "maximum effect" (and corresponding dose) for each variable, which could potentially be misleading, because it does not distinguish situations in which all three doses produce a significant effect and those for which only a single dose produces a significant effect.

Results

A. Bone marrow CFU data (Tables 1 and 2)

We agree with the authors' conclusion that "A statistically significant [$p<0.01$] dose-dependent decrease of 33% in CFU/GM x 10^5 cells was seen in the F1 generation male rats at 250 and 1250 ppm. A statistically significant decrease of 41% in cells/femur was observed in the F1 generation female rats at the high dose."

However, there are two other effects related to the lowest (25 ppm) dose that are not mentioned in the Executive Summary:

(i) The authors found (and we agree) a significant ($p<0.05$) 17% increase in CFU/GM x 10^5 in the 25 ppm F1 female rats.

(ii) Although the authors did not report this as statistically significant (and we agree), there was a suggestive ($p=0.06$) 28% decrease in cells/femur in the 25 ppm F1 males. This change may or may not be biologically important, but a very similar 28% decrease was also seen at this dose (25 ppm) in F1 females. A more global statistical analysis considering males and female together would find a significant ($p<0.01$) decrease at the 25 ppm dose level.

Finally, we note that the three page Executive Summary Tables make no mention of the CFU data, although these data were deemed by the Organizing Committee (along with the anti-CD3 data) to be the most important data from this study.

B. Anti-CD3 stimulation data (Tables 13-15)

We agree in principle with the authors' conclusion regarding these data. However, the Executive Summary and associated tables do not note the following effects.

(i) For the F0 females, the authors correctly report in their summary table that the 77% increase in unstimulated cultures found in the 1250 ppm animals was significant ($p<0.01$). However, they do not mention that the 36% increase seen at the 250 ppm group was also significant ($p<0.05$).

(ii) For the F1 males, the authors correctly report that there was a significant ($p<0.01$) increase in stimulated cultures in the 1250 ppm group, although they mis-report the magnitude of the increased response, which based on the summary statistics in their Table 13 was 59%, not 68% as is reported in their Executive Summary Table. Moreover, the Executive Summary Table makes no mention of the highly significant ($p<0.01$) increases in this variable also observed in the 250 ppm (47% increase) and 25 ppm (57% increase) groups.

Similarly, the authors' Executive Summary table correctly notes the significant ($p<0.01$) 33% increase in total cells in the 1250 ppm group, but makes no mention of the similar significant ($p<0.05$) increases observed in the 250 ppm (28% increase) and 25 ppm (29% increase) groups.

(iii) For the F1 females, the Executive Summary table correctly notes the significant ($p<0.01$) 41% increase in stimulated cultures observed in the 1250 ppm group. However, it makes no mention of the significant ($p<0.05$) increases seen in the 250 ppm (33% increase) and 25 ppm (26% increase) groups.

(iv) Although it made no difference to the final conclusions, the raw data given to us had only six (not seven) control values for spleen cell cultures unstimulated for the F1 males, and the resulting mean was slightly lower than the value reported by the authors in Table 13.

Comment

These data were received on September 9, well beyond the second requested submission date. Because of the high priority assigned to these data by the Organizing Committee and the lateness of the submission, the Chairman of the Statistics Subpanel decided to analyze these data himself. This (and the two companion Germolec studies of methoxychlor and nonylphenol discussed below) is the only dataset for which there could be a perceived conflict of interest, in that both Drs. Germolec and Haseman are NIEHS scientists. However, Dr. Haseman had no prior involvement with this study and did not provide the original statistical analysis.

5. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of methoxychlor in male and female Sprague Dawley rats." (Unpublished Final Report).

We were provided with the Executive Summary. More detailed written reports for this study were later received from the authors.

Raw data provided: Bone marrow cell number and colony forming unit (CFU) data and anti-CD3 data for male and female Sprague Dawley rats receiving dietary exposure of 0, 10, 100 or 1000 ppm methoxychlor for up to 82 days.

Comments on statistical methodology:

(1) We agree with the statistical methods used, except that Dunnett's test does not require an overall ANOVA to be significant.

(2) The Executive Summary tables list only the "maximum effect" (and corresponding dose) for each variable, which could potentially be misleading, because it does not distinguish situations in which all three doses produce a significant effect and those for which only a single dose produces a significant effect.

Results

A. Bone marrow CFU data (Tables 3 and 4)

We agree with the authors' conclusion that "A statistically significant increase was observed [for F1 males] with the CFU-GM/1 x 10^5 cells at the 100 ppm and 1000 ppm doses of 18% and 22%.

In F1 generation female rat study, a significant decrease was observed in CFU-M/1 x 10^5 at the 1000 ppm dose of 23%."

However, there are several significant ($p<0.01$) effects related to the lowest (10 ppm) dose in F1 males that are not mentioned in the Executive Summary:

(i) The 10 ppm group produced a significant (38%; $p<0.01$) reduction in cells/femur relative to controls; and

(ii) The 10 ppm dose produced a significant (43%; $p<0.01$) decrease in CFU-GM/femur and in CFU-M/femur relative to controls.

These significant low dose effects were not seen in the F1 females.

We had one minor numerical disagreement: For male F1 CFU-E/femur controls, the raw data provided to us had ten animals rather than nine, as given in the authors' Table 3. We found the mean control response to be 9.45, not 9.7, and both the 10 and 1000 ppm reductions were significant ($p<0.05$). The authors reported no pairwise significance for this variable, although they noted a significant ($p<0.05$) trend.

B. Anti-CD3 stimulation data (Tables 14-16)

We agree with the authors' interpretation of these data as given in the Executive Summary and in Tables 14-16. Perhaps the most notable effect is the significant ($p<0.01$) increase in unstimulated spleen cell cultures and significant ($p<0.05$ or $p<0.01$) increases in CD3 stimulated spleen cell cultures observed in F1 males that were significant at all three dose levels, including the lowest dose level of 10 ppm. The significant low dose (10 ppm) effects were not seen in the F1 females.

Table 15 mis-reports the number of animals in the 10 and 100 ppm F1 male groups. There were ten animals in these groups, not nine.

Comment

These data were received on September 9, well beyond the second requested submission date. Because of the high priority assigned to these data by the Organizing Committee and the lateness of the submission, the Chairman of the Statistics Subpanel decided to analyze these data himself. This (and the two companion Germolec studies) is the only dataset for which there could be a perceived conflict of interest, in that both Drs. Germolec and Haseman are NIEHS scientists. However, Dr. Haseman had no prior involvement with this study and did not provide the original statistical analysis.

6. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of nonylphenol in male and female Sprague Dawley rats." (Unpublished Final Report).

We were provided with the Executive Summary. More detailed written reports for this study

were later received from the authors.

Raw data provided: Bone marrow cell number and colony forming unit (CFU) data and anti-CD3 data for male and female Sprague Dawley rats receiving dietary exposure of 0, 25, 500 or 2000 ppm nonylphenol for 77 days.

Comments on statistical methodology:

(1) We agree with the statistical methods used, except that Dunnett's test does not require an overall ANOVA to be significant.

(2) The Executive Summary tables list only the "maximum effect" (and corresponding dose) for each variable, which could potentially be misleading, because it does not distinguish situations in which all three doses produce a significant effect and those for which only a single dose produces a significant effect.

Results

A. Bone marrow CFU data (Tables 4 and 5)

We agree with the authors' conclusion that "A statistically significant decrease of 29% in CFU-E/2 x 10^5 cells was seen in the F1 generation male rats at 2000 ppm. With the F1 generation female rats, a significant increase in the CFU-GM/1 x 10^5 cells of 25% was observed at the high dose along with an increase of 34% in the DNA synthesis." We also agree that no other CFU variables showed significant pairwise differences with the following exceptions (correctly noted in the authors' Tables 4 and 5 and Executive Summary tables, but not reported in the Executive Summary text):

(i) a significant ($p<0.05$) 18% decrease in DNA synthesis in the 25 ppm F1 males; and

(ii) a significant ($p<0.05$) 11% decrease in CFU-M/1 cells in 500 ppm F1 females.

The biological judgement is made by the authors (see Executive Summary Table) that neither of these two significant decreases is biologically meaningful. The Statistics Subpanel defers to expert scientific judgement on this matter.

B. Anti-CD3 stimulation data (Tables 15-17)

We agree in general with the authors' interpretation of these data as given in the Executive Summary and in Tables 15-17. Perhaps the most notable effect is the significant ($p<0.05$ or $p<0.01$) increase in unstimulated spleen cell cultures and significant ($p<0.05$ or $p<0.01$) increases in CD3 stimulated spleen cell cultures observed in F1 females that were significant at all three dose levels, including the lowest dose level of 25 ppm. In fact, the increases in the 25 ppm group (64% and 27% respectively) were even greater than the increases seen at the 2000 ppm group (57% and 23% respectively). The increases were greatest in the mid dose (500 ppm) group:

92% and 37%.

F1 males also showed significant increases in these two variables in the 500 and 2000 ppm groups, but there was no evidence of an increase in the 25 ppm group.

One minor numerical discrepancy: Table 16 reported 10 control animals for unstimulated spleen cultures. However, only nine control values were provided in the raw data. This had little effect on the reported statistical results.

One minor statistical discrepancy: we found the 36% increase in total cells for the F1 females in the 500 ppm group to be significant (p=0.04), while the authors report no significance for this dose. The reason for this difference is that the authors required a significant overall ANOVA before applying Dunnett's test (which is unnecessary as noted above), and the overall ANOVA p value was p=0.06. It is debatable if this marginal increase in the mid dose group is biologically important in any case.

Comment

These data were received on September 9, well beyond the second requested submission date. Because of the high priority assigned to these data by the Organizing Committee and the lateness of the submission, the Chairman of the Statistics Subpanel decided to analyze these data himself. This (and the two companion Germolec studies) is the only dataset for which there could be a perceived conflict of interest, in that both Drs. Germolec and Haseman are NIEHS scientists. However, Dr. Haseman had no prior involvement with this study and did not provide the original statistical analysis.

7. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats." In prep (Unpublished Abstract).

Raw data provided: In this multigenerational study, male rats were dietarily exposed to 0, 25, 200 or 750 ppm nonylphenol. The primary variable of interest was neonatal serum testosterone.

Statistical Methodology

Polynomial regression models were separately applied to the three generations (F1, F2, F3). Litter membership information was only available for the F1 generation. In addition, mixed effects anova were used to determine which concentrations differed from the control group. Therefore, mixed effects models were considered for F1 while ordinary multiple regression models were considered for F2 and F3.

Results:

(1) While a quadratic effect is suggested in generation F1, it appears that no dose-related changes are observed in either generation F2 or generation F3 (figure not included).

(2) Generation 1 modeling: A model with linear and quadratic dose effects was a significant improvement over a model with only linear dose effects. In this model, both linear and quadratic effects were significant (P=0.014 for the linear term; P=0.29 for the quadratic term). A categorical analysis did not provide significantly better fit (X2=5.0, df=3, P=0.17). In the categorical data analysis, Dunnett's procedure was used to compare the dose groups with the controls (via SAS PROC MIXED LSMEANS statement specifying all differences relative to controls). All dosed groups differed from controls: dose=25 ppm (P=.06), dose=200 ppm (P=.007) and dose=750 ppm (P=.013). Litter was a important source of variability accounting for over 50% of the total variability.

(3) Generation 2 and 3 modeling: No effects of dose on neonatal testosterone were observed in either generation 2 or generation 3.

Commentary/Limitations:

There was no way to cross reference data with published results since only abstract information was provided to summarize these studies

8. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression." In prep (Unpublished Abstract).

Chemicals: Nonylphenol, Genistein, Ethinyl Estradiol

Raw data provided: In this study, female rats were dietarily exposed to ethinyl estradiol (0, 1, 25, 200 ppm), genistein (0, 25, 250, 1250 ppm) or nonylphenol (0, 25, 500, 2000 ppm). The primary variables of interest were hepatic testosterone metabolism, CYP450 expression, and ERa levels in liver following exposure.

Statistical Methodology

Regression models were applied using the hepatic testosterone metabolism variables, CYP450 expression, and ERa levels in liver following exposure as the response variables. Initially, simple linear regression models were employed using chemical concentrations as the predictor variable. Residual analyses suggested some heterogeneity in the responses. Log10 transformations were employed to address this heterogeneity. Linear regression models were then applied to the log-transformed responses. In addition, one-way ANOVA models were applied to the log-transformed responses. All comparisons of the exposed groups to the control groups were conducted using Dunnett's procedure.

Results:

Ethinyl Estradiol

1. The one-way ANOVA models suggested that no concentration-related differences were present for testosterone (F=2.75, P=.09), or DHT Diol. (F=1.42, P=.286). The remaining

responses all exhibited significant patterns in which at least one concentration condition differed from the other conditions. For andro. (F=4.09, P=.03), DHT (F=10.03, P=.0014), 2a (F=5.22, P=.016), 7a (F=13.28, P=.0004), 16a (F=17.35, P<.0001), and ERa (F=17.70, P<.0001). For the responses where differences where observed the following concentrations differed from controls: for DHT, 1 ppm; for 2a, 200 ppm; for 7a, 25 ppm and 200 ppm; for 16a, 200 ppm and for ERa, 1 ppm, 25 ppm and 200 ppm differed.

2. Trends were evaluated in these responses using the linear regression with the log-transformed responses. Significant negative trends were observed for andro. (b1=-.002, P=.03), 2a (b1=-.001, P=.003), 7a (b1=-.002, P=.001), and 16a (b1=-.003, P<.001), while significant positive trends were observed for DHT (b1=.002, P=.004)and ERa (b1=.002, P=.02).

Genistein

1. The one-way ANOVA models suggested that no concentration-related differences were present for 6b (F=1.91, P=.18). The remaining responses all exhibited significant patterns in which at least one concentration condition differed from the other conditions. For DHT/ DHT. diol. (F=3.64, P=.04), 2a (F=6.03, P=.01), 7a (F=12.06, P=.001), 16a (F=11.18, P=.001), and ERa (F=6.83, P=.007). For the responses where differences where observed the following concentrations differed from controls: for DHT, 250 ppm; for 2a, 1250 ppm; for 7a, 250 ppm; for 16a, 1250 ppm and for ERa, 1250 ppm differed.

2. Trends were evaluated in these responses using the linear regression with the log-transformed responses. Significant negative trends were observed for 2a (b1=-.0004, P=.002), 7a (b1=-.0003, P=.038), and 16a (b1=-.0005, P<.0001) and ERa (b1=-.0001, P=.001).

Nonylphenol

1. The one-way ANOVA models suggested that no concentration-related differences were present for andro. (F=.74, P=.55). The remaining responses all exhibited significant patterns in which at least one concentration condition differed from the other conditions. For testosterone (F=10.34, P=.0012), DHT (F=13.04, P=.0004), DHT. diol. (F=8.04, P=.0033), 2a (F=7.66, P=.0004), 7a (F=24.18, P<.001), 16a (F=19.82, P<.001), and ERa (F=7.51, P=.0043). The 2000 ppm was the only condition significantly different than the controls for DHT (>controls), 2a (<controls) and 16a (<controls). All tested doses were significantly greater than the controls for testosterone and DHTdiol. while only the 25 ppm and 500 ppm groups were significantly greater than the controls for 7a.

2. Trends were evaluated in these responses using the linear regression with the log-transformed responses. Significant negative trends were observed for 2a (b1=-.0001, P=.001), and 16a (b1=-.0003, P<.0001) while significant positive trends were observed for DHT (b1=.0004, P<.001) and testosterone (b1=.0001, P=.008).

Commentary/Limitations:

There was no way to cross reference data with published results since only abstract information

was provided to summarize these studies.

9. Meredith, J. M., C. Bennett, et al. (2000). "Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats." Society for Neuroscience Abstracts (in press). 10. Scallet, A. C., C. Bennett, et al. (1999). "Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure." Society for Neuroscience Abstracts 25: 227.

Raw data provided: the following chemicals and doses were used:

Ethinyl estradiol/ 0, 0.1, 1, 5, 25, 100, 200 ppb
Genistein, 0, 5, 25, 100, 250, 625, 1250 ppm
Vinclozolin, 0, 2, 10, 50, 150, 300, 750
Nonylphenol , 0, 5, 25, 200, 500, 1000, 2000 ppm

Each chemical was applied at various doses to male and female 50 PND F1 Sprague Dawley rats. Eight pups per dam were kept with equal numbers of males and females. Dam information was not available. Treatment was presented in food to dams from GD 7 through birth and then to pups through PND 50. Paper is a little vague about how this was done since there is indication that pups were divided into three separate groups of 70 animals each (35M and 35F) then put on dosed chow. No discussion was provided of nonylphenol rates or application.

Comments on Statistical Methodology

Manuscript indicated that at ANOVA based tests were used to test for gender and dose main and interaction effects for the response SDNPOA (sexually dimorphic nucleus of the medial preoptic area) for each chemical. Student-Newman-Keuls post hoc multiple comparison procedures were used to separate gender and dose mean effects. One factor General Linear Models were fit to each combination of chemical and gender separately for the response SDNPOA, and the natural log of SDNPOA to identify gender by dose response differences.

Reanalysis used both one and two way ANOVA approaches as discussed above but with Dunnett's, Fisher LSD and the Waller-Duncan procedures additionally computed to examine dose effect differences, and linear and quadratic regressions of dose on response by gender to gain some idea of the nature of the trend.

Results:

(1) Vinclozolin: Authors identified strong gender effects that were consistent across all doses (i.e. no interaction). Reanalysis confirms this. Conclusion is vinclozolin has no effect.

(2) Ethinyl Estradiol: Authors identified no effect in females bur report a significant dose response for males. The general trend is to decrease volume of SDNPOA with increasing dose, with the exception of a small increase for the 0.1 dose. In comparing gender by dose means, author noted no gender differences at the 100 dose level even though male volume was over 8x female volume. In reanalysis, the dose response for males is confirmed but shown not to be very

large. Dunnett's test shows all effects not significantly different from the zero dose level but other MCPs show some difference between the low-dose treatments and the high-dose treatments. Regressions for transformed and untransformed response are significant, with a quadratic component for transformed response. Means suggest a decrease in SDNPOA with increasing dose up to a point at which a smaller response change with increasing dose is observed.

(3) Genistein: Authors identified clear gender dimorphism for this chemical, with no female dose response and a non-linear dose response for males (significant gender by dose interactions). In particular, they noted that the middle range of doses produced significantly smaller SDNPOA responses that either the zero dose or the higher doses. Reanalysis indicated strong gender by dose interactions with response but non-significant interactions with log transformed response. Multiple comparison procedures, including Dunnett's recreated the response pattern suggested by the author.

(4) Nonylphenol: Authors did not discuss this chemical in the paper provided. Reanalysis indicated strong gender, dose and gender by dose interactions. The interaction significance was lost when examining log response. As before, female response was overall 8X less than male response. Dunnett's test suggested that the 200 dose response was significantly different from doses below or above. Other multiple comparisons tend to blur the significance by suggesting that dose neighbors to 200 are not significantly different.

(5) Other Analyses: Normal quantile plots of residuals from all analysis models were examined for obvious non-normality and outliers. Residuals from male data were very normal while data from female analyses were less normal. This may have been due to the bias caused with fitting non-significant factors in the model. No outliers were identified.

Commentary:

The results reported in Meredith et. al were essentially reproduced.

O'CONNOR DATASETS

1. Biegel, L. B., J. C. Cook, et al. (1998). "Effects of 17β-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats." Toxicological Sciences 44: 143-154.

Raw data provided: Rats were exposed to dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm 17beta estradiol. Endpoints of interest included

(i) serum hormone concentrations from P1 female rats at the one week (Table 1), 28 days (Table 2) or 90 days (Table 3)

(ii) serum hormone concentrations from F1 female rats at postnatal day 98 (Table 4).

Comment on Statistical Methodology

The authors use Jonckheere's trend test, a widely used and appropriate nonparametric procedure. However, we have some reservations regarding its use in this particular setting. Perhaps our greatest concern with Jonckheere's test is that it assumes a monotonic (i.e., consistently increasing or consistently decreasing) dose-response trend. While it is not clear if "low-dose effects" are suspected in this particular study, such effects that are associated with a U (or inverted U) shaped dose-response curve might not be detected by Jonckheere's test. Two possible examples of this occurred in this study (see discussion of LH and prolactin below).

Results (from Tables 1-4)

Estradiol

We agree with the authors that serum E2 concentrations were increased in a dose-dependent manner at all time points in rats fed 2.5, 10, and 50 ppm 17beta estradiol. This increase was seen regardless of stage of estrous cycle.

Progesterone

We agree with the authors that serum p4 concentrations were not affected in the 1 day and 28 day groups. We also agree that at the 90 day time point serum p4 concentrations were decreased in a dose-related manner, and were reduced even at 0.05 ppm. These decreases appeared to be essentially independent of stage of estrous cycle. We also agree that F1 female rats showed a similar decrease on Postnatal Day 98 in the 2.5 ppm (but not the 0.05 ppm) dosed group.

LH

For the 1-week and 28 day time points, we agree that the decrease in the 50 ppm group is significant ($p<0.05$), but by Dunnett's test the slight decrease at 10 ppm was not statistically significant. We agree that at 90 days, there is a significant ($p<0.05$) decrease in the top three dosed groups (Table 3).

For the F1 data we may have some disagreement. The authors conclude that "LH concentrations were not affected by 17beta estradiol administration in the F1 generation" (page 147). The data suggest otherwise. The increased response in the low dose (0.05 ppm) group (2.4 vs. 2.1; see Table 4) is even more impressive that the decreases flagged by the authors as significant in Tables 2 (1.9 vs. 1.7) and 3 (2.3 s. 2.1 in two dosed groups). This increase is significant ($p<0.05$) by Dunnett's test.

The most likely explanation for this finding is that Jonckheere's test (which is somewhat insensitive to such a U shaped dose-response curve) was not significant, so the authors concluded that no significant effects occurred in any dosed group. While an increase of this magnitude may or may not be biologically important, it should have been identified as statistically significant by a methodology sensitive to such patterns of response. That is one reason why the Statistics Subpanel has concerns about Jonckheere's test being the sole statistical

methodology for these data.

Note: In response to this concern, the authors stated that "we are testing specifically for a decrease in these endpoints", and thus (paraphrasing the words of the authors) since Jonckheere's test indicated a significant decreasing trend, and the test assumes monotonicity of response, any apparent "significant" increase in LH at a low dose is inconsistent with the assumed pattern of dose-response and thus biologically unimportant.

FSH

We also have some possible disagreement regarding the interpretation of this variable. This is the one variable in which there appears to be a significant interaction between 17beta estradiol and estrous cycle. This can be clearly seen in Tables 2, 3 and 4 and is even suggestive in Table 1. This significant interaction means that the effect of 17beta estradiol is significantly different during some stages of the estrous cycle than during others.

A closer look at these data suggests (especially for 28 and 90 days) that there is a significant ($p<0.05$ by Dunnett's test) decrease in FSH during the estrous cycle in the 10 and 50 ppm groups at 28 days and in the 50 ppm group at 90 days. There was also a slight (but non-significant) decrease in FSH in the top dose group at one week (Table 1) and in the F1 animals (Table 4). The authors note these reductions, but dismiss them.

The Statistics Panel has had little direct experience with this variable and do not know whether it is biologically plausible for high doses of 17beta estradiol to have an inhibitory effect on FSH that is limited to animals in the estrous cycle. However, that is what the data are suggesting.

Prolactin

The Statistics Subpanel may have some disagreement regarding this variable. At one week, the authors report that a very modestly increased response in the high dose group (26.9 vs 22.5) is statistically significant. Here again, this is a result of a non-monotonic, U shaped dose-response curve being evaluated by Jonckheere's test. This "significantly elevated" response in the top dose group is the lowest of the mean responses in any of the dosed groups, even lower than the (apparently non-significant) response of 28.6 in the 0.05 ppm group. By Dunnett's test, only the response in the 10 ppm group is significantly elevated.

Note: In response to this concern the authors stated that in their experience the procedure used to collect blood "introduces a great deal of stress in the animals and will result in higher variability in the prolactin data. Unfortunately, this was not discussed in the current manuscript." They conclude that "the interpretation of the prolactin data is likely confounded by increased variability." It is unclear whether or not this interpretation of the data was given in the manuscript.

We agree with the authors that in the 28 and 90 day top dosed group, prolactin is significantly ($p<0.05$) increased. We also agree that there is no significant effect in the F1 group (Table 4).

2. Biegel, L. B., J. A. Flaws, et al. (1998). "90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol." Toxicological Sciences 44: 116-142.

Raw data provided: Rats were exposed to dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm 17beta estradiol. Endpoints of interest included

(i) body weights (Table 5) and epididymides, accessory gland and testes weights (Table 6) for P1 adult male rats.

(ii) body weights (Table 5) and uterus and ovary weights (Table 7) for P1 adult female rats.

(iii) ovaries and uterus histopathology for P1 adult female rats (Table 9)

(iv) mammary gland differentiation data (mammary gland labeling indices) for P1 and F1 female rats (Table 11)

(v) uterine marker data (uterine stromal cell and epithelial cell proliferation labeling indices for P1 and F1 female rats (Table 12)

(vi) reproductive indices (mating index; fertility index) for P1 generation (Table 13).

(vii) anogenital distance, preputial separation,and vaginal opening data for F1 animals (Table 17)

(viii) body weights (Table 19) and epididymides, accessory gland and testes weights (Table 20) for F1 adult male rats.

(ix) body weights (Table 19) and uterus and ovary weights (Table 21) for F1 adult female rats.

Comments on Statistical Methodology

The authors' Statistical Methodology Section is quite extensive, and we agree in general with the methods used. We have the following comments:

(1) Scheffe's test is a rather conservative multiple comparisons procedure. It is unclear why the authors chose this particular method for ovarian follicle number, while choosing alternative procedures for other variables of interest.

(2) Dunnett's test does not require the significance of an overall ANOVA.

Note: In response to this concern the authors agreed and stated that there was a "mis-statement in the paper" and that Dunnett's test was in fact carried out independently of an ANOVA. We agree with this approach.

(3) The authors state that "Incidences of clinical observations were evaluated by the Fisher's exact test with a Bonferroni correction, and, when significant, the Cochran-Armitage trend test." This seems to mean that the trend test was carried out only when the pairwise comparisons were

significant, which is not the usual approach. Perhaps the wording is just unclear. Also, we see no reason to use the Bonferroni correction for the Fisher's exact test, especially if the Cochran-Armitage trend test is used to control the experiment-wide error rate. While there is nothing "wrong" with using a Bonferroni adjustment, it results in a rather conservative test.

Note: In response to this concern the authors replied: "only the analysis from the Cochran-Armitage test was included in the manuscript. Therefore, the text in the manuscript was not correct."

Results

Table 5

We agree with the authors that for both males and females, the top three 17beta-estradiol doses significantly ($p<0.05$) reduced body weight.

Table 6

The only biologically important organ weight effects (by ANCOVA) appear to be markedly reduced testis, epididymides, and accessory sex gland weights in the 50 ppm dosed group. Additionally, the reduced accessory sex gland weight in the 10 ppm dosed group is significant ($p<0.05$). Organ and body weights were significantly ($p<0.05$) correlated. These conclusions are consistent with those of the authors.

Table 7

Ovary, but not uterus, weight is significantly ($p<0.05$) correlated with body weight (within groups). We agree with the authors that the top two doses (10 and 50 ppm) increase uterus weight and decrease ovary weight, although the statistical significance obtained by Dunnett's test for the top dose uterus weight was $p=0.06$. The reduced ovary weight in the 2.5 ppm group (following ANCOVA and Dunnett's test) is also borderline: $p=0.05$.

There is a minor typographical error in this table: the correct SD for 10 ppm uterus weight should be 0.526, not 0.256.

Table 9

We agree with the authors that epithelial hypertrophy of the uterus is significantly ($p<0.01$) increased in the 10 and 50 ppm groups. The authors fail to mention that a similar increase in uterine hypertrophy (9/9 vs. 1/9) was also observed (and is also significant) in the lowest dose (0.05 ppm) group. Ovarian atrophy is also significantly ($p<0.01$) increased in the 10 and 50 ppm dosed groups.

Note: In response to this concern, the authors stated that the 9/9 response that they reported in the manuscript for the 0.05 ppm group was a typographical error, and that the correct response was 0/9.

Table 11

We agree with the authors' interpretation of these data: generally negative findings with two marginally significant ($p<0.05$) decreases in mammary gland labeling indices the middle two dosed group for P1 female rats at one week. Because of the lack of dose-response and the lack of consistency of this effect at other time points, we agree that this could be a spurious finding.

Table 12

We agree with the authors' finding that the only statistically significant ($p<0.05$) effects were decreases in uterine epithelial cell proliferation labeling indices for 10 ppm and 50 ppm P1 groups at 28 days. We note another typographical error: the mean response for the P1 90 day top dose group for this variable should be 4.13, not 413. The SD for this group should be 1.957, not .957.

Table 17

We agree with the authors' finding that (i) anogenital distance in both males and females is unaffected by either 0.05 or 2.5 ppm 17beta estradiol; (ii) the 2.5 dose significantly ($p<0.05$) increases the day of preputial separation in males; and (iii) both doses significantly ($p<0.05$) reduce the day of vaginal opening in females.

Table 19

We agree with the authors that 2.5 ppm (but not 0.05 ppm) 17beta estradiol significantly ($p<0.05$) reduces body weight in F1 adult males and females.

Table 20

The decreased accessory sex gland, testis, and epididymides weights in 2.5 ppm 17beta estradiol F1 adult male rats were all secondary to reduced body weight, and were all not significant when evaluated by ANCOVA. There is another typographical error: the mean accessory gland weight for top dose males should be 1.66, not 0.660.

Table 21

We agree with the authors that 2.5 and 0.05 ppm 17beta estradiol had no effect on uterus weight in F1 adult female rats. The ovary weight effect is more interesting.

The authors correctly report that by Dunn's test (a valid nonparametric multiple comparisons procedure) that absolute ovary weight is significantly reduced in the top dose group relative to controls despite the similarity of means (0.141 +- 0.020 vs. 0.136 +- 0.183). However, they incorrectly report on page 129 that the ovary/body weight ratio was significantly increased in this group (0.052 +- 0.006 vs. 0.062 +- 0.078). Actually, despite the increase in the mean response in the high dose group, the statistical significance from Dunn's test is for a significant DECREASE, not an increase. How can this occur?

What these investigators failed to recognize (which was strongly suggested by the large SD in the dosed group) was that the dosed group contained a single ovary weight (0.879 g.) that was approximately ten times the value of the group mean. We strongly suspect a decimal point error in this value, but the investigators were unable to confirm this. As a result, although the mean response was indeed slightly elevated in the dosed group relative to controls, the preponderance of the individual animal data showed the opposite trend, and Dunn's test (which was apparently carried out correctly) actually identified a statistically significant decrease, not a significant increase, in the ovary/body weight ratio in the dosed group.

3. Cook, J. C., L. Johnson, et al. (1998). "Effects of dietary 17β-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats." Toxicological Sciences 44: 155-168.

Raw data provided: In a 90/day/one-generation study, male and female rats were exposed to dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm 17beta estradiol. Endpoints of interest included

(i) Testis and epididymis weights (Table 2)

(ii) Hormone levels of P1 male rats (Table 3)

(iii) Hormone levels of F1 male rats (Table 4)

(iv) Hormone levels of P1 sperm parameters (Table 5)

(v) Hormone levels of F1 sperm parameters (Table 6)

Comments on Statistical Methodology

The authors' Statistical Methodology Section is quite extensive, and we agree in general with the methods used. We have the following comments:

(1) Dunnett's test does not require the significance of an overall ANOVA.

(2)The authors use Jonckheere's trend test, a widely used and appropriate nonparametric procedure. However, we have some reservations regarding its use in this particular setting, since it assumes a monotonic (i.e., consistently increasing or consistently decreasing) dose-response trend. While it is not clear if "low-dose effects" are suspected in this particular study, a U (or inverted U) shaped dose-response curve would likely not be detected as significant by Jonckheere's test (especially if the test is applied in a "step-down" fashion) with significant low dose effects being missed.

Table 2

(1) Since body weights were so markedly reduced for these studies, we feel that in Table 2 the authors should have presented body weight data and adjusted the organ weights for body weight

differences in their assessment of organ weight changes.

(2) For the P1 group, there were marked decreases in body weight in the 2.5, 10, and especially the 50 ppm groups. In the 50 ppm dose group, the animals were on average approximately half the size of the controls. This made assessments of changes in organ weight more difficult.

Nevertheless, there was a 4-fold reduction in mean testis weight and a 5-6 fold reduction in mean epididymis weight in the top dose group that could not be explained by body weight changes. Similarly, the 10 ppm group showed a significantly ($p<0.05$) reduced epididymis weight after adjusting for body weight differences. Both testis weight and epididymis weight were significantly ($p<0.01$) correlated with body weight in this study.

(3) For the P1 Recovery group, body weights were still significantly ($p<0.05$) reduced in the 2.5, 10, and 50 ppm groups, but the reductions were much less than that seen in the P1 animals. For example, in the top dose group, the reduction in body weight was 17% rather than 50%.

We agree with the authors that none of the testis and epididymis weights in the dosed groups were significantly different from controls. Moreover, there was not a significant ($p<0.05$) correlation between organ and body weight in this study.

(c) For the F1 group there were significantly ($p<0.05$) reduced body weights in the 0.05 ppm (8% reduced) and in the 2.5 ppm (28% reduced) groups. Although epididymis weights were not significantly affected after adjustment for body weight differences, the increased testis weight in the 0.05 ppm group was significant ($p<0.05$) by ANCOVA/Dunnett's test. That is, the 9% increase in testis weight, coupled with an 8% decrease in body weight, resulted in a significantly elevated (adjusted) testis weight in the 0.05 ppm group. Analysis of the testis/body weight ratio would have produced similar results for this dosed group.

(4) For the F1 recovery group, body weight was still significantly (14%, $p<0.05$) reduced in the 2.5 ppm group, but there were no significant effects on testis or epididymis weight, after adjustment for body weight differences.

Tables 3-6

We are in agreement with the authors' interpretation of these data.

4. O'Connor, J. C., S. R. Frame, et al. (1998). "Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17β-estradiol." Toxicological Sciences 44: 169-184.

Raw data provided: Male and ovariectomized female Crl:CD BR rats received intraperitoneal injections of 0.0, 1.0, 2.5, 7.5, or 50 ug/kg/day 17-beta estradiol. The following endpoints were evaluated

(i) Uterine weights (Table 1);

(ii) Uterine stromal cell proliferation and epithelial cell height (Table 2);

(iii) Estrogen receptor concentrations (Table 3);

(iv) Serum hormone concentrations in females (Table 4);

(v) Reproductive organ weights in males (Table 5);and

(vi) Reproductive hormone concentrations in males (Table 6)

Comments on Statistical Methodology

(1) Dunnett's test does not require the significance of an overall ANOVA.

(2) The authors use Jonckheere's trend test, a widely used and appropriate nonparametric procedure. However, we have some reservations regarding its use in this particular setting, since it assumes a monotonic (i.e., consistently increasing or consistently decreasing) dose-response trend. While it is not clear if "low-dose effects" are suspected in this particular study, a U (or inverted U) shaped dose-response curve would likely not be detected as significant by Jonckheere's test (especially if the test is applied in a "step-down" fashion) with significant low dose effects being missed. Two possible examples of this occurred in this study (see discussion of follicle stimulating hormone and luteinizing hormone in Table 4 below).

Results

Table 1

We have a single, but important disagreement with this table. The increased uterine weight in the low dose (1 ug/kg/day) group is highly significant ($p<0.01$). It is significant by Dunnett's test (the method of analysis the authors used) with or without adjusting for body weight and with or without a logarithmic transformation applied to the data. This low dose increase in uterine weight is also highly significant ($p<0.01$) by a Mann-Whitney U test (the nonparametric method of statistical analysis the authors use for other data).

Note: The authors responded that "This was an oversight in the preparation of the manuscript. The authors agree that the uterine weight for the 1.0 mg/kg/day [sic] group was significant."

Table 2

We agree with the authors' interpretation of these data.

Table 3

We agree in general with the authors' interpretation, but we note that the relatively high estrogen receptor response in the control group was due primarily to a single extreme value (355.4) that was nearly 7 times greater than the average of the other five control values (Note the large

control SE). With this extreme value excluded, the high dose effect looks far less impressive. In fact, with this value excluded, the increase in the lower two dosed groups (145 or 151 vs. 52) becomes more impressive than the decrease in the top two dosed groups (35 or 32 vs. 52). Even with the extreme control value included, the pairwise comparison of high dose vs. control (by either Dunnett's test or a Mann-Whitney U test) is not significant. Thus, this decreased response may be less impressive that might first appear. The significant trend that the authors' analysis detects is due primarily to the high responses in the two lowest dosed groups vs. the low responses in the two highest dosed groups. The control response is intermediate between these two extremes.

Note: The authors' response: "The authors acknowledge that the interpretation of the data could have been confounded by the variability of the data.....the authors feel very strongly that these data suffer from a large amount of variability and are of 'poor quality' to allow for fair evaluation."

It is unclear whether or not this interpretation of the data was given in the manuscript.

Table 4

We agree with the authors' interpretation of the estradiol and prolactin data; however, we may have some disagreement with the interpretation of the hormone concentration data.

For both follicle stimulating hormone (fsh) and luteinizing hormone (lh), the dose-response curve appears to be non-monotonic (and U-shaped), and thus Jonckheere's test is insensitive to any low dose effects. For fsh the increased response in the lowest dose group (from 116.1 to 153.9) is of greater magnitude than the decreased response in the top dosed group (from 116.1 to 89.2), a decrease that the authors regard as significant. By a Mann-Whitney U test (the nonparametric pairwise comparison method preferred by these authors) for these variables, the low dose increase is even more significant ($p<0.01$) than the high dose decrease. A similar significant ($p<0.05$) lowest dose increase occurs for lh. Dunnett's test produces similar results.

Thus, there appears to be a significant ($p<0.05$) increase in both fsh and lh for the 1 ug/kg/day 17beta estradiol group in this study. Whether or not these increases are biologically important is a matter beyond the scope of our analysis.

Note: The authors' response: "The analysis that was performed on the data was Jonckheere's trend test. The statistics that were performed did not 'flag' this data as statistically significant (as expected since it is looking for a monotonic dose-response)."

See our comment regarding the use of Jonckheere's test above.

Table 5

Our analysis confirms the significant decreases in organ weights observed in the higher dosed groups and reported by the authors in this table.

Table 6

We agree with the authors' interpretation of these data.

VOM SAAL DATASETS

1. Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October.

Raw data provided: uterus weight data for female CD-1 mice receiving fetal exposure to 0.1 or 1000 ug/kg/day diethylstilbestrol (DES) or to 10 or 10,000 ug/kg/day methoxychlor (MXC). At seven months of age each female was ovariectomized and implanted with a Silastic capsule containing 0.5 ug 17beta-estradiol.

In one experiment MXC or DES were given in conjunction with adult estradiol capsule doses of either 0, 0.25, 0.5 or 1.0 ug.

These data are summarized in Figures 1-6.

Comments on Statistical Methodology

The authors appear to have used a Protected Fisher's LSD analysis to make pairwise comparisons, although they describe their procedure slightly differently. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Results

Figure 1 (Experiment 1)

(1) Since the two control groups (feed controls for MXC; s.c. injection controls for DES) showed no significant differences, they were pooled in our analysis.

(2) We agree with the study authors that body weights were significantly ($p<0.01$) elevated in the top dose DES group relative to controls.

(3) We also agree that uterine weights were significantly ($p<0.01$) reduced in the top dose DES group relative to controls.

(4) We do not agree that body weight is significantly correlated with uterine weight in this study. In fact, the significant correlation found by the authors is a negative correlation as discussed in the main portion of this report (see Figure 3), which is in the opposite direction of the association found by most other investigators (see discussion below). After adjusting for the treatment effect (i.e., the increase in body weight and the independent decrease in uterine weight produced by

DES), there is no significant correlation between uterus weight and body weight with the DES and control groups. Thus, there is no need to adjust for body weight when evaluating uterus weight for these data.

(5) For the Figure 1 data, the low dose DES group was 5% heavier on average than the control group, and the uterine weights were on average 19% greater than controls. Commonly used methods of adjustment such as the organ/body weight ratio would "weaken" the difference in organ weights between the two groups, since the impact of adjusting for the heavier body weights in animals with heavier uterus weights would be to reduce the difference in the (now adjusted) uterine weight.

However, in this example, the authors' use of ANCOVA actually increases the difference in uterine weight between the low dose DES and control groups, since the negative correlation would imply that heavier animals in general tend to have smaller uteri. However, as noted above (and illustrated in Figure 3), this negative correlation (within groups) is not statistically significant in any case in this particular experiment (but see discussion below for other data). Thus, our analysis did not adjust for body weight and did not find a significant increase in uterus weight in the low dose DES group.

(6) The raw data indicate that the control groups were necropsied on a different date than the DES animals. Ideally, these groups should be necropsied at the same time, as noted in the body of this report.

Figure 2 (Experiment 1)

We agree with the study authors that MXC did not significantly alter body weight or uterus weight relative to controls. Importantly, the $p<0.05$ result given in Figure 2 is a comparison of the high dose and low dose groups, not a comparison of the high dose and control groups. We note that the MXC and control animals were apparently necropsied on the same day in this study.

Figure 5 (Experiment 2)

(1) We agree with the study authors that the 100 mg/kg dose of DES significantly ($p<0.01$) increased body weight, while the low DES dose had no effect on body weight. We further agree that estradiol had no effect on body weight. There were also significant ($p<0.01$) litter effects for body weight.

(2) We agree with the study authors that the high dose of DES produced a significant reduction in uterine weight, and that the magnitude of this reduction increased with increasing doses of estradiol (i.e., there was a significant estradiol x DES interaction for this group).

(3) The effect of the low dose of DES is not as clear. The authors report that the increase in mean uterine weight response seen in the low dose DES group from zero to 1 ug estradiol (from 72 to 357 mg) was significantly greater than the corresponding control increase (from 61 to 268 mg). Our summary mean uterine weights for the low dose DES group (from 77 to 359 mg) differ slightly than those reported by the authors (from 72 to 357 mg), which probably reflects

the authors' use of adjusted uterine weights, as discussed below. Importantly, our analysis does not confirm that this difference in increase between low dose DES and control groups is statistically significant. There are several reasons for this.

Figure 5 clearly shows a higher variability in uterine weight response in the the high dose estradiol groups relative to the lower dosed estradiol groups. This significant heterogeneity indicates that a log transformation is needed to equalize the variances, and the corresponding increases in uterine weight response from zero to 1 ug estradiol (4.7 fold increase for the low dose DES group vs. 4.4 fold increase for controls) was not significantly different. That is, the estradiol x DES interaction was not significant, indicating a parallel (on a log basis) dose response curve for the low dose DES and control groups.

Equally important is the adjustment for the highly significant (even within groups) negative correlation between uterus weight and body weight seen in this study. Although we found the apparent negative association between uterine weight and body weight for the Figure 1 data to be not significant, here we agree with the authors that the negative association is statistically significant ($p<0.01$). The authors report the significance of this negative correlation as $p<0.0001$, and it is likely that the uterus weights they report for the low dose DES groups (given above) are adjusted uterus weights. Since the low dose DES groups were on average lighter than the controls, "correcting" for the negative correlation would produce lower adjusted uterus weights, consistent with the mean uterine weight responses reported by the study authors.

Importantly, the difference between the low dose DES and control groups in the change in uterine weight across levels of estradiol (i.e., the DES x estradiol interaction) remains non significant, regardless of whether or not the data are adjusted for body weight. However, the overall difference in uterus weight between the low dose DES and control groups (averaged over estradiol levels, i.e., the "main effect" of low dose DES treatment) is significant at the $p<0.05$ level.

The authors defend their adjustment for the negative correlation between uterus weight and body weight, stating that "it is very dangerous and certain to lead to future heartburn for anyone... to try to decide what is 'right' or 'wrong' in the functioning of biological systems. Data are data and must be approached in a neutral manner. You accept the direction of positive relationship between between 2 variables but not negative. I respectfully strongly disagree with this position."

The "right" and "wrong" statement above referred to a comment in an earlier draft (a comment subsequently removed by us) that the negative correlation between uterus and body weight was a correlation that appeared to be in the "wrong" direction. We agree that "right" and "wrong" are subjective judgements that should not have been used in this context.

However, the primary statistical evaluator for this study has had experience with dozens of other studies evaluating changes in uterus weight (in addition to this Appendix, see Kanno, J., Onyon, L., Haseman, J., Fenner-Crisp, P., Ashby, J. and Owens, W. The OECD Program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase one - submitted for publication). In all of these studies, there was either a significant positive correlation between uterus weight and body weight or no significant correlation. Thus, a

significant negative correlation is unique is his experience.

Time does not permit a more detailed examination of this interesting negative association. It is a fairly pervasive association that is not limited to certain DES/estradiol combinations. Nor is this association due solely to the presence of a few animals with high uterine weights.

Figure 6 (Experiment 2)

There is no significant overall difference in uterine weight between either MXC group and controls over the four different estradiol dosages. Note the high variability in response in the highest dose estradiol group.

4. Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives 105(1): 70-76.

Raw data provided: prostate and body weight data from offspring of pregnant mice fed 2 or 20 ug/kg/day bisphenol A or octylphenol. These data are summarized in Figure 2.

Comments on Statistical Methodology

The statistical methods used by the authors in this study were appropriate for the data.

Results

(1) We agree with the study authors that the low dose of octylphenol produced a significant ($p<0.05$) reduction in body weight. However, a similar reduction in body weight reported as significant ($p<0.05$) by the study authors for the low dose bisphenol A group was only marginally so ($p=0.07$) by Dunnett's test. Neither of the high dose groups showed a significant body weight effect.

(2) We disagree with the study authors that body weight is unrelated to prostate weight. Within the experimental groups, there is a significant ($p<0.05$) correlation as well as evidence of a body weight x treatment group interaction (this interaction due primarily to a single datapoint). Bisphenol A apparently both increases prostate weight and (independently) decreases body weight. Not adjusting for these treatment effects masks the association between prostate weight and body weight. These data illustrate the problem noted in the body of this report about the possible difficulties of ANCOVA in the presence of treatment effects on both body weight and prostate weight. Thus, we feel that body weight should be taken into account in the analysis of prostate weight in this study.

(3) As can be seen in Figure 2, there is also significant ($p<0.05$) heterogeneity in these data, suggesting that a logarithmic transformation is needed to equalize the variances.

(4) There are also minor errors in the error degrees of freedom reported by the study authors in several places in the paper.

(5) The Statistics Subpanel found it interesting that the lower doses of bisphenol A and octylphenol had an impact on body weight while 10-fold higher doses of these compounds did not.

(6) Importantly, despite these concerns, our reanalysis found that after log transforming the uterine weights and adjusting for body weight (which appeared to eliminate the heterogeneity), the elevated prostate weights in the two bisphenol A groups remained significant ($p<0.05$, rather than $p<0.01$) by Dunnett's test, while the slightly elevated prostate weights in the octylphenol groups were not significant. Thus, our conclusions regarding elevated prostate weights are essentially in agreement with the study authors'.

For a comparison of these results with Ashby's attempt to replicate the prostate weight BPA effects, see the final comments for Ashby Study 1 given earlier in the Appendix.

6. Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." (in press).

Raw data provided: Prostate weights and daily sperm production for CF-1 male mice prenatally exposed to 0, 0.002, 0.02, 0.2 or 2 ug/kg 17 alpha-ethinyl estradiol (EE2) at 50 days or 5 months of age. These data are summarized in Tables I and II.

Statistical Methods

Analysis of variance (ANOVA) was carried out using PROC GLM in the Statistical Analysis System (SAS). Because only one animal from each litter was used in the experiment, correction for litter effects was not necessary. Levene's test for variance homogeneity across dose groups was conducted for each response variable. For response variables whose variances appeared heterogeneous across dose groups, logarithmically transformed responses were analyzed, if they showed a more stable variance. If body weight was significantly ($p<0.05$) correlated with a response variable when adjusted for the effect of EE2, then analysis of covariance (ANCOVA) was conducted, with body weight as the covariate. To help in the interpretation of any statistically significant ($p<0.05$) effects of EE2 observed in ANOVA or ANCOVA omnibus tests, both linear and quadratic dose effects were tested using contrasts whose coefficients were determined with the ORPOL function in PROC IML. Two-tailed Dunnett's tests were used to compare each dose group to the control group. For the latter tests, SAS least-squares means rather than unadjusted means were used for the comparisons.

Results

Daily sperm Production (DSP) and DSP/g testis (Tables I and II)

Daily sperm production (DSP) and efficiency (EFF) of DSP measured as DSP/g testis were analyzed at both 2 months (50 days) and 5 months of age. Neither DSP nor EFF was significantly correlated with body weight at 2 months of age. The average DSP response at 2 months was significantly different among EE2 groups ($p<0.01$). There was no statistical evidence of either a linear or a quadratic dose effect, but each of the dosed groups showed

statistically reduced DSP relative to the control group ($p<0.01$ for the 2 ug/kg and 0.2 ug/kg groups, and $p<0.05$ for the 0.02 ug/kg and 0.002 ug/kg groups).

DSP variances at 5 months appeared somewhat erratic, but Levene's test did not indicate significant heterogeneity. There was no apparent relationship with either dose of EE2 or mean DSP response; untransformed and log-transformed responses gave qualitatively similar results. Hence, only results for the untransformed DSP response are reported. DSP was significantly ($p<0.05$) correlated with body weight, but none of the tests showed any significant effect of EE2 on DSP.

The mean EFF response at 2 months was significantly different among EE2 groups ($p<0.001$). There was no statistical evidence of either a linear or a quadratic dose effect, but average EFF was significantly reduced relative to control in all four dose groups ($p<0.01$).

Like DSP, variances of EFF at 5 months were erratic, but not in any apparent dose- or mean-related fashion; Levene's test did not indicate significant heterogeneity among dose groups. Because untransformed and log-transformed responses gave qualitatively similar results, only results for the untransformed EFF response are reported. EFF was marginally ($p<0.06$) correlated with body weight, but none of the statistical tests showed any significant effect of EE2 on EFF, whether or not body weight was included as a covariate.

Prostate weight (Tables I and II)

Mean prostate weights at both 2 months (50 days) and 5 months were analyzed. At 2 months of age, there was no statistical evidence of correlation between prostate weight and body weight. Average prostate weights differed significantly ($p<0.01$) among the treatment groups; there was no statistical evidence of a linear dose effect, but there was evidence of a quadratic dose effect ($p<0.05$). While the 0.002 ug/kg group and the 0.2 ug/kg group were not statistically different from control, the 0.02 ug/kg group was significantly ($p<0.01$) increased relative to control as was the 2 ug/kg group ($p<0.05$).

At 5 months of age, there was no statistical evidence of correlation between prostate weight and body weight. Average prostate weights differed significantly ($p<0.01$) among the treatment groups, but there was no evidence of either a linear or a quadratic dose effect. Significant increases relative to control were observed for the 2 ug/kg group ($p<0.01$) and the 0.2 ug/kg group ($p<0.05$), with less significant increases in the 0.02 ug/kg group ($p<0.06$) and the 0.002 ug/kg group ($p<0.08$).

Commentary

The results reported in Thayer et al. (2000) were essentially reproduced, with only slight (nonessential) variations, probably resulting from the use of slightly different statistical methods. Only the differences between this analysis and the Thayer analysis are now discussed.

1. Dunnett's test was used here for pairwise comparisons to controls instead of the LSD procedure reported in Thayer et al. (2000). Because of this, some comparisons reported

here did not achieve the same level of significance as the LSD results (e.g., DSP at 2 months, prostate weights at 2 and 5 months), although qualitatively the results were the same.

2. Linear and quadratic contrasts were tested here, unlike in Thayer et al. (2000). Hence, the significant quadratic effect ($p<0.05$) on prostate weight at 2 months reported here is not necessarily in disagreement with Thayer et al.

3. This analysis did not find a significant correlation between prostate weight and body weight at 5 months of age, as was reported in Thayer et al. (2000). This was most likely because this analysis adjusted for the effect of EE2 prior to adjusting for body weight, whereas the statistical significance for the covariable reported by Thayer et al. adjusted for body weight first. However, this is an insignificant point, as the demonstrated effects of EE2 are qualitatively similar whether or not body weight is included as a covariate.

4. The data reported by Thayer et al. (2000) for animals 50 days of age are identical to those in the submitted data file for animals 2 months of age. Hence, 2 months is the age discussed here, whereas 50 days is the age discussed by Thayer et al.

9. vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health 14 (1/2): 239-260.

Raw data provided: seminal vesicles, epididymis, testis, and preputial gland weights, and body weight from offspring of pregnant mice fed 2 or 20 ng/g bisphenol A (BPA) or octylphenol (OP). These data are summarized in Table 2. Daily sperm production data were requested, but were not provided until after the meeting. These data are summarized in Table 1.

Comments on Statistical Methodology

Many of the statistical methods used by the authors in this study were appropriate for the data. However, in the application of the LSD test, the authors apparently did not require the overall ANOVA for among-group differences to be significant before carrying out the pairwise comparisons. As noted in the text, the (protected) LSD test is an acceptable method for making pairwise comparisons. However, without the "protection" of an overall ANOVA, the LSD test is prone to false positive outcomes, and these could have occurred in the current study.

Tables 1 and 2 do not indicate if the measure of variability is the standard deviation or the standard error. It appears to be the standard error.

The daily sperm production data were provided without animal numbers, so it was not possible to "link" directly these responses to those of organ weight. Moreover, for four of the five groups, the daily sperm production data were evaluated for only a subset of the animals examined for organ weights, and it was unclear how the subsets were selected. For the fifth group (low dose OP), all animals with organ weights were evaluated for sperm production, and in fact an additional animal was evaluated for sperm production that was not evaluated for organ weight.

Results

A. Organ weights (Table 2)

(1) We agree with the study authors that the low dose of octylphenol produced a significant ($p<0.05$) reduction in body weight. However, a similar reduction in body weight reported as significant ($p<0.05$) by the study authors for the low dose bisphenol A group was only marginally so ($p=0.07$) by Dunnett's test. Neither of the high dose groups showed a significant body weight effect. These are the same body weight data evaluated previously in the Nagel et al. study (Vom Saal Study 4). The Statistics Subpanel found it interesting that the lower doses of bisphenol A and octylphenol had an impact on body weight while 10-fold higher doses of these compounds did not.

(2) The authors report a marginally reduced ($p=0.08$) seminal vesicle weight in the low dose BPA group. By Dunnett's test, this reduction is not significant ($p>0.20$). We agree with the authors that body weight was uncorrelated with seminal vesicle weight in this study.

(3) We agree with the authors that (i) neither BPA nor OP significantly affected testis weight, and (ii) testis and body weight were significantly ($p<0.05$) correlated in this study.

(4) We agree with the authors that there appears to be an association between epididymis weight and body weight. However, by Dunnett's test, the slight reduction in (adjusted) epididymis weight in the two BPA groups was not significant ($p>0.15$).

(5) We agree with the authors that preputial gland weight is not significantly correlated with body weight. The authors report a significant ($p<0.05$) elevation in preputial gland weight in the low dose BPA group, whereas by Dunnett's test, this difference not significant.

B. Daily Sperm Production (Table 1)

(1) We agree with the authors that the low dose OP group shows a significantly ($p<0.05$) reduced daily sperm production relative to controls. However, had the authors used a protected LSD test, the p value associated with this comparison would have been only $p=0.06$ - the significance of the overall difference among the five groups as determined by ANOVA.

(2) We were unable to confirm any of the summary statistics reported in Table 1 for sperm efficiency based on the raw data provided to us. However, the discrepancies in mean values were all less than 2%. By Dunnett's test none of the reductions in (log-transformed) sperm efficiency were statistically significant, using alpha=0.05, although the p value for the low dose OP effect was $p=0.057$.

(3) Interestingly, although we could not link directly the sperm production data to the organ weight data, there was a common indirect link - testis weight (sperm efficiency was simply the daily sperm production divided by the right testis weight). When the right testis weights were calculated indirectly from the sperm production data, for four of the five groups the resulting mean values were very consistent (within 1-7%) with those calculated from the individual right

testis weights provided to us. However, for reasons that are unclear, the high dose OP group had calculated right testis weights that were significantly ($p<0.01$) lower than those calculated directly from the raw data, averaging 23% less. In fact, there was no overlap between the "calculated " testis weights and the actual testis weights in this group.

Comments

In theory, the results of our reanalysis of organ weights are not necessarily in conflict with those of the authors, given the different statistical test procedures used (Dunnett's test vs. Fisher's LSD). Importantly, however, none of the overall relevant ANOVA's or ANCOVA's for organ weights revealed significant differences among groups, even at the $p<0.10$ level, so the standard application of the protected LSD test (described in the text) would not have flagged any of these organ weight changes as statistically significant. Since none are significant by ANCOVA/Dunnett's test, we cannot confirm the significant BPA effects on organ weight reported by the authors.

In contrast, by Dunnett's test, we were able to confirm the significant effect on daily sperm production reported by the authors for the low dose OP group, and this dose's effect on sperm efficiency was strongly suggestive ($p=0.057$). However, we were unable to confirm the significant ($p<0.05$) high dose BPA effect on sperm efficiency reported by the authors. In this latter case, the lack of concordance is simply reflecting a difference in statistical methodology used in the data analysis (i.e., Dunnett's test vs. Fisher's LSD).

In summary, our reanalysis was able to confirm (only) the following effects reported by the authors in Tables 1 and 2:

(1) The low (2 ng/g) dose of OP and probably the low dose (2 ng/kg) of BPA reduce body weight; and

(2) The low (2 ng/g) dose of OP reduces sperm production and probably reduces sperm efficiency as well.

10. vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November.

Raw data provided: prostate weight and body weights for male CF-1 mice who, when three months old, were castrated and implanted with a Silastic capsule containing either 0.5 mg. testosterone (T) or 0.5 mg. 5alpha-dyhrotestosterone (DHT). The effect of fetal position - located between two females (2F) or located between two males (2M) was of interest.

Results

The data are summarized below

		Body weight (g.)			Prostate weight (mg.)		
Chemical	Position	N	Mean	SD	N	Mean	SD
T	2F	5	34.3	2.7	5	51.9	16.8
T	2M	7	30.1	3.5	7	36.2	13.6
DHT	2F	7	31.4	3.0	7	36.4	7.6
DHT	2M	6	36.7	2.1	6	25.0	4.9

(1) There was no consistent chemical or position effect on body weight. In fact, there was a significant ($p<0.01$) chemical x position interaction, as can be seen in the table above. For Chemical T the 2F animals were heavier than the 2M animals, whereas the reverse was true for the DHT animals.

(2) For (log transformed) prostate weight, there was a significant ($p<0.05$) chemical and a significant ($p<0.05$) position effect, indicating a higher response for Chemical T than for DHT and a higher response for 2F than for 2M animals. There was no significant chemical x position interaction.

(3) Body weights and prostate weights were not significantly correlated for these data.

Comment

An Abstract of this presentation was provided. These unpublished data were not part of the original data requested from this investigator. Moreover, the Abstract reports that "fetal exposure to a very low dose (10 ug/kg/day) of the plastic monomer bisphenol A produced all of the effects on the developing prostate observed with E, EE, and DES..." However, we were unable to confirm this, since these data were not provided to us for re-analysis.

11. Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health 15: 12-25.

Data provided: prostate, seminal vesicle, testis, and body weights for male mice receiving fetal exposure (by feeding pregnant females) to 20 or 2000 ug/kg body weight methoxychlor. These data are summarized in Table 4.

Comments on Statistical Methodology

The authors use of ANCOVA to adjust organ weight for body weight is appropriate. The authors appear to have used a Protected Fisher's LSD analysis to make pairwise comparisons. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Results

(1) We agree with the study authors that methoxychlor has no effect on body weight or testis weight.

(2) We have a minor disagreement regarding seminal vesicle weight in the top dose group which by Dunnett's test is $p=0.09$ rather than $p<0.05$ as reported by the study authors,who used different (but acceptable) statistical methodology - Fisher's LSD test;

(3) We agree that the elevated prostate weights in the two methoxychlor groups is significantly ($p<0.01$) elevated after adjusting for body weight differences among groups. Thus, we agree with the major conclusion of the study authors.

Comment

(1) We note that the means and SE's summarized in Table 4 are for adjusted (for body weight) organ weights rather than for observed organ weights.

(2) We note that certain raw data requested from this study were not provided to us by the study authors. These included intrauterine position data and data on male fetal serum estradiol concentration and androgen binding.

VOM SAAL DATASETS REQUESTED BUT NOT PROVIDED

The following studies were requested from vom Saal, but no raw data were provided:

2. Howdeshell, K. L., A. K. Hotchkiss, et al. (1999). "Exposure to bisphenol A advances puberty." Nature 401: 763-764.

3. Howdeshell, K. L. and F. S. vom Saal (2000). "Developmental exposure to bisphenol A: interaction with endogenous estradiol during pregnancy in mice." American Zoologist 40(3). (in press).

5. Palanza, P., S. Parmigiani, et al. (1999). "Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and o,p'-DDT alters aggressive behavior of male and female house mice." Pharmacology Biochemistry and Behavior 64(4): 665-672.

7. Timms, B. G., S. L. Petersen, et al. (1999). "Prostate gland growth during development is stimulated in both male and female rat fetuses by intrauterine proximity to female fetuses." Journal of Urology 161: 1694-1701.

8. vom Saal, F. S., B. G. Timms, et al. (1997). "Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses." Proceedings of the National Academy of Sciences 94: 2056-2061.

CHAHOUD STUDY

1. Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." (3 Abstracts).

Raw data provided: Sprague Dawley rats treated by gavage with either 0.02, 0.1, or 50 mg/kg

BPA (in 2% Mondamin vehicle), 0.02 or 0.2 mg/kg/day 17alpha-estradiol (EE) (in peanut oil vehicle) or 2% Mondamin (vehicle control) on days 6-21 post conception. Parameters of interest included

Anogenital Distance (AGD) on days 3, 15, and 21
Preputial separation time (not re-evaluated due to lack of time)
Prostate weight at PND 70 and 170.
Daily sperm production (in millions) at PND 70 and 170
Level of testosterone at PND 70 and 170

Experimental Design and Statistical Methods:

One of the 23 questions asked each investigator was whether or not concurrent controls were used, and Dr. Chahoud's response indicated that there were no concurrent controls in his study. Instead, the data were generated sequentially according to the following time frame:

BPA 50 mg/kg/d (Sept. 98-Oct 98)
Mondamin (control group) (Oct 98-Dec 98)
0.1 mg/kg/d BPA (Dec 98-Jan 99)
0.2 mg/kg/d EE (Feb 99 - Feb 99)
0.02 mg/kg/d BPA (Sept 99-Oct 99)
0.02 mg/kg/d EE (Sept 99- Oct 99)

The Statistics Subpanel feels that concurrent controls are an essential experimental design requirement, and thus the lack of concurrent controls in this study was a serious design deficiency. The lowest dose (0.02 mg/kg/d) BPA group was examined approximately 11 months after the control group, while the other two BPA groups were examined 2 months later (0.1 mg/kg/d BPA) or one month earlier (50 mg/kg/d) than controls. Thus, possible treatment effects are confounded with time-related changes.

Note: In response to this concern Dr. Chahoud stated "I believe that the endpoints that were evaluated are robust enough to withstand a 0 to 2 month time difference in treated times" although he provides no evidence to support this speculation. He further stated that "the 0.02 BPA group has an additional vehicle control which was not sent to the committee because the data were not yet in the databank." The Statistics Subpanel confirms that these additional control data were not provided to us. We also note that these control data were not included in the three Abstracts noted above that summarized the study results, nor were they included in Dr. Chahoud's presentation at the Low Dose Endocrine Disruptor meeting.

The Draft Report of the Statistics Subpanel distributed at the Low Dose Endocrine Disruptor Meeting included a statistical reanalysis of these data, based on the critical assumption that there were no time-related changes in the responses of interest. However, after further discussion, the Statistics Subpanel has decided that no statistical re-analysis can really compensate for the lack of concurrent controls. We concluded that this confounding of possible treatment effects with time-related changes precludes any reliable assessment of the effects of EE and BPA on the various parameters evaluated. Thus, our Final Report has no statistical reanalysis of these data,

and we cannot confirm the low dose effects reported by Dr. Chahoud in his three Abstracts.

EMA STUDY

1. Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." (Unpublished Study Report).

Raw data provided: In this multi-generation study, groups of 25 male and female Crj: CD (SD) IGS rats were given bisphenol A (BPA) at 0.2, 2, 20, or 200 ug/kg/day by gastric intubation throughout the study beginning at the onset of the 10- and 2-week pre-mating period, in F0 males and F0 females , respectively, and continuing through the mating, gestation, and lactation period, for two generations. Organ weights, anogenital distance, vaginal opening, preputial separation, and sperm count data were provided for F0, F1, and F2 groups.

Comments on Statistical Methodology

In general we agree with the statistical methodology used, subject to the following comment: Dunnett's test does not require the significance of an overall ANOVA.

Results

A. Sperm count in F0 and F1 generation (Table 11)

We agree with the authors that BPA had no significant effect on sperm count

B. Age of preputial separation and age of vaginal opening in F1, F2, and F2 (satellite) animals (Table 13)

We agree with the authors that there were no significant BPA effects on either of these two variables in any generation, regardless of whether or not the data were adjusted for body weight on days of preputial separation/vaginal opening. This latter variable (body weight) also showed no significant BPA effects.

C. Testis, prostate, epididymis, and seminal vesicles weight in F0 and F1 males (Table 24)

(1) We agree with the authors that there were no significant BPA effects on body weight, including the slight (4%) decrease seen in the top dose (200 ug/kg) group in the F0 generation and the even smaller 2% decrease seen at that dose in the F1 generation..

(2) We agree with the authors that the organ weight data are basically negative. The authors reported a significant ($p<0.05$) decrease in right testis weight in the 20 ug/kg/BPA dose in the F1 generation. Our Dunnett's test did not confirm this, but did indicate a significant ($p<0.05$) reduction in overall (right + left) testis weight in this dosed group. However, after adjusting for body weight by ANCOVA, this difference was not significant at the $p<0.05$ level, and thus this one isolated effect is most likely a spurious finding of no biological significance.

(3) Generally, organ weight and body weight were significantly ($p<0.05$) correlated for these data.

D. Seminal vesicle weights in F2 parental male rats (Table 25)

(1) We agree with the authors that there were no significant BPA effects on body weight in either the F2 or F2 satellite animals.

(2) We agree with the authors that absolute seminal vesicle was marginally reduced ($p<0.05$) in the 200 ug/kg group. However, these animals were also 4% lighter than controls, and after adjusting for body weight differences by ANCOVA, the slight decrease in seminal vesicle weight was not significant. A similar reduction was not seen in the F2 satellite animals. Thus, it is most likely a spurious finding.

(3) Seminal vesicle weight was highly correlated ($p<0.01$) with body weight for these data.

E. Ovary and uterus weights in F0 and F1 females (Table 26)

(1) We agree with the authors that there were no significant BPA effects on body weight in either the F0 or F1 animals.

(2) We agree with the authors that there was no BPA effect on uterus weight for these data.

(3) We also agree that there was a statistically significant ($p<0.05$) decrease in ovary weight in the lowest (0.2 ug/kg) BPA group in the F1 generation. By Dunnett's test, this 10% reduction in ovary weight was significant both before ($p=0.03$) and after ($p=0.04$) an adjustment for body weight. This one single effect may or may not be biologically important, but it was not supported by a corresponding decrease at this dose in the F0 animals.

(4) Ovary weight was significantly ($p<0.01$) correlated with body weight in these studies, but uterus weight was not.

F. Anogenital distance (AGD) in F1 and F2 males and females on Days 0, 4, 7, 14, or 21 of lactation (Table 32)

(1) We agree with the authors that two of the four subsets of data showed no significant effects: Male and female F2 animals.

(2) We also agree that for F1 males there was a marginally significant ($p=0.045$) reduction in AGD in 0.2 ug/kg males on Day 14, but this reduction was not significant ($p=0.2$) after adjusting for body weight and is thus of questionable biological significance.

(3) We also agree that for F1 females the 200 ug/kg BPA dose produced a significant ($p<0.05$) reduction in AGD on day 4 and that the 2 and 20 ug/kg BPA doses produced a significant ($p<0.05$) increase in AGD on Day 7. These differences were significant both with and without an adjustment for body weight.

The biological significance of these changes is uncertain at best. Importantly, (i) the significant changes go in opposite directions; (ii) they are not supported by similar changes in any of the other four time points; and (iii) they are not supported by any similar changes in males or in the F2 generation. Thus, especially in light of the large number of comparisons made, they could simply be reflecting random variability.

(4) Anogenital distance and body weight were generally highly correlated ($p<0.01$) for these data.

G. Testis, prostate, epididymis, and seminal vesicle weights in F1 and F2 weanlings of rats treated with BPA (Table 42)

(1) We agree with the authors that for the F1 animals the only significant body weight effect was a significant (9%; $p<0.05$) reduction in the 20 ug/kg BPA group.

(2) We also agree that for F1 males there were no significant body weight differences, with or without adjusting for body weight.

(3) We also agree that there were no significant body weight effects for F2 males.

(4) We also agree that the only significant organ weight change for F2 males was a 17% reduction in seminal vesicle weight in the 2 ug/kg BPA group, a reduction that was significant both with ($p<0.01$) and without ($p<0.05$) a body weight adjustment. However, this single isolated effect among the 32 pairwise comparisons made for these data (4 BPA groups x 4 organs x 2 generations) is of questionable biological significance.

(5) There were highly significant ($p<0.01$) correlations between organ weight and body weight for all four organs in the F1 and F2 generations.

H. Uterus and ovary weights in F1 and F2 weanlings of rats treated with BPA (Table 43)

(1) We agree with the authors that there were no significant body weight or organ weight changes in any of the BPA groups for either the F1 or F2 animals.

(2) There were highly significant ($p<0.01$) correlations between organ weight and body weight for uterus and ovary in the F1 and F2 generations.

I. Anogenital distance (AGD) in F1 and F2 parental male and female rats treated with bisphenol A (Tables 18 and 19).

(1) These were the most difficult data to interpret in this study. Anogenital distance data were evaluated at multiple time points for both sexes in the F1, F2, and F2 satellite animals. A total of 96 pairwise comparisons were made for males and 156 for females.

(2) By the authors' statistical analysis (summarized in Tables 18 and 19) 33 of the 96 comparisons in males and 16 of the 156 comparisons in females (unadjusted for body weight

differences) showed significant ($p<0.05$) reductions in anogenital distance relative to controls. These reductions were most apparent in the 20 and 200 ug/kg BPA groups, but the lowest dose (0.2 ug/kg) BPA male group also showed significant ($p<0.05$ or $p<0.01$) reductions in AGD for the F1 group at 57, 64, 71, and 78 days after birth.

(3) Complicating interpretation of these data were the corresponding reductions in body weight that were observed in the BPA animals. The statistical significance of the reductions in AGD depend upon how the body weight adjustment is carried out.

The authors adjusted by taking the AGD/body weight ratio and concluded that "these changes [AGD] were not considered compound-related since these changes were within 5% of control values and relative AGD (AGD/body weight) in these groups were not significantly different from the control values." This characterization of the magnitude of the AGD effect is essentially accurate. However, using an ANCOVA adjustment for body weight (the method of adjusting for body weight that we feel is more appropriate), many of the AGD reductions remained statistically significant ($p<0.05$).

To summarize, the following factors argue against the biological significance of these AGD reductions: (i) the magnitudes of the decreases are consistently small (<5%); (ii) few if any of the changes are significant based on an analysis of the AGD/body weight ratio; and (iii) AGD data in F1 and F2 males and females on Days 0, 4, 7, 14, or 21 of lactation (Table 32; see F. above) are not supportive of this effect.

The factors arguing for the biological significance of these reductions include: (1) The consistency and large number of statistically significant AGD reductions seen at multiple dose levels, generations, and in both males and females; (see authors' Tables 18 and 19); and (ii) ANCOVA cannot account totally for these AGD reductions.

Comments

(1) Ultimately, the reader must decide if the slight reductions in AGD discussed above are biologically important. The authors feel that they are not, and we agree that this is a reasonable interpretation of the data. These decreases, if real, appear to be the only toxicological effect of BPA under the conditions of this study, based on the data provided to us.

(2) In their Abstract the authors refer to significant increases in the AGD for 20 and 200 ug/kg/day F1 and F2 females. These were actually decreases, not increases (see Table 19).

GRAY STUDY

1. Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health 15: 48-64.

Raw data provided:

(i) anogenital distance (AGD) data for male rat offspring receiving perinatal exposure to vinclozolin (V) at doses of 0, 3.125, 6.25, 12.5, 25, 50, and 100 mg/kg/day. Data provided to us were limited to post natal day (PND) 2 only. These data (from Blocks 1-4) are summarized in Table 2.

(ii) prostate, testis, seminal vesicle, and cauda epididymal weights, and cauda epididymal sperm, epididymal sperm, and fertility data for male rat offspring receiving perinatal exposure to vinclozolin (V) at doses of 0, 3.125, 6.25, 12.5, 25, 50, and 100 mg/kg/day. These data (from Blocks 1) are summarized in Table 3 and are for animals 55-56 days of age

(ii) prostate, testis, seminal vesicle, and cauda epididymal weights, and cauda epididymal sperm, epididymal sperm, and fertility data for male rat offspring receiving perinatal exposure to vinclozolin (V) at doses of 0, 3.125, 6.25, 12.5, 25, 50, and 100 mg/kg/day. These data (from Blocks 1 and 4) are summarized in Table 4 and are for animals about 12 months of age.

(iv) proportion of pups with nipples for male rat offspring receiving perinatal exposure to vinclozolin (V) at doses of 0, 3.125, 6.25, 12.5, 25, 50, and 100 mg/kg/day. These data are summarized in Table 4 and are reported as being for animals about 12 months of age, but actually include also the 55-56 day old animals from Table 3, as is discussed below.

Statistical Methodology

Statistical methods applied were generally appropriate for the data subject to the follow comments:

(1) The authors report that pairwise comparisons were made by two different methods: Fisher's protected LSD (which they denote as "LSMEANS") or Dunnett's test. These tests are used for different purposes, and it is unclear which variables are being evaluated by which procedure.

(2) Dunnett's test does not require the significance of an overall ANOVA.

(3) The reported summary statistics for certain variables pooled values over blocks, which suggests that in some instances differences among blocks may not have been taken into account in the statistical analyses.

(4) The title to Table 4 does not appear to accurately describe certain variables summarized in this table, as is discussed in more detail below.

(5) Tables 2 and 3 did not indicate if the reported measure of variability was standard deviation (SD) or standard error (SE). It appeared to be SE.

Table 2

(1) We agree with the authors' decision to use the litter as the basic experimental unit, given the significant "litter effects" that were present in the data. We note that this resulted in relatively small sample sizes in the 50 mg/kg/day (N=3) and 100 mg/kg/day (N=2) groups.

(2) We agree with the authors that vinclozolin (even the higher doses) did not significantly effect PND 2 body weight (data not shown in paper).

(3) We agree in general with the authors' conclusions that at PND 2 AGD was reduced in the groups receiving vinclozolin. However, by our analysis (Dunnett's test adjusting for litter effects, block effects, and the correlation of AGD with body weight) the very slight (3.5%) reduction in AGD (adjusted for body weight and block effect) observed in the 6.25 mg/kg/day dosed group is not statistically significant, even at the $p<0.10$ level. The authors reported $p<0.05$ for this comparison. The SE's reported for AGD in Table 2 for these groups (assuming the measures of variability are SE's; we obtained slightly different, but consistent values) also do not support the statistical significance of this dosed group. However, we agree that the AGD reduction at the lowest dosed group (3.125 mg/kg/day) is significant ($p<0.05$).

(4) While this is admittedly a secondary matter, we do not agree with the authors' assessment of litter effects for AGD. First of all, we would not have limited our analysis to the control data only. Secondly, the authors report on page 54 that the number of degrees of freedom (df) is 1.97 for their test, whereas that value (we think) is the value of the F statistic, not the degrees of freedom. Thirdly, and most importantly, since there are only 20 control litters among the four blocks (see Table 2), a true test for litter effects nested within blocks should have only 16 degrees of freedom for litter (and 3 for blocks), not 19 as reported by the authors on page 54. The authors' reported test for litter effects $F(19,92)$ is possible only if blocks are ignored, which would not be appropriate, since the block effect is significant. Finally, even if the block effect is ignored, the F value for the $F(19, 92)$ test is $F=5.29$, $p<0.0001$, not $F=1.97$. The Statistics Subpanel viewed this as a relatively minor issue, especially since we agree with the authors that significant litter effects are present in the data.

(5) Finally, we do not understand the authors' statistical analysis of "% controls" or understand how a control response can exceed 100%, but this should not be the primary statistical analysis of interest for AGD in any case.

Table 3

(1) We agree with the authors' decision to use the litter as the basic experimental unit, since body weight and all organ weights (with the exception of seminal vesicle weight) showed a significant ($p<0.01$) "litter effect." However, using the litter as the basic experimental unit resulted in rather small "N's": only 2-3 litters for each dosed group and 5 control litters.

(2) The last column of this table has an array of p values that are undefined. We suspect that they may be associated with an overall ANOVA, or possibly a tend test, but this is unclear.

(3) We agree with the authors that vinclozolin had no significant effect on body weight in this study.

(4) We agree with the authors that the 50 and 100 mg/kg/day vinclozolin doses significantly ($p<0.01$) reduced ventral prostate weight, both with or without an adjustment for body weight. We also agree that the 100 mg/kg/day dose significantly ($p<0.01$) reduced seminal vesicle

weight. Finally, we agree that testis weight and paired epididymides weight were unaffected by vinclozolin.

(5) We also agree that testis sperm and epididymal sperm counts were unaffected by vinclozolin.

Table 4 - Organ weights

(1) We agree with the authors' decision to use the litter as the basic experimental unit, given the significant ($p<0.05$) "litter effects" that were present in the data. However, we note that this resulted in relatively small sample sizes in the 25 mg/kg/day (N=4), 50 mg/kg/day (N=3) and 100 mg/kg/day (N=2) groups.

(2) We agree with the authors that there is no vinclozolin effect on body weight, testis weight, or testis spermatids.

(3) We also agree that there is a significant ($p<0.01$) high dose effect on seminal vesicle and cauda epididymal weights.

(4) We also agree that the top three vinclozolin doses produce a significant ($p<0.05$) reduction in ventral prostate weight, but we find the reduction at the 6.25 mg/kg/day dose (discussed later on) to not quite be statistically significant: $p=0.06$.

(5) We agree that fertility is significantly ($p<0.05$) reduced in the top two dosed groups.

(6) We do not agree that the reduction in cauda epididymal sperm in the top dose group is significant ($p<0.05$). The high dose mean response of 67 is based on two highly variable values (0.01 and 134), the second of which is well within the control range (127.4 to 171.5) for the five control litters in this study. A single aberrant litter is insufficient evidence to conclude that a statistically significant effect exists.

(7) It is unclear if the authors' analysis of the Table 4 data adjusted for block effects: this table pooled results for Blocks 1 and 4, and the means values given in the table are simply the averages of all the data, ignoring block. Significant differences were observed between these two blocks, that should be taken into account in the statistical analysis. For example, although animals were 6% heavier on average in Block 4 than in Block 1 (720.5 vs. 678.75), the seminal vesicle weights were 32% heavier on average in Block 1 (2104 vs. 1597). This difference between blocks for seminal vesicle weight is highly significant ($p<0.001$).

While this made no difference for seminal vesicle weight (which clearly showed no vinclozolin effect), ventral prostate weights were also significantly ($p<0.05$) heavier in Block 1 than in Block 4. This affected the statistical significance of the low dose (6.25 mg/kg/day) vinclozolin effect on prostate weight, since the 6.25 mg/kg/day group was disproportionately represented in Block 4, as indicated below (also note the striking Block effect). Note that the average difference in prostate weights between the 6.25 mg/kg/day and control groups averaged across the two blocks (112.5 mg.) is less than is indicated in Table 4 (149 mg.), which ignores block differences.

Ventral prostate weight (mg.)	N	Mean	SD
Controls, Block 1	5	679	99
6.25 mg/kg B, Block 1	3	526	85
Controls, Block 4	4	422	124
6.25 mg/kg B, Block 4	5	350	67

While there is still evidence of an effect of 6.25 mg/kg/day vinclozolin on prostate weight, our reanalysis found this reduced weight to be not as significant ($p=0.06$) as the authors' analysis in Table 4 ($p<0.05$). Since for the various organ weights this is the only potential "low dose effect" in this study for which we were given raw data, we decided to discuss this result in some detail. This single low dose effect may or may not be biologically important.

(8) Finally, we note that the top two dosed groups (50 and 100 mg/kg/day) have litters only for Block 1, not Block 4.

Table 4 - Nipple data

This table lists the percentage of animals with nipples, but not the actual proportion of responding animals. The raw data provided to us indicate the following proportions of responders, which are also consistent with the reported rates in Table 4:

Dose (mg/kg/day)	Number of litters	Proportion (%) pups with nipples
0	19	0/77 (0.0)
3.125	16	1/96 (1.0)
6.25	19	3/117 (2.6)
12.5	12	3/84 (3.6)
25	11	3/56 (5.4)
50	3	10/11 (90.9) [Table 4 gives 91.0]
100	2	7/7 (100.0%)

There were certain discrepancies for these data that we were able to resolve, which include these:

(1) It appears that there were in fact 11 pups evaluated at the 50 mg/kg dose, not 7 as given in Table 4. It is not possible to produce a 91.0% response rate from 7 pups, and the corrected sample sizes also agree with the number of pups reported to be evaluated for ectopic testes and for hypospadias (see Table 4).

There appear to be similar discrepancies for the number of litters evaluated for the 6.25 and 12.5 mg/kg/day dosed groups in Table 3 for this variable.

(2) Although Table 4 indicates that the data in that table are for Blocks 1 and 4 only, the percentage nipple data actually include data from all four blocks. This is apparently clear from the text, but not from the table.

(3) Most importantly, the raw data provided to us indicate that the nipple data from Table 4 include as a subset the nipple data from Table 3 for younger (55-56 day old) animals. Thus, for example, the one low dose (3.125 mg/kg/day) animal reported to have nipples in Table 3 and in Table 4 is really the same animal, despite the titles to the table that suggest that these are different animals in groups with different ages. It is unclear why the authors pooled these particular data, while reporting separately the organ weight changes for the younger (Table 3) and older (Table 4) animals. It is certainly not clear from the paper that the nipple data in Table 4 are pooled data. We suspect this pooled was unintentional.

We agree with the authors that the data show a striking and statistically significant ($p<0.01$) increase in the percentage of animals with nipples at the 50 mg/kg/day and 100 mg/kg/day groups and that the data as presented in Table 4 show a suggestion of an effect at lower doses. However, we would have preferred a statistical analysis and data presentation that clearly separated the nipple responses in the younger and older animals, as was done for organ weight.

LEE STUDY

1. Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine 9(1): 105-111.

Raw data provided: testis, epididymis, seminal vesicle, and prostate weight and anogenital distance (A/G) data for newborn male rats given 0.08, 0.8, or 8 mg/kg body weight nonylphenol (NP) daily from days 1-15 after birth. Animals were sacrificed on Day 31. Data are summarized in Figure 1 of his paper.

Comment on Statistical Methodology

The authors appear to have used a Protected Fisher's LSD analysis to make pairwise comparisons, although they describe their procedure slightly differently. This is an acceptable method of statistical analysis for making all possible pairwise comparisons, as noted in the body of this report.

Results

This study had several limiting features. First of all, there was the small sample size (e.g., only 3 high dose and 3 mid dose animals). Secondly, the study pooled data across three replicates that showed significant differences in control body weight.

The dates of the three replicates were 3-21-97, 5-30-97, and 9-13-97. The three high dose animals were examined on 3-21-97 (along with 4 controls); the three mid dose animals were examined on 5-30-97 (along with two controls); four low dose animals were also part of this replicate; the third replicate contained 3 low dose animals and three controls. It would have been preferable to have had all dosed groups represented in each replicate.

This is especially true since the control group body weights were quite different from replicate to replicate. For example, the three controls in the 9-13-97 replicate weighted 128, 128, and 134 g.,

while the six control animals from the other replicate were much lighter, ranging in weight from 87-98 g. (3-21-97) to 101-102 (5-30-97). These differences in control body weight across replicates is highly significant ($p<0.01$).

Perhaps the most critical problem was that NP had striking effects on body weight: on average a 34% decrease in body weight relative to controls in the top dose group (from 106.56 g. to 70.67 g.) and an 18% decrease (from 106.56 to 87.00) in the mid dose group. As noted in the body of this report, it is very difficult to interpret decreased organ weights when body weight decreases of this magnitude occur.

We agree with the study authors that NP was associated with a decreased absolute organ weight (testis, seminal vesicles, prostate, and epididymis) and anogenital distance in this study. However, for all five of these variables, the effect of NP is confounded with the impact of reduced body weight. For example, as shown in Figure 2 of our main presentation, a simple linear model relating testis weight to body weight (ignoring NP dose) fits the data very well, and the NP effect is not significant after adjusting for body weight by ANCOVA. This result also holds for the other organ weights as well and for anogenital distance.

This is not a problem that can be solved by simply calculating organ/body weight ratios. What would be required (and is unavailable from the data provided to us) would be control animals of equivalent size (and age) to the high dose animals to see whether or not the testis (and other organ) weights are indeed smaller than would be expected in equivalently sized controls.

The Statistics Subpanel concludes while NP clearly decreased organ weights in this study, it is not possible with the data provided to rule out the possibility that this was only a secondary effect of reduced body weight, i.e., that the NP animals were showing organ weights and A/G distances that were "normal" for control animals of equivalent size. The small samples sizes and the significant replicate effects also contribute to the uncertainty of these findings.

SPEAROW DATASETS

1. Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science 285: 1259-1261. 2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August.

Agent: 17beta-estradiol (E2)
Doses: 0, 2.5, 10, 20, 40 ug/animal

Animals: B6, C17, CD-1, S15 Male mice

Variables: Testes weight, Accessory gland weight, Percent tubules with elongated spermatids

Statistical Methods

Analysis of variance (ANOVA) was carried out using PROC GLM in the Statistical Analysis System (SAS). No information on litter identity was provided, so it was assumed that only one animal (male) from each litter was used. Analyses were conducted for each strain separately (B6, C17, CD-1, S15). Levene's test for variance homogeneity across dose groups was conducted for each response variable. Because of evidence of variance heterogeneity for testes weight and accessory gland weight in several cases, both logarithmically transformed variables as well as untransformed variables were analyzed. In addition, the rank transformation was explored. For the response variable "percent tubules with elongated spermatids" (PTES), angle-transformed (arcsin(square root)) variables were analyzed as well as untransformed percents, irrespective of the outcome of the variance homogeneity test. For testes and accessory gland weights, if body weight was significantly ($p<0.05$) correlated with the organ weight when adjusted for the effect of E2, then analysis of covariance (ANCOVA) was conducted, with body weight (or log-transformed or rank-transformed body weight) as the covariate. Body weight was not considered as a covariate in the analysis of PTES. To help in the interpretation of any statistically significant ($p<0.05$) effects of E2 observed in ANOVA or ANCOVA omnibus tests, both linear and quadratic dose effects were tested using contrasts whose coefficients were determined with the ORPOL function in PROC IML. Two-tailed Dunnett's tests of SAS least-squares means were used to compare each dose group to the control group.

Results

Testes weight

Although there was evidence of variance heterogeneity for strains B6 ($p<0.02$) and CD-1 ($p<0.01$), the log transformation did not help to stabilize the variance; in fact, it made things slightly worse. Neither was the rank transformation uniformly helpful, actually worsening the situation for B6 and even inducing heterogeneity for S15. The relationship between mean and variance was somewhat nonspecific for B6, but the mean and variance appeared inversely related for CD-1. Results of statistical tests of untransformed testes weights are reported, as the results based on transformed variables appeared qualitatively similar. Body weight (at 43 days) was significantly ($p<0.0001$) correlated with testes weight for all four strains, and hence was included as a covariate. The effect of dose of E2 was statistically significant ($p<0.0001$ for B6, C17 and S15; $p<0.01$ for CD-1). For B6, C17 and S15, testes weights were significantly ($p<0.0001$) reduced at all dose levels relative to controls. For CD-1, mean testes weights at 20 and 40 ug E2 were reduced ($p<0.01$) as was the mean at 10 ug ($p<0.05$); the mean at 2.5 ug was not statistically reduced ($p=0.08$), although it was numerically smaller than control. All strains except CD-1 showed a linear dose effect ($p<0.001$), while all four showed a quadratic effect ($p<0.0001$, except for CD-1, $p<0.01$); however, the dose-response relationship appeared essentially monotone.

Accessory gland weight

There was evidence of variance heterogeneity for strains B6 ($p<0.02$), C17 ($p<0.03$) and CD-1 ($p<0.001$). However, the log transformation was not uniformly helpful in stabilizing the variance. Here again, the rank transformation was not uniformly helpful, actually worsening the situation for some strains. Various patterns of mean-variance relationships were apparent.

Results of statistical tests of untransformed accessory gland weights are reported. Body weight (at 43 days) was significantly ($p<0.0001$) correlated with accessory gland weight for all four strains, and hence was included as a covariate. The effect of dose of E2 was statistically significant ($p<0.0001$) for all four strains with all strains showing significantly ($p<0.0001$) reduced accessory gland weights at all dose levels relative to controls. All strains except CD-1 showed a strong ($p<0.0001$) linear dose effect. All four exhibited a significant ($p<0.0001$, except for B6, $p<0.01$) quadratic effect, with the highest dose having a numerically larger mean accessory gland weight than one or more lower doses.

Percent tubules with elongated spermatids (PTES)

The results for PTES were qualitatively the same, whether untransformed PTES or angle-transformed percents were analyzed. There was no evidence of variance heterogeneity either before or after the angle transformation. For the B6 and C17 strains, the effect of dose was highly statistically significant ($p<0.0001$), with the response in each dose group (B6: 2.5, 10, 40 ug E2; C17: 2.5, 10, 20 ug E2) being significantly ($p<0.0001$) reduced relative to control. PTES in the 10 and 40 ug dose groups of strain B6 and in the 20 ug dose group of strain C17 were reduced to zero for every animal. In contrast, the omnibus ANOVA test showed no statistically significant (5% level) effect on PTES in the CD-1 strain. None of the CD-1 dose groups (2.5, 10, 20, 40 ug E2) were significantly reduced relative to control, although at 20 and 40 ug, PTES was numerically lower than at 0, 2.5 and 10 ug (e.g., 75.6% at 40 ug versus 91.0 % at 0 ug).

Commentary

The results reported in Spearow et al. (1999) and Spearow et al. (2000) for the particular variables analyzed here were essentially reproduced. Some pertinent issues are now discussed.

1. As was stated in the original draft of this report, the present analysis was based on the strong assumption that all animals were from different litters (i.e., one male per litter). It was further stated that, if this were not the case, then some of the observed dose effects could be due to failure of the analysis to account for littermates. Dr. Spearow was invited to supply feedback to the original draft report and he presented the following reply.

Reply: Mice from the same litter were born on the same day and implanted on the same date. The submitted data lists the birth date and/or implant date. Note that mice born or implanted on different days were from different litters. An effort was made to insure that many litters were represented with each strain x treatment group. Whenever possible larger litters were randomly assigned to several different treatment groups to alleviate litter specific effects. Over the full data set, the mice within each strain x treatment group came from an average of 6.895 different litters, with an average of 2.37 littermates per strain x treatment group. Furthermore, the responses were very consistent within each strain, i.e., they did not jump up and down in an irregular manner with increasing E2 doses. In a statistical model in which the effects of Strain x Dose and effects of Litter (Strain x Dose) on testes weight were fitted, Strain x Dose showed an $F=110.405$ with 18 df; Litter (Strain x Dose) showed an $F=5.028$ with 112 df (with 179 df for Residual). Thus, these data show that only a very small portion of the observed dose effects were due to the failure to account for littermates in the original analysis.

2. Spearow et al. (1999) reported a comparison of a separate CD-1 data set data of four additional strains not included in the data file submitted for this analysis, and hence not analyzed here.

3. Slight differences in methodology between this analysis and that of Spearow et al. are as follows. The Spearow analysis included a formal test for interaction between strain and dose effects in the analysis of testes weights, whereas this analysis did not. However, the conclusions of Spearow et al. regarding interactions are supported by this analysis. Spearow et al. (1999) reported analyses of testes weights with dose of E2 expressed both in ug and in ug/g body weight (with similar results and interpretations), while the present analysis used only dose in ug. In the present analysis, statistically significant differences at the 5% level were noted whereas in Spearow et al. (2000), differences were noted only if significant at the 1% level.

4. Unfortunately, neither the log nor the rank transformation was satisfactory in stabilizing the variances for testes weight or accessory gland weight. However, all of the analyses gave qualitatively similar results, irrespective of whether transformed or untransformed data were analyzed. In addition, although it did not include an adjustment for body weight, the nonparametric Kruskal-Wallis test gave comparable results when used to test for a dose effect of E2.

2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August.

Agent: Estradiol Benzoate
Doses: 0, 0.1, 1, 10 ug/kg

Animals: C17, C57BL, CD1, CD9, SPR Female mice

Variable: Uterus weight on day 4

Statistical Methods

Analysis of variance (ANOVA) was carried out using PROC GLM in the Statistical Analysis System (SAS). The response variable was uterus weight on day 4 (UWD4). No information on litter identity was provided, so it was assumed that only one animal from each litter (female) was used. Analyses were conducted separately for each strain (C17, C57BL, CD1, CD9, SPR). Levene's test for variance homogeneity across dose groups was conducted. Because of evidence of variance heterogeneity in several cases, both logarithmically transformed variables as well as untransformed variables were analyzed. In addition, when the log transformation did not improve variance homogeneity, ANOVA on ranked UWD4 data was explored. If body weight was significantly ($p<0.05$) correlated with uterus weight when adjusted for the effect of estradiol benzoate, then analysis of covariance (ANCOVA) was conducted, with body weight (or log-transformed or rank-transformed body weight) as a covariate. To help in the interpretation of any statistically significant ($p<0.05$) effects of estradiol benzoate observed in ANOVA or

ANCOVA omnibus tests, both linear and quadratic dose effects were tested using contrasts whose coefficients were determined with the ORPOL function in PROC IML. Two-tailed Dunnett's tests of SAS least-squares means were used to compare each dose group to the control group.

Results

Uterus weight on day 4 (UWD4)

There was evidence of variance heterogeneity for strains C57BL ($p<0.02$), CD1 ($p<0.02$) and SPR ($p<0.01$), but not for C17 or CD9. However, the log transformation was not helpful in stabilizing the variance. No specific mean-variance pattern was apparent. Analysis of variance on ranks was explored, but results were qualitatively similar to those based on untransformed and log-transformed variables. In the opinion of the analyst, results based on untransformed data, although similar, appeared generally conservative relative to results based on transformed data. Hence, results of statistical tests using untransformed UWD4 values are reported. Body weight (day 4) was significantly correlated with UWD4 for strains C57BL ($p<0.04$) and CD9 ($p<0.001$), but not for strains C17, CD1 and SPR. Hence, for C17, CD1 and SPR, ANOVA tests were conducted, while for strains C57BL and CD9, ANCOVA tests were conducted.

For strains C17, C57BL and SPR, there was a statistically significant ($p<0.0001$) effect of dose of estradiol benzoate, with a lesser, but still statistically significant ($p<0.05$), effect for strain CD9. The dose effect was not statistically significant at the 5% level for strain CD1. For strain C17, mean UWD4 was statistically elevated at doses 10 ug/kg ($p<0.0001$) and 1 ug/kg ($p<0.05$) but not at 0.1 ug/kg. Both linear and quadratic effects were significant ($p<0.001$), but the dose-response was monotone. For strain C57BL, mean UWD4 was statistically elevated at doses 10 ug/kg and 1 ug/kg ($p<0.0001$), but not at 0.1 ug/kg. Both linear and quadratic effects were significant ($p<0.0001$), but the dose-response was monotone. For strain CD1, although there was not a statistically significant dose effect, the mean uterus weight was numerically elevated relative to control in the highest (10 ug/kg) group. For strain CD9, only the mean uterus weight at 10 ug/kg was statistically ($p<0.05$) elevated relative to control. There was no linear effect of dose, but there was a significant ($p<0.01$) quadratic (apparently monotone) effect. For strain SPR, mean UWD4 was elevated relative to control at both 1 ug/kg ($p<0.001$) and 10 ug/kg ($p<0.0001$), but not at 0.1 ug/kg. Both linear and quadratic effects were significant ($p<0.0001$) in the monotone dose-response.

Commentary

The results of this analysis are in close agreement with the results of Spearow et al. (2000).

1. One minor disagreement occurred in comparing mean uterus weights for strain CD9. As the report states, for strain CD9, only the mean uterus weight at 10 ug/kg was statistically ($p<0.05$) elevated relative to control. The manuscript of Spearow et al. (2000) indicates that uterine weight in the CD9 strain did not increase significantly from 0 to 10 ug EB/kg body weight, but did increase from 1 to 10 ug EB/kg body weight ($p<0.05$ by Tukey-Kramer). The present analysis used Dunnett's test rather than Tukey-Kramer, so in the present analysis, only pairwise

comparisons to controls were made. However, that does not appear to be the cause of the disagreement. The apparent reason for this disagreement is that least-squares means (i.e., means adjusted for body weight differences) were compared in the present analysis while unadjusted means were compared in the analysis of Spearow et al. (2000). The unadjusted mean for 0 ug/kg is numerically larger than the unadjusted mean for 1 ug/kg, but the reverse is true for the least-squares means.

2. The assumption that all animals were from different litters is critical to the present analysis, as was stated in the original draft of this report. Dr. Spearow was invited to comment on that draft and he replied as follows with regard to this issue.

Reply: As in the previous experiment, an effort was made to randomly assign litter mates to several different Estradiol-Benzoate dose treatments (not a single dose). Furthermore, each strain x treatment group in B6, Spr., CD-1 and CD-9 strain mice contained an average of 7.68 mice per strain x treatment from an average of 4.68 different litters. Fitting a model composed of the main effects of Strain x Dose and of Litter (Strain x Dose) revealed that the effect of Strain x Dose was highly significant ($p<0.0001$; $F=16.941$; 19 df) while the effect of Litter (Strain x Dose) was not significant ($p=0.2446$; $F=1.179$; 66 df) (with 75 residual df). Thus, the data suggests that the data analysis is nevertheless appropriate and that strains of mice differ dramatically in Uterotrophic assay dose responses to estradiol benzoate.

3. Modest differences in notation occurred as follows: The C57BL strain in this report corresponds to the B6 strain in Spearow et al. (2000) and the SPR strain in this report is referred to as Mus spretus in Spearow et al. (2000).

4. It is unfortunate that none of the transformations was satisfactory in stabilizing the variance of the response variable. Although it demonstrated some improvement, even the rank transformation was not completely satisfactory. Nevertheless, all of the analyses gave qualitatively similar results, irrespective of whether transformed or untransformed data were analyzed. In addition, although it did not include an adjustment for body weight, the nonparametric Kruskal-Wallis test gave comparable results when used to test for a dose effect of estradiol benzoate.

TYL DATASETS

1. Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology 30: 81-95.

Raw data provided: In this multigeneration study octylphenol (OP) was administered ad libitum to five groups of rats at dietary concentrations of 0, 0.2, 20, 200, or 2000 ppm. The following data were provided

(i) Male reproductive organ weights in F0 parental animals (Table 1)

(ii) Daily Sperm production data for F0 parental animals (Table 1)

(iii) Female reproductive organ weights in F0 parental animals (Table 1)

(iv) Male reproductive organ weights in F1 parental animals (Table 2)

(v) Daily Sperm production data for F1 parental animals (Table 2)

(vi) Female reproductive organ weights in F1 parental animals (Table 2)

(vii) Reproductive organ weights in F2 male offspring (Table 3)

(viii) Daily Sperm production data for F2 male offspring (Table 3)

(ix) Reproductive organ weights in F1 male offspring (pups) (no table)

(x) Reproductive organ weights in F1 female offspring (pups) (no table)

(xi) Reproductive organ weights in F2 male offspring (pups) (no table)

(xii) Reproductive organ weights in F2 female offspring (pups) (no table)

(xiii) Vaginal patency and preputial separation for F1 litters (Table 4)

(xiv) Vaginal patency and preputial separation for F2 litters (Table 5)

Comments on Statistical Methodology

(1) This study (and the companion bisphenol A study) were arguably the most comprehensive of the studies we evaluated. Large sample sizes and a wide range of doses were used. This was a multi-generational study involving many endpoints. The statistical methods were well thought out and appropriate for the data.

(2) On page 85 Jonckheere's test is referenced in a statement citing methods used as "a test for linear trend." Jonckheere's test is a nonparametric procedure that assesses monotonic trends, and cannot be used to assess linear trends.

Results

A. Male reproductive organ weights in F0 parental animals (Table 1)

We agree with the authors that the highest OP dose (2000 ppm) significantly ($p<0.01$) reduced body weight. However, after adjusting for body weight differences by ANCOVA, there were no significant OP effects on organ weight. For these data there was a significant ($p<0.01$) correlation between organ weight and body weight.

B. Daily Sperm production data for F0 parental animals (Table 1)

We agree with the authors that OP had no significant effect on sperm production.

C. Female reproductive organ weights in F0 parental animals (Table 1)

We agree with the authors that OP had no effect on body weight in females. We also agree that uterine weight (with or without adjustment for body weight) was significantly reduced in the 2000 ppm OP group. Uterus weight was significantly ($p<0.05$) correlated with body weight in this study; ovary weight was not.

D. Male reproductive organ weights in F1 parental animals (Table 2)

We agree with the authors that the 2000 ppm OP dose significantly ($p<0.01$) reduced body weight. However, after adjusting for body weight differences by ANCOVA, there were no significant OP effects on organ weight. For these data there was a significant ($p<0.01$) correlation between organ weight and body weight, with the exception of seminal vesicles weight, which showed no significant correlation with body weight.

E. Daily Sperm production data for F1 parental animals (Table 2)

We agree with the authors that OP had no significant effect on sperm production.

F. Female reproductive organ weights in F1 parental animals (Table 2)

Interestingly, our analysis confirmed the authors' finding that the lowest OP dose (0.2 ppm) significantly ($p<0.05$) increased body weight. However, organ weights were unaffected at this or any other OP dose. For these data ovary weight was significantly correlated with body weight, while uterus weight was not.

G. Reproductive organ weights in F2 male offspring (Table 3)

We agree with the authors that the 2000 ppm OP dose significantly ($p<0.05$) reduced body weight. However, after adjusting for body weight differences by ANCOVA, there were no significant OP effects on organ weight. For these data there was a significant ($p<0.01$) correlation between organ weight and body weight, with the exception of dorsal prostate weight, which showed no significant correlation with body weight.

H. Daily Sperm production data for F2 male offspring (Table 3)

We agree with the authors that OP had no significant effect on sperm production.

I. Reproductive organ weights in F1 male offspring (pups) (no table)

For these data the 2000 ppm OP dose significantly ($p<0.01$) reduced (by 17% on average) body weight. After adjusting for body weight differences by ANCOVA, there was only one significant ($p<0.01$) OP effect on organ weight: a significant increase in testis weight in the 2000 ppm group. However, the observed mean testis weight was actually 4% lower in the top

dose group, with the statistical significance arising because of the lower body weight in the top dose group and the strong correlation between body weight and testis weight. Thus, this one significant organ weight change may not be biologically important (but see K. below).

There was a significant ($p<0.01$) correlation between organ weight and body weight for all of the organ weights evaluated. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

J. Reproductive organ weights in F1 female offspring (pups) (no table)

For these data the 2000 ppm OP dose significantly ($p<0.01$) reduced (by 14% on average) body weight. After adjusting for body weight differences by ANCOVA, there were no significant OP effect on uterus or ovary weight. There was a significant ($p<0.01$) correlation between organ weight and body weight for both uterus and ovary. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

K. Reproductive organ weights in F2 male offspring (pups) (no table)

For these data the 2000 ppm OP dose significantly ($p<0.05$) reduced (by 10% on average) body weight. After adjusting for body weight differences by ANCOVA, there were no significant OP effect on epididymis or seminal vesicles weight. However, after adjusting for body weight and litter effects, there was a significantly elevated ($p<0.05$) testis weight in all four dosed groups. Since these data were not summarized in the paper, they are given below (based on liter means).

Dose of OP	Body weight			Testis weight		
	N	Mean	SD	N	Mean	SD
0.0	23	47.8	6.1	23	0.219	0.029
0.2	25	49.0	8.2	25	0.240	0.048
20	27	44.6	5.4	26	0.218	0.038
200	29	47.0	6.1	29	0.234	0.034
2000	28	43.0*	3.1	28	0.211	0.021

*$p<0.05$ vs. controls (Dunnett's test)

The biological significance of these "significant" increases in testis weight is uncertain. Only two of the four dosed groups show numerically elevated mean testis weights, with the significance resulting only after adjusting for the reduced body weights and the strong correlation between testis and body weight. Moreover, the increases in (body-weight adjusted) mean testis weights are relatively modest (from 7-9%) and show no evidence of a dose-response trend. On the other hand, the increase is significant in all four dosed groups and is somewhat consistent with the increase reported above for the high dose F1 group.

There was a significant ($p<0.01$) correlation between organ weight and body weight for all three organs. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

This dataset contained an outlier: a single pup in the 2000 ppm group was reported to have a seminal vesicle weight of 1.0195, compared with a range of 0.0063 to 0.0531 for the 79 other animals in this group. We suspect the pup in question should have had a value 100-fold less than reported. In any case, OP had no effect on seminal vesicle weight in this dataset.

L. Reproductive organ weights in F2 female offspring (pups) (no table)

For these data the 2000 ppm OP dose significantly ($p<0.05$) reduced (by 9% on average) body weight. After adjusting for body weight differences by ANCOVA, there were no significant OP effect on uterus or ovary weight. There was a significant ($p<0.01$) correlation between organ weight and body weight for both uterus and ovary. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

M. Vaginal patency and preputial separation for F1 litters (Table 4)

We agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there is a significant ($p<0.01$) increase in the age of acquiring vaginal patency in the 2000 ppm group. However, we do not agree that the slight increase in the 20 ppm group is significant. This difference in interpretation is surprising, since we in theory used exactly the same statistical method (ANCOVA/Dunnett's test) in our analysis.

We agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there is a significant ($p<0.01$) increase in the age of acquiring preputial separation in the 2000 ppm group.

N. Vaginal patency and preputial separation for F2 litters (Table 5)

We agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there is a significant ($p<0.01$) increase in the age of acquiring vaginal patency in the 2000 ppm group.

We also agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there is a significant ($p<0.05$; the authors reported $p<0.01$) increase in the age of acquiring preputial separation in the 2000 ppm group. Interestingly, this is the only set of such data in this study for which body weight at acquisition was not significantly associated with the day of preputial separation.

2. Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 (Draft Final Report).

Raw data provided: In this multigeneration study bisphenol A (BPA) was administered ad

libitum to five groups of rats at dietary concentrations of 0, 0.015, 0.3, 4.5, 75, 750, and 7500 ppm. The following data were provided

(i) Male reproductive organ weights in F0 parental animals (Table 28)

(ii) Daily Sperm production data for F0 parental animals (Table 28)

(iii) Female reproductive organ weights in F0 parental animals (Table 30)

(iv) Male reproductive organ weights in F1 parental animals (Table 59)

(v) Daily Sperm production data for F1 parental animals (Table 59)

(vi) Female reproductive organ weights in F1 parental animals (Table 61)

(vii) Reproductive organ weights in F2 male offspring (Table 90)

(viii) Daily Sperm production data for F2 male offspring (Table 90)

(ix) Reproductive organ weights in F1 male offspring (pups) (Table 20)

(x) Reproductive organ weights in F1 female offspring (pups) (Table 21)

(xi) Reproductive organ weights in F2 male offspring (pups) (Table 51)

(xii) Reproductive organ weights in F2 female offspring (pups) (Table 51)

(xiii) Vaginal patency and preputial separation for F1 litters (Table 33)

(xiv) Vaginal patency and preputial separation for F2 litters (Table 64)

(xv) Reproductive organ weights in F2 female offspring (Table 92)

Comments on Statistical Methodology

(1) This study (and the companion OP study) were arguably the most comprehensive of the studies we evaluated. Large sample sizes and a wide range of doses were used. This was a multi-generational study involving many endpoints. The statistical methods were well thought out and appropriate for the data.

(2) This study had several clear data "outliers" which suggests that the statistical check for outliers described in the OP study above (as well as in the draft BPA report) apparently was not used or was not used carefully. However, this is a draft report only.

A. Male reproductive organ weights in F0 parental animals (Table 28)

We agree with the authors that the highest BPA dose (7500 ppm) showed a significantly (22%; $p<0.01$) reduced body weight relative to controls. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on organ weight, with the following two exceptions:

(i) testis weight was significantly ($p<0.05$) increased in the top dose (7500 ppm) BPA group. This "increase" is of questionable biological significance, since it is due entirely to reduced body weight. Actual mean testis weight in the 7500 ppm dosed group was essentially identical to controls;

(ii) prostate weight was significantly ($p<0.05$) reduced in the 7500 ppm dosed BPA group. This decrease (which averaged >30% relative to controls) may be somewhat more important.

For these data there was a significant ($p<0.01$) correlation between organ weight and body weight for all five organs: epididymis, testes, prostate, seminal vesicles, and preputial gland..

B. Daily Sperm production data for F0 parental animals (Table 28)

We agree with the authors that BPA had no significant effect on sperm production.

C. Female reproductive organ weights in F0 parental animals (Table 30)

We agree with the authors that the top dose (7500 ppm) BPA group showed a significant (13%; $p<0.01$) reduction in body weight relative to controls. We also agree that uterine and ovary weights (with or without adjustment for body weight) were significantly reduced in the 7500 ppm BPA group.

We also agree with the authors that uterus weight was significantly ($p<0.05$) reduced in the lowest dose group (0.015 ppm) relative to controls. This reduction was 16% even though the animals in the 0.015 ppm group were 2% heavier on average than controls. Adjusting for body weight did not materially affect the statistical significance of this decrease.

We agree with the authors' statement on page 52 that "absolute uterine weights were also significantly reduced at 0.015 ppm." However, we do not agree with the subsequent statement that "Relative .....uterine weights were equivalent across all groups." As noted above, the low dose animals were actually slightly heavier than the controls, so the difference in relative uterus weight (19%) is even more impressive than the difference in absolute uterus weight (16%).

Although we used ANCOVA rather than the uterus/body weight ratio, a Dunnett's test based on the uterus/body weight ratio still reveals a significant ($p=0.02$) reduction in relative uterus weight in the low dose group relative to controls. Whether or not this organ weight effect is a biologically important change is a matter of scientific judgement. We suspect that the difference of interpretation is due to the authors' requirement of a significant overall ANOVA before carrying out Dunnett's test. As noted elsewhere, this extra requirement on Dunnett's test is

unnecessary.

Ovary weight was significantly ($p<0.01$) correlated with body weight in this study; uterus weight was not.

D. Male reproductive organ weights in F1 parental animals (Table 59)

We agree with the authors that body weights were significantly reduced in the 750 ppm (6%; $p<0.05$) and 7500 ppm (26%; $p<0.01$) BPA groups. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on organ weight. All five organ weights were significantly ($p<0.05$) correlated with body weight for these data.

E. Daily Sperm production data for F1 parental animals (Table 59)

We agree with the authors that BPA had no significant effect on sperm production.

F. Female reproductive organ weights in F1 parental animals (Table 61)

We agree with the authors that body weights were significantly reduced in the 750 ppm (6%; $p<0.05$) and 7500 ppm (16%; $p<0.01$) BPA dosed groups relative to controls. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on uterus weight. However, the 29% reduction in ovary weight observed in the 7500 ppm group remained significant ($p<0.01$), after adjusting for body weight differences by ANCOVA. For these data there was a significant ($p<0.01$) correlation between organ weight and body weight for both uterus and ovary.

G. Reproductive organ weights in F2 male offspring (Table 90)

We agree with the authors that the 750 and 7500 ppm BPA dosed groups significantly ($p<0.01$) reduced body weight (12% and 29% respectively) relative to controls. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on organ weight. However, an outlier was detected: In the top dose group a prostate weight was reported to be 5.248 g., while the other 27 values for that group ranged from 0.2285 to 0.7913 g. We strongly suspect that a 10-fold (decimal point) error was made for this value. With this value included, there is significant ($p<0.05$) heterogeneity. Removing it eliminates the heterogeneity.

For these data there was a significant ($p<0.01$) correlation between organ weight and body weight for all five organs.

H. Daily Sperm production data for F2 male offspring (Table 90)

We agree with the authors that BPA had no significant effect on sperm production.

I. Reproductive organ weights in F1 male offspring (pups) (Table 20)

For these data the 7500 ppm BPA dose significantly ($p<0.01$) reduced (by 26% on average) body

weight. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on organ weight. There was a significant ($p<0.01$) correlation between organ weight and body weight for all of the organ weights evaluated. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

J. Reproductive organ weights in F1 female offspring (pups) (Table 21)

We agree with the authors that the 7500 ppm BPA dose produced a significant (26%, $p<0.01$) reduction in body weight relative to controls. After adjusting for body weight differences by ANCOVA, there were no significant BPA effect on uterus or ovary weight. There was a significant ($p<0.01$) correlation between organ weight and body weight for both uterus and ovary. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

K. Reproductive organ weights in F2 male offspring (pups) (Table 51)

We agree with the authors that significant reductions in body weight were observed in the 75 ppm BPA (7%, $p<0.05$) and 7500 ppm (20%, $p<0.01$) groups. After adjusting for body weight differences (and litter effects) by ANCOVA, there were no significant BPA effects on organ weight, with one exception: adjusted testis weights were significantly ($p<0.05$) increased in the 7500 ppm group. However, this "increase" is of questionable biological significance, since it is due primarily to the strong correlation between testis weight and body weight and the reduced body weight observed in the 7500 ppm group. The mean absolute testis weight in the 7500 ppm group is actually 12% below control levels, but the ANCOVA "adjusts" it (since body weights are even more reduced - 20%) to be significantly above control levels.

There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

L. Reproductive organ weights in F2 female offspring (pups) (Table 51)

For these data the 7500 ppm BPA dose significantly ($p<0.01$) reduced (by 22% on average) body weight. After adjusting for body weight differences by ANCOVA, there were no significant BPA effects on uterus or ovary weight. There was a significant ($p<0.01$) correlation between organ weight and body weight for both organs evaluated. There were also highly significant ($p<0.01$) litter effects, which supports the authors' decision to use the litter rather than the individual pup as the basic experimental unit.

M. Vaginal patency and preputial separation for F1 litters (Table 33)

We agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there is a significant ($p<0.01$) increase in the age of acquiring vaginal patency in the 7500 ppm BPA group.

We agree with the authors that after adjusting for body weight at acquisition by ANCOVA, there

is a significant ($p<0.01$) increase in the age of acquiring preputial separation in the 750 and 7500 ppm BPA groups.

N. Vaginal patency and preputial separation (PPS) for F2 litters (Table 64)

The preputial gland separation data had significant ($p<0.01$) heterogeneity, due to the presence of two outliers in the 7500 ppm group: 79 and 88 days; the range of the remaining 26 values was 43-49 days. With the two outliers removed, there is a significant ($p<0.01$) increase in preputial separation (after adjustment for body weight on day of acquisition), not only for the 7500 ppm group, but also for the 0.3, 75 ($p<0.05$ only), and 750 ppm groups. The authors' analysis included the two outliers, which inflated the error term and masked the effects in the lower dosed groups (see discussion of heterogeneity in the body of this report).

Note: The authors subsequently confirmed that these two outliers were in fact "real" values. They stated that "the two F2 males in BPA who acquired PPS very late (at approximately 75 and 82 days of age) were very tiny relative to the other pups, but they mated early in the cohabitation period and sired live, healthy litters, so their data were retained." Including these two values (which would require the use of nonparametric methods) would not materially affect the results given above, and would actually make the 7500 ppm PPS effect even more significant.

The vaginal opening data also had a significant outlier in the 4.5 ppm BPA group: 98 days (range of other values: 27-38 days). This value was deleted from our analysis as an outlier. After adjusting for body weight on day of acquisition, the increase in the day of vaginal opening was significantly ($p<0.01$) increased in the 7500 ppm group (only).

O. Reproductive organ weights in F2 female offspring (Table 92)

We agree with the authors that the 7500 ppm BPA dose produced a significant (14%, $p<0.01$) reduction in body weight relative to controls. After adjusting for body weight differences by ANCOVA, there were significant ($p<0.05$ or $p<0.01$) reductions in ovary weight in the 0.015 ppm group (12% reduced), 4.5 ppm (16%), 75 ppm (12%) and 7500 ppm (34%) groups. The (ANCOVA-adjusted) reductions were 13%, 16%, 12%, and 31% respectively, and all were significant ($p<0.05$).

There was significant heterogeneity for uterus weight, due to a single outlier in the control group: 5.318 g., while the range of the other 29 control uterus weights was 0.39 to 1.21 g. We strongly suspect a 10-fold decimal point error. With this value excluded, there is no BPA effect on uterus weight. Neither uterus weight nor ovary weight showed a significant ($p<0.05$) correlation with body weight in this study, although the ovary weight association was suggestive ($p=0.06$).

WAECHTER DATASETS

1. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences 50: 36-44.

Raw data provided:

(1) prostate, seminal vesicle, epididymis, right caudal epididymis, and testis weights from the male offspring of female CF-1 mice exposed orally to low doses of bisphenol A (BPA): 0, 0.2, 2, 20, or 200 ug/kg/day or to 0.2 ug/kg/day DES. These data are summarized in Table 3.

(2) daily sperm production from male offspring of female CF-1 mice exposed orally to low doses of bisphenol A (BPA): 0, 0.2, 2, 20, or 200 ug/kg/day or to 0.2 ug/kg/day DES. These data are summarized in Figure 4.

Comments on Statistical Methodology

Pooling the two control groups is reasonable, since there did not appear to be significant differences between them. In general, the statistical methods used by the authors were appropriate, subject to the following comments:

(1) There is no need to require a significant overall ANOVA when using Dunnett's test.

(2) Applying a Bonferroni correction to pairwise comparisons made by the Wilcoxon-rank sum test results in a rather conservative procedure. Requiring significance of an overall nonparametric ANOVA (as the authors do) controls the experiment-wide error rate and should be sufficient protection against false positive outcomes without introducing the conservatism associated with the Bonferroni correction.

Results

Table 3

For body weight and most of the organ weights, there was a significant ($p<0.01$) litter effect, so we agree with the authors' decision to use the litter as the basic experimental unit. Body weight and organ weight were significantly ($p<0.05$) correlated for these data.

We agree with the authors that the 20 ug/kg/day and 200 ug/kg/day BPA exposures significantly ($p<0.01$) increases body weight. DES had no significant effect on body weight.

We also agree with the authors that neither BPA nor DES had a significant effect on testis, prostate, epididymis, right caudal epididymis, or seminal vesicle weight.

Figure 4

For daily sperm production, there was a significant ($p<0.01$) litter effect, so we agree with the authors' decision to use the litter as the basic experimental unit.

We also agree with the authors that neither BPA nor DES had a significant effect on daily sperm production.

Comments

(1) We agree with the authors' interpretation of these data.

(2) The set of raw data provided to us by the authors had several unrelated numerical errors (pups mis-assigned to litters, litters mis-assigned to dosed groups) that were ultimately corrected but made the analysis of the data more difficult. This study illustrates the advantage of having summary tables available as a reference to identify potential data discrepancies. For studies having only Abstracts, data discrepancies such as those encountered and corrected in this set of data would likely go undetected.

2. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology 30: 130-139.

Raw data provided:

(1) prostate, seminal vesicle, epididymis, right caudal epididymis, and testis weights from the male offspring of female Han-Wistar albino rats exposed via drinking water to low doses of bisphenol A (BPA): 0, 0.01, 0.1, 1, or 10 ppm, or to 0.1 ppm DES. These data are summarized in Table 3.

(2) daily sperm production from male offspring of female Han-Wistar albino rats exposed via drinking water to low doses of bisphenol A (BPA): 0, 0.01, 0.1, 1, or 10 ppm, or to 0.1 ppm DES. These data are summarized in Figure 3.

Comments on Statistical Methodology

Pooling the two control groups is reasonable, since there did not appear to be significant differences between them. In general, the statistical methods used by the authors were appropriate, subject to the following comments:

(1) There is no need to require a significant overall ANOVA when using Dunnett's test.

(2) Applying a Bonferroni correction to pairwise comparisons made by the Wilcoxon-rank sum test results in a rather conservative procedure. Requiring significance of an overall nonparametric ANOVA (as the authors do) controls the experiment-wide error rate and should be sufficient protection against false positive outcomes without introducing the conservatism associated with the Bonferroni correction.

Results

Table 3

For body weight and most of the organ weights, there was a significant ($p<0.01$) litter effect, so we agree with the authors' decision to use the litter as the basic experimental unit. Body weight

and organ weight were significantly ($p<0.01$) correlated for these data.

We agree with the authors that there were no significant body weight effects of either BPA or DES for these data.

We also agree with the authors that neither BPA nor DES had a significant effect on testis, prostate, epididymis, right caudal epididymis, or seminal vesicle weight.

Figure 3

For daily sperm production, there was a significant ($p<0.01$) litter effect, so we agree with the authors' decision to use the litter as the basic experimental unit.

We also agree with the authors that neither BPA nor DES had a significant ($p<0.05$) effect on daily sperm production, although the 8% reduced sperm production in the DES group was suggestive ($p<0.10$)

Comments

(1) We agree with the authors' interpretation of these data.

(2) This set of raw data, like the previous study with mice, had several unrelated numerical errors (pups mis-assigned to litters, litters mis-assigned to dosed groups) that were ultimately corrected but made the analysis of the data more difficult. This study illustrates the advantage of having summary tables available as a reference to identify potential data discrepancies. For studies having only Abstracts, data discrepancies such as those encountered and corrected in this set of data would likely go undetected.

WELSCH DATASET

1. Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology 14: 359-367. 2. Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A.

Raw data provided: prostrate weight from male Sprague-Dawley rats receiving bisphenol A (BPA) from gestation day 2 through PND 21 in drinking water at concentrations of 0, 0.005, 0.05, 0.5, 5, or 50 mg/l. Estimated daily intake ranged from approximately 0.001 to approximately 10 mg/kg day.

A. Statistical Analysis I (RK)

Statistical Analysis and Results

Analysis of variance (ANOVA) was carried out using PROC GLM in the Statistical Analysis System (SAS). Because replicate 1 had 2 pups/litter while replicate 2 had only 1 pup/litter,

(ventral) prostate weights and body weights were averaged within litters in replicate 1 prior to analysis. Levene's test for homogeneity of variance of prostate weights across dose groups was conducted, and was found to be non-significant. Body weight was not significantly correlated with prostate weight (in rep 1 average values were not correlated), so ANCOVA with body weight as covariate was not used. Both linear and quadratic dose effects were tested using contrasts whose coefficients were determined with PROC IML using the ORPOL function. Two-tailed Dunnett's tests were used to compare each dose group to the control group. For the latter tests, SAS least-squares means were used for the comparisons.

The ANOVA was essentially a split-plot analysis, with Dams as main plots and Litters as sub-plots. The Dam-within-Dose effect was used to test for a Dose effect of BPA (including linear and quadratic dose contrasts), while the Rep-by-Dam-within-Dose effect was used to test for a Rep effect and a Dose-by-Rep interaction.

There was no indication of a Rep effect ($p=0.69$) or of a Dose-by-Rep interaction ($p=0.21$). Hence, BPA tended to have a similar effect in the two replicates. The test of an effect of Dose of BPA was statistically significant ($p<0.02$), but neither the linear nor the quadratic dose effect was significant. Dunnett's test for individual differences between dosed groups and controls indicated that the 0.05 mg/l ($p<0.01$), the 5 mg/l ($p<0.0001$) and the 50 mg/l ($p<0.02$) were significantly increased relative to control. The 0.005 mg/l and 0.5 mg/l were not significantly different from control, although mean prostate weights (averaged over reps) were numerically higher than the control mean in those groups.

Commentary

The results and conclusions of this analysis appear to agree with those of Elswick et al. (2000). First, ventral prostate weight was not correlated with body weight. Most importantly, a significant effect of BPA was found when data on the two replicates (referred to as "blocks" by Elswick et al.) were combined for analysis. In addition to finding a significant treatment effect when analyzing the combined data, Elswick et al. also analyzed the data from the two replicates separately, finding significant differences in replicate 2 but not in replicate 1. However, the specific dose-group differences from control noted in the Elswick analysis of replicate 2 (i.e., 0.05, 5 and 50 mg/l) were reproduced in the present analysis of the combined data. Because no replicate differences and no replicate-by-dose differences were found, it appears that replicate 1 gave a qualitatively similar result as replicate 2, but simply did not achieve statistical significance.

B. Statistical Analysis II (KP)

Procedure:

Pregnant CD-Sprague-Dawley rats (~10 weeks old) consumed drinking water with given doses of BPA from gestation day 2 through PND 21 with estimated daily intakes of ~0.001 to ~10 mg/kg/day. Study was performed in a replicate block design to obtain a target number of 16 dams/dose group. In block one, two male pups were necropsied whereas in block two only one male pup was necropsied at PND 177. Prostate gland was separated into dorsolateral prostate

and VP. The micromorphology of VP from all treatment groups was examined in H & E stained sections.

Parameters of Interest:

Ventral prostate weight (grams) at PND 177
Dorsal prostate weight (grams) at PND 177
Total prostate weight (grams) at PND 177

Statistical Methods:

Abstract provides treatment group means and standard errors (Table 4), No other indication of statistical test methodology or manipulations of responses are provided. Text suggests that some form of covariate adjustment using body weights was provided. Analysis was performed by block separately then with blocks combined.

Reanalysis first attempted to recreate Table 4 of means and standard errors. Data from the two pups per dam for the first block (replicate) were averaged, this included all prostate measurements and body weight. This data was combined with that of the second block for subsequent analysis. Two analysis models were explored. The first was a split-plot ANOVA fixed effects model with dam within treatment as the main plot error and reps within dam by treatment as the split-plot error. Treatment effects were tested against main plot error and rep main effects and rep by treatment effects were tested against split plot error. Analysis of expected mean squares indicated that the main plot test for treatment effects was not exact. Next a mixed effects general linear model was run with treatments, rep and rep by treatment interactions as fixed effects and dam within treatment as a random effect as well as the rep within treatment by dam residual. Both models incorporated a simple linear regression of body weight as a covariate. In each model, Dunnett's multiple comparison procedure was run to compare treatments to the zero dose control. Only the results of the mixed model analysis are reported below. Residuals from the selected analysis model were examined for normality.

Results:

Ventral prostate weight:

Reanalysis was able to reproduce the VP means given in Table 4 (which apparently were pooled over reps and dams), but not the standard errors.

A mixed effect model with treatments adjusted for body weight was also fit. The body weight covariate was not statistically significant ($p=0.76$).
With or without the body weight adjustment, the 0.05, 5 and 50 treatment levels were found to be statistically different from control. Dam within treatment variability was significant ($p=0.025$). Residuals were normal looking. There was no significant replicate or replicate by treatment interaction in the data.

A reanalysis for each rep separately found, after adjusting for body weight, that the 5 level was

significantly different from control for rep 2 and none of the treatments were significant for rep 1. This suggests that the significant treatment effects observed in the overall analysis are present in the individual reps but that it requires the combined power of the two reps to actually identify these significant effects. Looking at the simple (adjusted) means for each rep one can see similar patterns.

Rep = 1

trt	MEAN	Standard Error
control	0.51759564	0.04142598
0.005	0.50079052	0.04163314
0.05	0.62340607	0.04162727
0.5	0.60878430	0.05243254
5	0.61057561	0.04500588
50	0.57772719	0.04140790

Rep=2

trt	MEAN	Standard Error
control	0.38846731	0.07428831
0.005	0.53028064	0.05858102
0.05	0.58114569	0.06009946
0.5	0.49596062	0.05864708
5	0.72377754	0.05968106
50	0.63905792	0.06782318

Dorsal prostate weight: Standard errors observed in reanalysis data set not the same as those reported in Table 4, but means are. The reanalysis following the general approach used for ventral prostate weight found no significant treatment effects.

Total prostate weight: Standard errors observed in reanalysis data set not the same as those reported in Table 4, but means are. No treatment effects were observed as significantly different from controls at the type I error probability of 0.05. But, the 5 and 50 levels treatments were close to significance (p=0.052 and 0.056 respectively) and the 0.05 level not significant at the p=0.097 level.

Commentary:

The original report suggested significant 0.05, 5 and 50 mg/L treatment differences as being significantly from the zero dose control for the ventral prostate (VP) lobe measurement. The researchers also found this was the case for the analysis with rep block two data analyzed separately. The restudy confirmed the overall combined data analysis findings but was unable to duplicate the rep block two findings.

The original study also reported significant differences in control group means between rep blocks for VP. The restudy was unable to find any significant differences, even when body

weight was taken into effect.

The original study concluded that large intra-litter variability of the VP weights was a confounding factor in finding treatment effects. Here, intra-litter variability was measured by the rep and rep by treatment interaction effects. Neither of these was significant in any of the overall models.

One aspect of the restudy analysis that was still an issue was the impact that averaging the data from pups within each dam for the rep block one experiment had on the final results. When this averaging is not done and all of the pup-to-pup variability is included in the analysis, the findings for VP still still stand providing additional strength for a treatment response.

There is very little evidence in these data for a body weight effect on the treatment effects.

Final comment on Welsch Study prostate weight data: Although slightly different methodologies were used in some cases by the two reviewing statisticians, both are in agreement that (i) an analysis of the two blocks combined revealed significantly ($p<0.05$) elevated ventral prostate weight in the 0.05, 5, and 50 mg/l BPA groups; (ii) BPA tended to have a consistent effect on ventral prostate weight in the two replicates, achieving statistical significance in one replicate, but not in the other when evaluated separately; and (iii) ventral prostate weight was not significantly correlated with body weight in this study.

C. Statistical Analysis III (JH)

In addition to providing an analysis of prostate weight, the authors present the results of some methodological work investigating the effect of different sampling designs on the outcome of endocrine disruptor studies. Based on the results of simulations studies, they conclude that because of the "large intralitter variability" in prostate weight response, the use of only one pup per litter leads to "a substantial percentage of incorrect conclusions about the presence or absence of treatment effects observed" and that such a design "should not be used when assessing effects on highly variable organ weights and other reproductive endpoints." Unfortunately, their simulation study is seriously flawed. Consider the following points.

The authors opine that the positive findings in several studies from vom Saal's lab (discussed above) as well as in a one-pup-per litter study from their own lab (discussed earlier in this section) may have been false positive outcomes resulting from the use of only one pup per litter. Other studies with more pups per litter did not reproduce these effects, and thus the authors concluded that this "problem" was associated with "the sampling of only one or two pups per litter in the original [positive] reports," results that presumably were reflecting false positive outcomes. They further concluded that their simulation study supported this point of view, revealing that "serious mistakes regarding treatment effects can be made when only one or two pups are selected from each litter." These incorrect conclusions can either be false positive or false negative outcomes, i.e., "incorrect conclusions can be made about the presence or absence of a treatment effect due to the sampling strategy."

For a fixed total number of litters, increasing the number of pups per litter will increase power

(will reduce the false negative rate), but it has no impact on the false positive rate. The false positive rate associated with a particular statistical methodology is fixed by the selection of alpha (typically 0.05). If the null hypothesis is true and there is no difference among the experimental groups, and if the statistical tests used are appropriate for the data (as was the case with the prostate weight analysis), then the p value (i.e., the actual false positive rate) should be essentially equal to alpha.

None of the authors' simulations indicated that the sampling strategy used (one, two or three pups per litter) had any impact on what they regarded to be the false positive rate. Nevertheless, the authors concluded that "the sampling of only one or two pups per litter" may have been a "contributing factor" to the positive low dose effects observed by some investigators, effects that were not confirmed by others who used more than two pups per litter. However, nothing in the Elswick et al. paper supports their speculation that these low dose effects were merely false positive outcomes resulting from the use of only one or two pups per litter.

Thus, while the "significant prostate weight effect" found by vom Saal's lab and by Welsch when using a one pup per litter experimental design may or may not be a biologically important finding, it was NOT a direct consequence of using only one pup per litter, as the authors imply.

If for a fixed total number of litters, an experimenter has the choice of using single or multiple pups per litter, with no difference in cost or time, then the multiple pup per litter strategy is preferred. As noted above, increasing the number of pups per litter decreases the false negative rate (i.e., increases power), but it has no impact on the false positive rate. The reason that power is increased as the number of pups per litter is increased is that variability in the litter-based average response is reduced. To suggest that using fewer pups per litter (thereby increasing the variability) would lead to increased findings of statistical significance (whether correct or false) is illogical.

Moreover, the authors' simulation studies were flawed in that they were based on sampling from a sample rather than sampling from an underlying population. A given sample of litters hopefully represents a randomly selected set of litters from a much larger underlying population of similar litters. While the characteristics of the sample can be used to help define or infer the characteristics of the underlying population from which it was selected, it should not itself be considered the population (for purposes of a simulation). Simulations should be based on populations, not samples. If the underlying populations are identical, then the simulations assess false positive rates; if they are different, then the simulations assess false negative rates (power).

The authors' simulations were based on comparisons of subsamples selected without replacement from two (or more) finite samples with different observed mean responses. In those instances in which the samples were not statistically different, the authors considered the samples themselves to be identical, and then used them as populations in their simulation study. However, since the samples (now regarded as populations) were in fact different, all of the authors' reported "false positive rates" (based on the comparisons of subsamples selected from the nonidentical samples) were in fact power calculations. Moreover, the use of small finite "populations" greatly limits the possible p values that could result from a multiple pup per litter sampling strategy, and in the extreme case of complete subsampling, leads to a single outcome that is reproduced exactly in

every "simulation".

The important point here is that NONE of the authors' many simulations were actually assessing false positive rates. All of them were addressing the issue of power for detecting either subtle or marked changes in prostate weight.

Finally, the authors' recommendation to increase the number of pups per litter may not be the best approach to optimize power. They conclude that "despite an adequate number of litters," serious mistakes can be made "when only one or two pups are selected from each litter."

As noted in the text of our report, if litter effects are present in the data, then the appropriate sampling unit is the litter, not the individual pup. In such cases it is easy to demonstrate (by simulation studies or by other means) that for a fixed total sample size, power is maximized by sampling single pups from multiple litters rather than multiple pups from a small number of litters. The stronger the litter effect, the greater the gain in power by sampling more litters. For example, when strong litter effects are present, 20 litters with one pup per litter will have more power than 5 litters of 4 pups per litter. Both strategies will have the same false positive rate, as discussed above.

Thus, while sampling multiple pups per litter clearly has more power than sampling single pups per litter for a fixed total number of litters, it may not be the optimal experimental design if additional litters are available.

In summary, the Statistics Subpanel believes that the simulation study is seriously flawed and gives a misleading impression of the statistical "benefits" associated with a multiple-pup-per litter experimental design. Moreover, we disagree with its implication that positive statistical findings based on single pup per litter studies (such as those conducted by Welsch, vom Saal, and possibly others) may be nothing more than false positive outcomes related to the use of only one pup per litter. Even the authors' flawed simulation study provides no evidence to support this speculation.

FINAL COMMENTS

As indicated in the body of this report, there were several low dose endocrine disruptor studies for which raw data were requested, but for which no statistical reanalysis was carried out. Some investigators chose not to provide us with the raw data that we requested. Other investigators did submit data, but their data were not selected for re-evaluation by the Statistics Subpanel for a variety of reasons, the primary one being simply lack of time. Importantly, both Drs. Chapin and Newbold submitted extensive multiple datasets, and the lack of statistical analysis of these particular studies should definitely NOT be inferred to mean that these investigators did not provide the requested data or that their studies were flawed in any way. We simply ran out of time.

Appendix B:
Investigators' Responses to "Issues Relative to the Evaluation of Endocrine Low-Dose Studies"
Index

39]

3. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep (Unpublished Abstract). **[PAGE B-43]**

4. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of genistein in male and female Sprague Dawley rats." (Unpublished Final Report). **[PAGE B-46]**

5. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of methoxychlor in male and female Sprague Dawley rats." (Unpublished Final Report). **[PAGE B-50]**

6. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of nonylphenol in male and female Sprague Dawley rats." (Unpublished Final Report). **[PAGE B-54]**

7. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats." In prep (Unpublished Abstract). **[PAGE B-57]**

8. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression." In prep (Unpublished Abstract). **[PAGE B-61]**

9. Meredith, J. M., C. Bennett, et al. (2000). "Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats." Society for Neuroscience Abstracts (in press).

10. Scallet, A. C., C. Bennett, et al. (1999). "Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure." Society for Neuroscience Abstracts 25: 227. **[PAGE B-64]**

John O'Connor

1. Biegel, L. B., J. C. Cook, et al. (1998). "Effects of 17ß-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats." Toxicological Sciences 44: 143-154. **[PAGE B-67]**

2. Biegel, L. B., J. A. Flaws, et al. (1998). "90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17ß-estradiol." Toxicological Sciences 44: 116-142. **[PAGE B-72]**

3. Cook, J. C., L. Johnson, et al. (1998). "Effects of dietary 17ß-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats." Toxicological Sciences 44: 155-168. **[PAGE B-76]**

4. O'Connor, J. C., S. R. Frame, et al. (1998). "Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17ß-estradiol." Toxicological Sciences 44: 169-184. **[PAGE B-80]**

Frederick vom Saal

1. Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October. **[PAGE B-83]**

4. Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives 105(1): 70-76. **[PAGE B-87]**

6. Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." (in press). **[PAGE B-90]**

9. vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health 14 (1/2): 239-260. **[PAGE B-94]**

10. vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November.
[PAGE B-97]

11. Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health 15: 12-25.
[PAGE B-100]

Ibrahim Chahoud

1. Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." (3 Abstracts).
[PAGE B-104]

Mokoto Ema

1. Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." (Unpublished Study Report). **[PAGE B-111]**

Earl Gray

1. Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health 15: 48-64. **[PAGE B-114]**

Ping Lee

1. Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine 9(1): 105-111.
[PAGE B-117]

Jimmy Spearow

1. Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science 285: 1259-1261. **[PAGE B-121]**

2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August. **[PAGE B-124]**

Rochelle Tyl

1. Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology 30: 81-95. **[PAGE B-127]**

2. Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 (Draft Final Report). **[PAGE B-133]**

John Waechter

1. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences 50: 36-44. **[PAGE B-139]**

2. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology 30: 130-139. **[PAGE B-145]**

Frank Welsch

1. Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology (in press).
2. Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A.
3. Welsch, F., B. A. Elswick, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A on female offspring of Sprague-Dawley rats." Toxicological Sciences 54(Supplement): 256A.
[PAGE B-151]

Appendix B:
Investigators' Responses to
"Issues Relative to the Evaluation of Endocrine Low-Dose Studies"

John Ashby

1. Ashby, J., H. Tinwell, et al. (1999). "Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate glad of CF1 mice exposed in utero." Regulatory Toxicology and Pharmacology **30**: 156-166.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To confirm previously reported effects of exposure to BPA *in utero* using larger group sizes, (Nagel et al 1997, EHP 105 : 70-76).	
2) Species, strain, and source of animals	CF1 mice purchased from Charles River, Portage, USA	
3) Diet/source	Pregnant and lactating females maintained on Rat and Mouse No 3 diet; All other mice maintained on Rat and Mouse No1 diet. Both diets purchased from Special Diet Services Ltd, Witham, Essex, UK.	
4) Caging protocols (single or multiple housing)	Acclimatization : 5/cage (single sex); Mating : 2 females : 1 male; Pregnancy and lactation : 1 dam/cage; Weaning (pnd 23) : Litter mates, according to sex; pnd 112 : 3 males/litter individually housed, all remaining males and females continued as group housed litter mates.	
5) Assignment of treatment groups to cage location on racks	In group order, according to stage in pregnancy	
6) Bedding/source	Sawdust (Wood Treatments Ltd, Macclesfield, Cheshire, UK); Shredded paper (SI Supplies, Poynton, Cheshire, UK)	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent	 BPA (Aldrich, Gillingham, Dorset, UK); DES (Sigma, Poole, Dorset, UK); Tocopherol stripped corn oil (TSCO; ICN, Aurora, Ohio, USA) BPA : 99+% pure; DES : 99% pure - Melting points performed : BPA : 158-159ºC; DES : 170-171ºC	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Analyses of dose Formulations Methods of analyses	Retrospective analyses have shown that at least 18mg BPA will dissolve in 1ml peanut oil Solubility of BPA in oil	
8) Age and weight of animals at start and end of study	Females used for mating were 27.1±1.4g and ~10-12 weeks at time of mating; no terminal weights are available. Female pups were terminated at 44 weeks and were 39-44g at termination. Weaning weights also given in Table 2 Male pups terminated at 6 months and were 41-46g at termination. No weaning weights are available.	Table 1 Table 2 Table 3
9) Method of assigning animals to dosed and control groups	Females with confirmed vaginal plug and/or significant body weight gain (ie >3.5g, based on previous experience) were distributed evenly throughout the 4 dosed groups such that equal numbers of females at the same stage in pregnancy were in each group. Females for which the exact day of gestation could not be precisely determined but which were obviously pregnant (as evidenced through body weight gain) were used as naïve controls.	
10) Type of control groups? Concurrent with dosed groups?	Vehicle controls receiving 1ml/kg body weight tocopherol stripped corn oil (TSCO) Naïve controls : no dosing and no handling Yes	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of	Stock solutions of 2mg/ml BPA and DES in TSCO were prepared on 30 Jan 1998. These were homogenised for ~5mins. Serial dilutions of the appropriate stocks were made to give final dosing solutions of 2 and 20µg/ml BPA and 0.2µg/ml DES. These dilutions were prepared on 30 Jan, and 1, 2, 4, 6 Feb 1998. Both stock solutions were homogenised before preparing the dilutions. Dilutions were mixed by vigorous pippetting and shaking. All dosing solutions shaken well before use. DES solutions (both stock and dosing) were wrapped in foil. 30 Jan-8 Feb 1998 Oral BPA : 2µg/kg and 20µg/kg; DES : 0.2µg/kg; TSCO : 1ml/kg. Dosing volume was 1ml/kg body weight for all groups 1 dose per day for 7 days (Gestation days 11-17) between 8.30-10am	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
dosing Light/dark cycle	 12h dark : 12h light	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used. All data are entered as individual entries grouped according to litter. 3 males per litter were individually housed at post-natal day (pnd) 112 through to termination. Identifications are clearly marked in accompanying tables : individually housed males have unique numbers. Remaining males from same litter were given the dam number followed by a letter (males were not ear-punched so as to avoid any unnecessary stress). All females are given a unique number.	Table 4 : VO Table 5 : Male organ weights Table 6 : DSP
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	Between birth and weaning (pnd 23) a total of 51/402 pups were lost, with 2 total litter losses. The losses were spread across all groups and were not considered abnormal. A number of males were lost post-weaning due to fighting	Table 7 : Litter survival data
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	The protocol of Nagel et al (1997) was adhered to as far as was possible. Thus the same strain of animals were used and all forms of stress were attenuated. The only variables were a: the diets used to maintain the animals b : The male pups were isolated at pnd 112 for 71 days rather than at 5 months for 1 month. This was endorsed by one of the authors of the original paper.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate using a single shipment of males and females for mating.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used?	The vast majority of observations were performed by two technicians with a 3rd person occasionally helping out during weaning and also weighing of the female pups. All pms were performed by 2 technicians; each isolating specific tissues ie. One isolated liver, kidneys, testes and epididymides and the other isolated prostate, seminal vesicles and	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Please give details.	coagulating glands.	
18) Were animals and tissue samples examined in a blinded fashion?	Female pups were examined and pm'd in numerical order (see below for more details). Male pups were pm'd as follows. Each group was colour coded thus allowing one animal from each group to be sampled in sequence without prior knowledge of exposure regimen.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Female pups were examined in numerical order. Females were given id numbers according to their birth date. Thus order of examination was : Birth date, Group, id number. Male pups were not examined except for routine gross clinical observations	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	ANOVA : organ and terminal body weights ANCOVA : differences among experimental groups/ dams/ housing of male pups/ correlation of organ weights with body weights. Dunnetts Test : female body weights during puberty. All data assessed by independent statistician	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All female data are complete. A number of males were used for training purposes. These were taken from litters in which either n<3 or, only 1 male remained following individually housing.	Table 8 : Use of males for training or inclusion into main database.

John Ashby

2. Ashby, J., H. Tinwell, et al. (2000). "Current issues in Mutation Research. DNA adducts, estrogenicity and rodent diets." Mutation Research **(in press)**.
3. Ashby, J., H. Tinwell, et al. (2000). "Uterotrophic activity of a "phytoestrogen-free" rat diet." Environmental Health Perspectives **108**(1): A12-A13.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	Comparison of various commercially available rodent diets in their ability to induce a uterotrophic response.	
2) Species, strain, and source of animals	Alpk : Ap$_f$SD rats obtained from Astrazeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, Cheshire UK.	
3) Diet/source	Rat & Mouse No 1 (Special Diet Services Ltd, Witham Essex, UK) AIN-76A & Purina 5001 (Purina Mills Inc., Richmond, IN, USA)	
4) Caging protocols (single or multiple housing)	Maximum of 6 rats/cage. Animals group-housed in numerical order (eg animals 1-5 in one cage, 6-10 in another etc).	Table 1 (5/cage) Table 2 (5/cage) Table 3 (6/cage)
5) Assignment of treatment groups to cage location on racks	In group order	
6) Bedding/source	Sawdust (Wood Treatments Ltd, Macclesfield, Cheshire, UK); Shredded paper (SI Supplies, Poynton, Cheshire, UK)	
7) Chemical analyses: Chemical(s)/source	Faslodex (Gift from Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire, UK) Rat & Mouse No 1 (Special Diet Services Ltd, Witham Essex, UK) AIN-76A & Purina 5001 (Purina Mills Inc., Richmond, IN, USA)	Table 2
Purity of test agent	Faslodex : >99%	
Identified contaminants, %	-	
Stability of test agent		
Analyses of dose Formulations	Several batches of the diets used in Study 3 were analysed for the presence of genistein and diadzein	Table 1
Methods of analyses	Extraction of isoflavones, hydrolysis to aglucones followed by GC-MS	
8) Age and weight of animals at start and end of study	Animals were 20-21 dys upon arrival and given 24h to acclimatize Animals were 21-22 days at start of dosing Animals were 24-25 days at termination See appropriate tables for body weights at start of dosing and at termination	Table 2-4

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
9) Method of assigning animals to dosed and control groups	Randomised by weight sort and Latin square to ensure a random weight distribution	
10) Type of control groups? Concurrent with dosed groups?	Animals maintained on Rat & Mouse No 1 diet Yes	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	Animals were allowed free access to the appropriate diet. Faslodex was prepared in arachis oil (vehicle) to give a final concentration of 10mg/5ml (66.8mg Faslodex + 33.4ml arachis oil). This was prepared and homogenised for ~5mins on 16 August '99 and was stored at +4°C throughout the three day dosing period. 17-19 August 1999 27-30 August 1999 14, 15 & 16 April 2000 Faslodex administered by oral gavage; all other compounds were diets Free access to diets throughout the duration of the experiment. Faslodex : 10mg/kg using a dosing volume of 5ml/kg body weight 1 dose per day for 3 days for Faslodex (Table 1). All diets freely available for 3 days (Tables 1 & 3) or 4 days (Table2). 12h light : 12h dark	 Table 2 Table 3 Table 4
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Unknown	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide	No	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
details		
14) Survival information: were there any early deaths or notable "competing risks"?	All animals survived the duration of the experiment	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Parameter measured (uterine weight) was to determine whether other diets would increase uterine weight (blotted and dry) over that observed for our standard RM1 diet	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Three experiments were performed. Each experiment was done in a single replicate using a single shipment of animals.	Table 2 Table 3 Table 4
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician prepared dosing solutions and provided diets for all 3 experiments. One technician weighed and dosed the animals for each experiment. This was not the same technician for all three studies Two technicians pm'd the animals in each experiment. The same technicians performed the pm'ing for all three studies.	
18) Were animals and tissue samples examined in a blinded fashion?	Groups were colour coded but specific diets unknown. This allowed sequential pm'ing of 2 animals/group in a blind fashion	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals were dosed and clinical observations performed in numerical order. Animals were pm'd 2/group as described above	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available	No	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
information.		
21) What statistical techniques were used to evaluate the data and why?	Standard 2-sided Student's t-test for increase/decreases in uterine (blotted and dry) weights. Comparisons performed using animals exposed to RM1 diet as controls.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data are complete for the groups of interest. Other dose groups were included in these experiments but have no relevance to the data requested.	

John Ashby

4. Ashby, J., H. Tinwell. (2000). "Activity of bisphenol A in pregnant SD and Alpk rats: preliminary data." **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To determine specific sexual maturity parameters (VO, PS, 1st and 2nd oestrus) following exposure to BPA or EE of two strains of rat in utero between GD6 and GD21	
2) Species, strain, and source of animals	Alpk : APfSD (AP) rats purchased from AstraZeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, UK Sprague Dawley (SD) rats purchased from Harlan, UK	
3) Diet/source	Pregnant and lactating females maintained on Rat and Mouse No. 3 diet. Weaned animals maintained on Rat and Mouse No 1 diet. Both purchased from Special Diet Services Ltd., Witham Essex, UK)	
4) Caging protocols (single or multiple housing)	Pregnant/lactating females : 1 dam ± litter/cage Weanlings : housed as littermates according to sex	
5) Assignment of treatment groups to cage location on racks	In group order	
6) Bedding/source	Sawdust (Wood Treatments Ltd, Macclesfield, Cheshire, UK); Shredded paper (SI Supplies, Poynton, Cheshire, UK)	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 BPA purchased from Aldrich (Gillingham, Dorset, UK) EE purchased from Sigma Chemical Company, Poole, Dorset, UK. BPA : 99+ % Pure EE : 98+% Pure - BPA : melting point = 158-159ºC EE : unknown Not performed but samples of dosing solutions stored at -80ºC in glass vials for future analysis. NB. It is known that 25.04mg BPA will dissolve in 1ml AO which is five times greater than the highest dose level used in this study (50mg/10ml).	
8) Age and weight of animals at start and end of study	Dams were approximately 8 weeks old at start of dosing	Table 1 Dam wts Table 2 Pup wts at pnd 2, 5 Table 3 AP pup wts at weaning (pnd 23) & pnd 25
9) Method of assigning animals to dosed and control groups	Body weights at gestation day (GD) 1 and 5 recorded and body weight gain calculated. Animals ranked in order of body weight gain and then assigned to groups in this order whilst also ensuring no significant differences in group mean GD5 weights	Table 4
10) Type of control groups? Concurrent with dosed groups?	Dams dosed with 10ml/kg body weight arachis oil (peanut oil) Yes	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure	 See Table 5 See Table 5 Oral gavage of pregnant dams between GD6 and 21	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Dose levels	AO : 10ml/kg; BPA : 50mg/kg, 100µg/kg; 20µg/kg; EE : 200µg/kg and then 100µg/kg (see Table 4 and the answer to question 15)	
Frequency and duration of dosing	Once daily in the morning (8am - 10am) for 16 days (GD6-GD21)	
Light/dark cycle	12h light : 12h dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Yes. Litters identified by dam no in appropriate tables. NB. Individual pups are not identified with unique numbers until weaning	
13) Was there any "culling" of litters? If so,	Yes	
When?	4 days after birth	
How much?	To give 8 pups/litter consisting of, where possible, 4 males and 4 females	
What was the method of selection?	Random selection although any pups which were unusually large (rare) or small were preferentially culled	
Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	Early deaths occurred in the EE dosed groups in the SD rats (2 removed from study before the end of dosing). Vaginal bleeding, clinical signs of mild toxicity (eg. Pale, hunched, piloerection) and body weight loss observed in both strains of rat when exposed to 200µg/kg EE.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	EE at 200µg/kg was toxic to the dams (both strains). Weight loss during dosing (see Table 1) and the observation of vaginal bleeding led to a reduction in the dose of EE from 200µg/kg to 100µg/kg. This was reduced at GD11 for AP rats and GD14 for SD rats. Several females exposed to EE did not litter. The uterus of females which did not litter were stained with ammonium polysulphide and examined for implantation sites as confirmation of pregnancy (see Table 5). A number of SD dams started to litter on the morning of the last exposure and so did not receive a final dose. This observation occurred in control and BPA dosed groups and so was not considered to be compound related	Table 6 litter details
16) Was the study done in a	Two independent studies were performed, one for each strain of rat. The studies were done	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	using a single shipment of animals in each case and in a single replicate fashion. The individual studies were separated by 1 week.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Two technicians were involved in dosing the rats. One technician performed all AG distance measurements except for 2 SD litters. Culling was performed by the same technician in both studies. One technician prepared all dosing solutions	
18) Were animals and tissue samples examined in a blinded fashion?		
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in group order and, in some cases, according to DOB (litters)	Table 7 AG dists in Sprague Dawley rats Table 8 AG dists in AP rats
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	A preliminary evaluation of the data was performed using a Student's t-test (2-sided)	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Data recorded by Chahoud et al have been used as a comparison to these data as the same strain of rat (SDs) and an identical dosing regime was employed.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as	These studies are on-going and are complete as far as possible. Only data that are absent are those where there were no litters or the dams started to litter before receiving the last dose (SD rats only). AG distances were measured in all pups 24h after birth (pnd 2). The litters were then culled 4 days after birth (pnd 5) hence leading to a reduction in the amount of data	Tables 7 & 8 (AG distances) Table 2 (pup wts at pnd 2 and pnd 5)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
part of this research effort? If some selectivity was involved, please provide the details	recorded thereafter	

John Ashby

5. Odum, J., P. A. Lefevre, et al. (1997). "The rodent uterotrophic assay: critical protocol features, studies with nonylphenols and comparison with a yeast estrogenicity assay." Regulatory Toxicology and Pharmacology **25**: 176-188.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To study the rat uterotrophic assay, investigate the response to a number of previously tested or untested chemicals and compare results with those obtained in the yeast estrogenicity assay.	
2) Species, strain, and source of animals	Alpk : Ap$_f$SD rats obtained from Astrazeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, Cheshire UK.	
3) Diet/source	Rats were weaned on Harlan Teklad TRM diet (Harlan UK, Bicester, Oxfordshire, UK) at 19-21 days of age and were maintained on PCD diet (Special Diet Services Ltd, Witham, Essex, UK) from 21 days of age onwards.	
4) Caging protocols (single or multiple housing)	5 rats/cage	
5) Assignment of treatment groups to cage location on racks	Treatment groups allocated to cages on racks in random order.	
6) Bedding/source	Shredded paper (SI Supplies, Poynton, Cheshire, UK)	
7) Chemical analyses: Chemical(s)/source Purity of test agent	Arachis oil, estradiol (E2) (>98% pure), ethinyl estradiol (EE) (>98% pure) was obtained from Sigma, Poole, Dorset UK. Estradiol benzoate (E2B) as a solution in peanut oil from Intervet, Cambridge UK (purity not specified). ICI 182780 (faslodex) was a gift from Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire, UK). p-Nonylphenol was obtained from two sources: Fluka Chemika-Biochemika [Gillingham, Dorset, UK; 85% para isomers: NP(F)] and Schenectady International [Freeport, Texas, USA; 94.2%: NP(S)]. NMR analysis of these two samples indicated them to be a mixture of branch chain isomers. 4-(n-nonyl)phenol (nNP; 98%) was obtained from Lancaster synthesis (Morecambe, Lancashire, UK).	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Identified contaminants, %	--	
Stability of test agent	Not tested in these studies	
Analyses of dose Formulations	Not performed	
Methods of analyses	N/A	
8) Age and weight of animals at start and end of study	Animals were 20-21 dys upon arrival and given 24h to acclimatize Animals were 21-22 days at start of dosing Animals were 24-25 days at termination See appropriate tables for body weights at start of dosing and at termination	Tables I-V in file "1 Odum RTP 1997.xls"
9) Method of assigning animals to dosed and control groups	Rats were randomised by weight to ensure a random weight distribution through the groups and that the inital group mean body weights were similar for all groups.	
10) Type of control groups?	Arachis oil at a dosing volume of 5ml/kg body weight	
Concurrent with dosed groups?	Yes	
11) Specifics of treatment regimens:		
Formulations/vehicle	All compounds were formulated in arachis oil at 5ml/kg body weight. All compounds were soluble. In Table 4, Group 7, ICI 182,780 was co-administered with NP in the same solution at 5ml/kg.	
Administration dates	See Table VI	Table VI in file "1 Odum RTP 1997.xls"
Route of exposure	sc injection or oral gavage (see Tables I-V)	
Dose levels	See Tables I-V	
Frequency and duration of dosing	All compounds were administered once daily for three days	
Light/dark cycle	12h light : 12h dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal	Not known	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
in the study		
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	No No	
14) Survival information: were there any early deaths or notable "competing risks"?	No deaths	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Uterine wet weight No	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The dose response curves for E2, E2B and EE utilize combined data from 15 studies. The NP dose response experiment was a single study. The four separate NP experiments were 4 studies carried out on separate occasions. The dates of all the studies are given in Table VI and the details of which studies contributed to which dose responses are detailed in Tables 1-V.	Tables I-VI in file "1 Odum RTP 1997.xls"
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician prepared dosing solutions for all 3 experiments. One technician weighed and dosed the animals for each experiment. This was not the same technician for all the studies Two technicians pm'd the animals in each experiment. The same technicians performed the pm'ing for all the studies.	
18) Were animals and tissue samples examined in a blinded fashion?	At PM the group numbers were known but the identities of the groups were not obvious.	
19) Were animals examined in a random order or were they examined in a systematic	At PM, two animals at a time were taken from each cage on the rack sequentially. The cages were distributed on the rack in roughly numerical order.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Analysis of variance was determined using the GLM procedure described in SAS (1989). Differences from control values in all cases were assessed statistically using a 2-sided Student's t-test based on the error mean square from the analysis of variance.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	--	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data are complete for the groups of interest. Other dose groups were included in these experiments but have no relevance to the data requested.	

John Ashby

6. Odum, J. and J. Ashby (1999). "Neonatal exposure of male rats to nonylphenol has no effect on the reproductive tract." Toxicological Science **(in press)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To confirm the report of Lee (1998) that i.p. administration of NP to male rats (up to 8 mg/kg/day) in the neonatal period caused reductions in the weights of male reproductive organs at post-natal day 31, and increased the frequency of cryptorchidism given that neither effect was observed in the earlier rat reproduction study of NP (Chapin et al 1999).	
2) Species, strain, and source of animals	Alpk : Ap_fSD rats obtained from Astrazeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, Cheshire UK.	
3) Diet/source	Pregnant and lactating females maintained on Rat and Mouse No 3 diet; after weaning rats were	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	maintained on Rat and Mouse No1 diet. Both diets purchased from Special Diet Services Ltd, Witham, Essex, UK.	
4) Caging protocols (single or multiple housing)	Pregnancy and lactation : 1 dam/cage. Weaning (pnd 28) : Litter mates, according to sex.	
5) Assignment of treatment groups to cage location on racks	Treatment groups allocated to cages on racks in random order.	
6) Bedding/source	Sawdust (Wood Treatments Ltd, Macclesfield, Cheshire, UK); Shredded paper (SI Supplies, Poynton, Cheshire, UK)	
7) Chemical analyses: Chemical(s)/source	p-Nonylphenol (p-NP) provided by Schenectady International and from the same batch as that tested in a multi-generation study (Chapin et al 1999) and a 90-day toxicity study (Cunny et al, 1997). Faslodex (Gift from Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire, UK). Arachis 0il and DMSO : Sigma, Poole, Dorset UK.	
Purity of test agent	p-NP (>95%); Faslodex : >99%	
Identified contaminants, %	p-NP: nonene (0.2%), lower alkylates (0.1%), o-NP (4%),	
Stability of test agent	Not tested in these studies	
Analyses of dose Formulations	Not performed	
Methods of analyses	N/A	
8) Age and weight of animals at start and end of study	Male pups were 1 day old at start of dosing and 34 (expt 1) or 34-36 (expt 2) days old at termination. Male pups were not given unique identification until the end of the study, up to this time pups were only identified per litter. Mean male pup weights for each litter through the study are given in Tables I and III and terminal body weights for individually identified pups are given in Tables II and IV.	Tables I-IV in file "5 Odum TS 2000.xls"
9) Method of assigning animals to dosed and control groups	Pregnant dams were randomised by weight to ensure a random weight distribution and assigned to the treated groups. All male pups in each litter were dosed according to the group to which the dam had been assigned.	
10) Type of control groups?	Expt 1 : Arachis oil at a dosing volume of 5ml/kg body weight Expt 2 : DMSO at a dosing volume of 1ml/kg body weight	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Concurrent with dosed groups?	Yes	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Expt 1: NP formulated in arachis oil (8mg/kg; 5ml/kg). Readily soluble. Expt 2: NP formulated in DMSO (8mg/kg; 1ml/kg). Readily soluble. Faslodex formulated in DMSO (0.5mg/kg; 1ml/kg). Readily soluble Expt 1: 27 May -5 June 1999 Expt 2: 18 Feb-27 Feb 2000 Intraperitoneal Expt 1: NP 8mg/kg. Expt 2: NP 8mg/kg. Faslodex 0.5mg/kg. Male pups were dosed each day from post natal day 1 to 10 (inclusive) where birth=d0 12h dark : 12h light	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used. The identities of the pups for each dam are given.	Tables II and IV in file "5 Odum TS 2000.xls"
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	No No	
14) Survival information: were there any early deaths or notable "competing risks"?	Pup survival from birth was 82-100%, no compound-related deaths occurred.	
15) Specific variables that are	Testis, Epididymis, Seminal Vesicle, Ventral Prostate weights adjusted for body weights.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis		
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Two separate experiments were performed at separate times. Animals for these experiments arrived as single shipments.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For each study two technicians weighed and dosed the animals. Postmortems were performed by two technicians each isolating specific tissues. The same technician isolated the same tissue for each experiment.	
18) Were animals and tissue samples examined in a blinded fashion?	At PM the group numbers were known but the identities of the groups were not obvious.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	At PM one animal at a time was taken from each cage on the rack sequentially. The cages were distributed on the rack in a random manner. In Expt 1 all animals were PMed on the same day. In Expt 2 animals were killed on 3 sequential days with equal numbers of animals in all groups being examined on each day.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Analysis of variance was determined using the GLM procedure described in SAS (1989). Differences from control values in all cases were assessed statistically using a 2-sided Student's t-test based on the error mean square from the analysis of variance.	
22) Any historical control data	--	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Data on all the animals which were PMed in the two experiments have been provided. All animals were PMed in Expt 1. In Expt 2, twenty-five males from each group were examined at post-mortem out of total numbers of male pups of 29-35 per group, animals closest to the mean group body weights were selected. Time constraints only meant that not all animals were PMed. Data was obtained on testis descent (as described in the paper) but this was not requested here and all animals showed normal testis descent anyway.	

John Ashby

7. Odum, J., I. T. G. Pyrah, et al. (1999). "Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays." Regulatory Toxicology and Pharmacology **29**: 184-195.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To evaluate the response of the Noble rat mammary gland and uterus to nonylphenol. Data from Colerangle and Roy (1996) indicated an approximate 600-fold superior sensitivity of the Noble rat mammary gland, compared to uterotrophic and multigeneration studies conducted in other rat strains. The dosing regime of Colerangle and Roy (1996) (administration via a subcutaneously implanted mini-pump for a period of 11 days) was used in case this was a contributory factor.	
2) Species, strain, and source of animals	Male and female Noble rats were obtained from Charles River USA via Charles River UK Ltd, Margate, Kent, UK. One group were bred at CTL to obtain sufficient females for use in an ovariectomized (OVR^-) uterotrophic assay. Females were ovariectomized at 4-5 weeks of age, and were used after 2-3 weeks recovery. Vaginal smears were taken from all OVR^- rats and only non-cycling animals were used for the uterotrophic assays.	
3) Diet/source	Pregnant and lactating females maintained on Rat and Mouse No 3 diet; after weaning rats were maintained on Rat and Mouse No1 diet. Both diets purchased from Special Diet Services Ltd, Witham, Essex, UK.	
4) Caging protocols (single or multiple housing)	Mating : 1 male to 2 females/cage. Pregnancy and lactation : 1 dam and litter/cage. Other studies: 5 rats/cage (maximum).	
5) Assignment of treatment groups to cage location on racks	Treatment groups allocated to cages on racks in random order.	
6) Bedding/source	Shredded paper (SI Supplies, Poynton, Cheshire, UK)	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
7) Chemical analyses: Chemical(s)/source	p-Nonylphenol (p-NP) provided by Schenectady International and from the same batch as that tested in a multi-generation study (Chapin et al 1999) and a 90-day toxicity study (Cunny et al, 1997). Estradiol (E2) (>98% pure), diethylstilbestrol (DES) (>99% pure), arachis oil and DMSO were obtained from Sigma, Poole, Dorset UK.	
Purity of test agent	p-NP (>95%);	
Identified contaminants, %	p-NP: nonene (0.2%), lower alkylates (0.1%), o-NP (4%),	
Stability of test agent Analyses of dose Formulations	Not tested in these studies Not performed	
Methods of analyses	N/A	
8) Age and weight of animals at start and end of study	Rats used in the first uterotrophic assay (Table 1 in file) were ovariectomized (OVR^-) at 4-5 weeks old, given 2-3 weeks recovery, and were therefore 6-8 weeks old when used in the OVR^- uterotrophic assay (Figure 1 in paper). Rats used in the second (OVR^-) uterotrophic assay (Figure 1 in paper , Table 2 in file) were also 6-8 weeks old when used. The intact rats used in the mammary gland study (Tables 3 onwards in file) were 5-6 weeks old when used.	Tables 1 and 2 in file "7 Odum RTP 1999.xls" Tables 4-8
9) Method of assigning animals to dosed and control groups	Rats were randomised by weight to ensure a random weight distribution through the groups and that the inital group mean body weights were similar for all groups.	
10) Type of control groups? Concurrent with dosed groups?	Control groups for uterotrophic assays received vehicle only. Control groups for mammary gland studies were implanted with minipumps containing DMSO only. Yes in all cases.	
11) Specifics of treatment regimens: Formulations/vehicle	For the uterotrophic assays all compounds were formulated in arachis oil at 5ml/kg body weight. All compounds were readily soluble. For the mammary gland experiment, compounds were dissolved in DMSO: DES 7.6mg/10ml, NP 7.6mg/10ml and NP 5.4g/10ml. Solutions were loaded into osmotic minipumps (Alzet : Type 2002, nominal fill volume 0.2ml and nominal pumping rate 0.5µl/hr; ac tual fill volume 0.242ml and actual pumping rate 0.55µl/hr).	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	BRDU for the mammary gland experiment was prepared in phosphate buffered saline pH7.4 at 2g/100ml and dosed at 5ml/kg body weight.	
Administration dates	Administration dates are given in Tables 1-3.	Tables 1-3 in file "7 Odum RTP 1999.xls"
Route of exposure	For the uterotrophic assays NP and DES were given by oral gavage, E2 was given by sc injection (see Tables 1 and 2). For the mammary gland experiment NP and DES were administered via subcutaneously implanted osmotic minipumps. BRDU was dosed ip.	
Dose levels	For the uterotrophic assays dose levels are given in Tables 1 and 2. The target and achieved dose levels for the mammary gland experiment are in Table 3. BRDU was dosed at 100mg/kg.	Tables 1-3 in file "7 Odum RTP 1999.xls"
Frequency and duration of dosing	For the uterotrophic assays compounds were administered once daily for 3 days (Table 1) or 11 days (Table 2). For the mammary gland experiment NP and DES were dosed via a subcutaneously implanted osmotic minipump for 11 days. BRDU was dosed exactly 2h before termination.	
Light/dark cycle	12h light : 12h dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	No	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection?	No	
Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	No deaths	
15) Specific variables that are	Uterine wet weights, mammary gland differentiation (numbers and areas of mammary gland	Tables 1,2,5-8 in file "7 Odum RTP

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	structures) and mammary gland cell proliferation. In the analysis of mammary gland differentiation it should be noted that not all mammary gland structures are always present eg lobules 3 are not normally present in virgin control animals. This is noted in the paper under "hierarchy of counting data".	1999.xls"
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Three experiments were carried out using 3 shipments of animals. Two OVR^- uterotrophic assays and the mammary gland study. The dates of all the studies are given above the data tables.	Tables 1-3 in file "7 Odum RTP 1999.xls"
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician prepared dosing solutions for all 3 experiments. One technician weighed and dosed the animals for each experiment. This was not the same technician for all the studies One technician surgically implanted all the minipumps in the mammary gland experiment. Two technicians pm'd the animals in each experiment. The same technicians performed the pm'ing for all the studies. Mammary gland structure determinations (numbers and areas) were determined by one person. Mammary gland cell counts (BRDU) were determined by one person (not the same as above).	
18) Were animals and tissue samples examined in a blinded fashion?	At PM the group numbers were known but the identities of the groups were not obvious. Mammary gland structure determinations (numbers and areas) were determined blind, cell counts were not.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	At PM, two animals at a time were taken from each cage on the rack sequentially. The cages were distributed on the rack in roughly numerical order. As mammary gland structure determinations were determined blind, this was completely random. Cell counts were done by taking one animal from each group in turn.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No.	
21) What statistical techniques	Variation within experiments for uterotrophic assays and for determination of cell proliferation	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
were used to evaluate the data and why?	was determined by analysis of variance using the GLM procedure described in SAS (1989). The number and areas of mammary gland structures were analysed using a random effects model, using the "mixed" procedure described in SAS (1989). Differences from control values were assessed statistically using a 2-sided Student's t-test based on the error mean square from the analysis of variance.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	--	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Data for the 2 uterotrophic assays is complete. Some animals were excluded as the vaginal smear data indicated that ovariectomy was incomplete in these animals, these are marked in Tables 1 and 2. The mammary gland data is part of a larger study where an identical experiment was performed on Alpk : Ap_fSD rats, this is reported in paper 3 (Odum, J., Pyrah, I.T.G., Soames, A.R., Foster, J.R., Van Miller, J.P., Joiner, R.L., and Ashby J. (1999). Effects of p-nonylphenol and diethylstillboestrol on the alderley park rat: comparison of mammary glad and uterus sensitivity following oral gavage or implanted mini-pumps. J. Appl. Toxicol., 19, 367-378.) However the Noble rat data is self-contained within this study (ie correct negative and positive controls). The mammary gland preparation for animal 35 was poor and therefore no results were obtained for this animal.	Tables 1& 2 in file "7 Odum RTP 1999.xls" Tables 4-8 in file "7 Odum RTP 1999.xls"

John Ashby

8. Odum, J., I. T. G. Pyrah, et al. (1999). "Effects of p-nonylphenol and diethylstilbestrol on the alderley park rat: comparison of mammary gland and uterus sensitivity following oral gavage or implanted mini-pumps." Journal of Applied Toxicology **19**: 367-378.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To investigate the estrogenic effects of nonylphenol and DES when administered orally or via subcutaneous implants and to compare the comparative sensitivity of the mammary gland and uterus of the AlPk rat.	
2) Species, strain, and source of animals	SD rats obtained from Charles River UK Ltd, Margate, Kent, UK. Alpk : Ap_fSD rats obtained from Astrazeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, Cheshire UK. Where appropriate, females were ovariectomized at 4-5 weeks of age, and were used after 2-3 weeks recovery. Vaginal smears were taken from all OVR^- rats and only non-cycling animals were used for the uterotrophic assays.	
3) Diet/source	Rats were maintained on Rat and Mouse No1 diet (Special Diet Services Ltd, Witham, Essex,	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	UK).	
4) Caging protocols (single or multiple housing)	5 rats/cage (maximum)	
5) Assignment of treatment groups to cage location on racks	Treatment groups allocated to cages on racks in random order.	
6) Bedding/source	Shredded paper (SI Supplies, Poynton, Cheshire, UK)	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 p-Nonylphenol (p-NP) provided by Schenectady International and from the same batch as that tested in a multi-generation study (Chapin et al 1999) and a 90-day toxicity study (Cunny et al, 1997). Estradiol (E2) (>98% pure), diethylstilbestrol (DES) (>99% pure), arachis oil and DMSO were obtained from Sigma, Poole, Dorset UK. p-NP (>95%); p-NP: nonene (0.2%), lower alkylates (0.1%), o-NP (4%), Not tested in these studies Not performed N/A	
8) Age and weight of animals at start and end of study	Rats used in the immature uterotrophic assay (Tables 1&2 in file) were 20-21 dys upon arrival, given 24h to acclimatize and were 21-22 days at start of dosing and 24-25 days at termination (Tables 1&2 in file for body weights at start of dosing and at termination). Rats used in the OVR^- uterotrophic assays (Tables 3&4 in file) were ovariectomized at 4-5 weeks old, given 2-3 weeks recovery, and were therefore 6-8 weeks old when used. The intact rats used in the mammary gland studies (Tables 6 onwards in file) were 5-6 weeks old when used (Tables 7&8 in file for body weights at start of dosing and at termination).	Tables 1&2 in file "3 Odum JAT 1999.xls" Tables 3&4 Tables 7&8
9) Method of assigning animals to dosed and control groups	Rats were randomised by weight to ensure a random weight distribution through the groups and that the initial group mean body weights were similar for all groups.	
10) Type of control groups?	Control groups for uterotrophic assays received vehicle only or were implanted with minipumps containing vehicle only. Control groups for mammary gland studies were implanted with minipumps containing DMSO only or were dosed with vehicle only .	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Concurrent with dosed groups?	Yes in all cases.	
11) Specifics of treatment regimens:		
Formulations/vehicle	For the oral (and E2 sc) uterotrophic assays all compounds were formulated in arachis oil at 5ml/kg body weight. All compounds were readily soluble. For uterotrophic assay experiment 4, compounds were dissolved in DMSO: DES 7.6mg/10ml, NP 7.6mg/10ml and NP 5.4g/10ml. Solutions were loaded into osmotic minipumps (Alzet : Type 2002, nominal fill volume 0.2ml and nominal pumping rate 0.5µl/hr ; actual fill volume 0.242ml and actual pumping rate 0.55µl/hr). For mammary gland experiment 5, compounds were dissolved in DMSO: DES 7.6mg/10ml, NP 7.6mg/10ml and NP 5.4g/10ml. Solutions were loaded into osmotic minipumps (Alzet : Type 2002, nominal fill volume 0.2ml and nominal pumping rate 0.5ml/hr; actual fill volume 0.242ml and actual pumping rate 0.55µl/hr). For mammary gland experiment 6 compounds were formulated in arachis oil at 5µl/kg body weight. All compounds were readily soluble. BRDU for the mammary gland experiments was prepared in phosphate buffered saline pH7.4 at 2g/100ml and dosed at 5ml/kg body weight.	
Administration dates	Administration dates are given above the data Tables 1-6.	Tables 1-6 in file "3 Odum JAT 1999.xls"
Route of exposure	NP and DES were given by oral gavage or via subcutaneously implanted osmotic minipumps, E2 was given by sc injection (see Tables 1 and 2). BRDU was dosed ip.	Tables 1-2 in file "3 Odum JAT 1999.xls"
Dose levels	Dose levels for experiments 1,2,3 & 6 are given in the data tables. The target and achieved dose levels for experiments 4&5 are in Tables 5&6. BRDU was dosed at 100mg/kg.	Tables 1-6 in file "3 Odum JAT 1999.xls"
Frequency and duration of dosing	For experiments 1&2 compounds were administered once daily for 3 days. For experiments 3 & 6 compounds were administered once daily for 11 days. For experiments 4&5 compounds were dosed via subcutaneously implanted osmotic minipumps for 11 days. BRDU was dosed exactly 2h before termination.	
Light/dark cycle	12h light : 12h dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	No	
13) Was there any "culling" of		

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	 No No	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Uterine wet weights, mammary gland differentiation (numbers and areas of mammary gland structures) and mammary gland cell proliferation. In the analysis of mammary gland differentiation it should be noted that not all mammary gland structures are always present eg lobules 3 are not normally present in virgin control animals. This is noted in the paper under "hierarchy of counting data".	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Six experiments were carried out using 6 shipments of animals. Two immature rat uterotrophic assays, 2 OVR^- uterotrophic assays and 2 mammary gland studies. The dates of all the studies are given above the data tables.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician prepared dosing solutions for all 3 experiments. One technician weighed and dosed the animals for each experiment. This was not the same technician for all the studies One technician surgically implanted all the minipumps in the mammary gland experiment. Two technicians pm'd the animals in each experiment. The same technicians performed the pm'ing for all the studies. Mammary gland structure determinations (numbers and areas) were determined by one person. Mammary gland cell counts (BRDU) were determined by one person (not the same as above).	
18) Were animals and tissue samples examined in a blinded fashion?	At PM the group numbers were known but the identities of the groups were not obvious. Mammary gland structure determinations (numbers and areas) were determined blind, cell counts were not.	
19) Were animals examined in a random order or were they	At PM, two animals at a time were taken from each cage on the rack sequentially. The cages were distributed on the rack in roughly numerical order.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	As mammary gland structure determinations were determined blind, this was completely random. Cell counts were done by taking one animal from each group in turn.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No.	
21) What statistical techniques were used to evaluate the data and why?	Variation within experiments for uterotrophic assays and for determination of cell proliferation was determined by analysis of variance using the GLM procedure described in SAS (1989). The number and areas of mammary gland structures were analysed using a random effects model, using the "mixed" procedure described in SAS (1989). Differences from control values were assessed statistically using a 2-sided Student's t-test based on the error mean square from the analysis of variance.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	--	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Data for the 4 uterotrophic assays is complete. Some animals were excluded as the vaginal smear data indicated that ovariectomy was incomplete in these animals, these are marked in Table 3. The mammary gland data is part of a larger study where an identical experiment was performed on Noble rats, this is reported in paper 5 (Odum, J., Pyrah, I.T.G., Foster, J.R., Van Miller, J.P., Joiner, R.L., Ashby, J. (1999). Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays. Reg. Tox. & Pharm., 29: 184-195) However the AlPk rat data is self-contained within this study (ie correct negative and positive controls). Mammary gland H&E preparations for animals 71&74 was poor and therefore no differentiation results were obtained for these animals.	

John Ashby

9. Tinwell, H., R. Joiner, et al. (2000). "Uterotrophic activity of bisphenol A in the immature mouse." Regulatory Toxicology and Pharmacology **(in press)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To determine the activity of BPA in the immature mouse uterotrophic assay including, in some studies, the onset of hyperplasia and hypertrophy.	
2) Species, strain, and source of animals	Mouse : Alpk:AP_fCD-1. Purchased from Astrazeneca Barriered Animal Breeding Unit, Alderley Park, Macclesfield, Cheshire, UK	
3) Diet/source	Rat & Mouse No 1 (Special Diet Services Ltd, Witham Essex, UK)	
4) Caging protocols (single or multiple housing)	Multiple housing, maximum of 6/cage	Table 1 : Caging protocols and treatment regime
5) Assignment of treatment groups to cage location on racks	In group order	
6) Bedding/source	Sawdust (Wood Treatments Ltd, Macclesfield, Cheshire, UK) Shredded paper (SI Supplies, Poynton, Cheshire UK)	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 BPA : Aldrich, Gillingham, Dorset UK DES and Arachis Oil : Sigma, Poole, Dorset UK BPA : 99+% Pure; DES : 99% Pure - Not performed -	
8) Age and weight of animals at start and end of study	All animals were purchased at 19-20 days old and allowed 24h acclimatization. Dosing commenced at 20-21 days old. Body weights of animals are given in the appropriate tables. In Studies 1-5, all females were ~15g on arrival; in Studies 6-9 they were <14g on arrival.	Tables 2-10
9) Method of assigning animals to dosed and control groups	Randomised by weight sort and Latin square to ensure a random weight distribution.	
10) Type of control groups? Concurrent with dosed groups?	Arachis oil at a dosing volume of 5ml/kg body weight Yes	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
11) Specifics of treatment regimens:		
Formulations/vehicle	See Table 1	
Administration dates	See Table 1	Table 1
Route of exposure	See Table 1	
Dose levels	See Table 1	
Frequency and duration of dosing	Arachis Oil, BPA and DES administered once daily for three days by sc injection (Studies 1-8) or oral gavage (Study 9). BudR given in drinking water for entire duration of study, including acclimatisation phase (Studies 6, 8, 9)	
Light/dark cycle	12h light : 12h dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Unknown	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	Majority of mice survived the duration of the study. 1 animal in Study 3 died 1 animal in Study 4 died - these were not thought to be compound related and are noted in the relevant tables	 Table 4 Table 5
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that	Body weight at start of dosing was considered to be of importance. Thus in initial experiments (Studies 1-5) females were 19-20 days old and ~15g on arrival. In the later experiments, the animals were still 19-20 days old on arrival but <14g to ensure that they were less than 18g at start of dosing. Data from any female with a body weight of 18g or more on the first day of dosing are shown in the relevant tables but excluded from the statistical analyses (highlighted	Tables 2-10

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
should be considered in the data analysis	by an * in tables)	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Each individual study (9 in total) was performed in a single replicate with a single shipment of animals	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For each study one technician weighed and dosed the animals. However, the same technician did not do all 9 studies. One technician pm'd all animals in each study. This was the same technician for all 9 studies	
18) Were animals and tissue samples examined in a blinded fashion?	Groups were colour coded but the dose levels were unknown. This allowed sequential pm'ing. Tissue samples for analysis of hypertrophy and hyperplasia were scored blind	Tables 2-10 : uterotrophic data Tables 11-13 : Hyperplasia data Tables 14-16 : Hypertrophy data
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals were dosed and clinical observations were performed in numerical order. Animals were pm'd 2/group according to colour code	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Organ weights : ANOVA and ANOCOVA on final body weight Organ : Body weight ratios : ANOVA Analyses conducted separately comparing BPA and DES treated groups with negative control, arachis oil, group for each individual study. Data for each dose level were combined across all studies (sc injection exposure only) and analysed in a similar fashion. Hyperplasia and Hypertrophy : ANOVA * : $p<0.05$; ** : $p<0.01$ (Tables 17 - 20)	Table 17 : Uterotrophic means & statistical analyses Table 18 : Combined uterotrophic data & statistical analyses Table 19 : Hyperplasia means & statistical analyses

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
		Table 20 : Hypertrophy means & statistical analyses
22) Any historical control data relevant to the interpretation of experimental results should be provided.	-	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Those females with a body weight of 18g or more on the first day of dosing have been included in the accompanying tables but were excluded from the data analyses (identified by an * in relevant tables). Animals 1-36 and 77-84 have not been included in Table 2 (Study 1) as they are not relevant to the requested data (other compounds investigated). Fewer uterii were used in Study 6 (Tables 7, 11 & 14) for histopathology (hyperplasia/hypertrophy) than for the uterotrophic assay because 5 uterii from each group were stored at -70°C for possible future investigations. A number of uterii could not be analysed for hyperplasia and/or hypertrophy. These are noted in the relevant tables (Hyperplasia : Tables 11-13; Hypertrophy : Tables 14-16)	

Barry Delclos

1. Delclos, K. B., T. J. Bucci, et al. (2000). "Effects of dietary genistein exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The objective of the protocol was to obtain data on a range of endpoints to be used to select doses for a multigeneration protocol.	NCTR Protocol E-2122.01. **Data in file "Delclos-Study3 Genistein.xls"**
2) Species, strain, and source of animals	Rat, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	NCTR Protocol E-2122.01.
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content.	NCTR Protocol E-2122.01.
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning; after weaning, pups were housed in same sex pairs	NCTR Protocol E-2122.01.
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	NCTR Protocol E-2122.01.
6) Bedding/source	Hardwood chips, Northeastern Products Corp., Caspian, MI	NCTR

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
7) Chemical analyses: Chemical(s)/source	Genistein, synthesized by Toronto Research Chemicals	
Purity of test agent	>99%	NCTR
Identified contaminants, %		
Stability of test agent	Stable > 6 months	
Analyses of dose Formulations	Wet extraction from chow, then analysis by HPLC/UV. Homogeneity and stability in chow demonstrated.	
Methods of analyses	Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR.	
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7.	NCTR Protocol E-2122.01.
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. 8-10 vaginal plug positive dams were assigned to each dose group to ensure 5 litters for the study. The first five litters with an adequate pup number (see question 12) were continued on the experiment.	NCTR Protocol E-2122.01.
10) Type of control groups? Concurrent with dosed groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	NCTR Protocol E-2122.01.
11) Specifics of treatment regimens: Formulations/vehicle	Compound was blended into 5K96 chow and administered in the diet.	NCTR Protocol E-2122.01.
Administration dates	Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 50.	
Route of exposure		
Dose levels	0, 5, 25, 100, 250, 625, 1250 ppm	
Frequency and duration of dosing	Chow was available *ad libitum*.	
Light/dark cycle	12 light/ 12 dark	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	4 males and 4 females from each litter were used. In the data spreadsheets, the birth mother and nursing mother of each pup are indicated. (While fostering was limited, there were some fostered pups, see below.) In life data was collected on all 4 pups of each sex. Three pups of each sex per litter were used for organ weights and histopath. The remaining pup of each sex per litter were used by Dr. Scallet (Study #9).	NCTR Protocol E-2122.01
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled on PND2 to 4 males and 4 females per litter. Pups were selected randomly, using a random number chart. Fostering was done within dose groups only when necessary to maintain 4 males and 4 females per litter. The pups brought into litters were +/- 1 day of the litter date of birth. For each pup (row) in the spreadsheet, both a birth dam and a nurse dam are indicated. The fostered pups can be identified from these columns (i.e., different birth and nurse dams indicate a fostered pup).	NCTR Protocol E-2122.01
14) Survival information: were there any early deaths or notable "competing risks"?	One female and 2 males, all in the high dose group, died prior to scheduled sacrifice. There were no notable competing risks.	NCTR
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Comparison of prostate weight and preputial separation with EE2 (study 2). Note: The mammary pathology was recently reviewed and the tables included contain the revised diagnoses. The male mammary gland was also affected at a relatively low dose in the EE2 experiment.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate.	NCTR Protocol E-2122.01
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A single technician read the anogenital distances. A separate team of technicians performed necropsies and collected organ weights. Necropsies were supervised by the Study Pathologist. Three pathologists were involved in reading the slides from this study. The initial pathologist left the Center and a second completed the study. The third pathologist reviewed the diagnoses.	NCTR

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
18) Were animals and tissue samples examined in a blinded fashion?	The cages have color-coded stickers indicating dose groups to ensure that the proper dosed feed is placed on each cage. The technicians could thus be aware of the dose group of the animal being examined. Data/observations are entered under the unique animal identifier. Technicians collecting organ weights are not aware of the treatment groups. Slides are not read blind; the procedure used by the pathologists is outlined under question 19.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	The order was random. Necropsies take place over several days and the removals are stratified so that all groups are represented on each day. The following procedure is used for histopathology: High dose and control slides are read first, alternating a few of each. Intermediate doses of reproductive organs and mammary glands are then read. Slides are arranged in an array (compartmentalized box) with control slides in one column, the next higher dose in the adjacent column,etc., across all dose groups. One slide from each dose group is then read across the columns. After this is completed, if there is an apparent effect, each suspect dose is scrambled with the control and the slides are read blind.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Pup organ weights (absolute and body weight ratio) were analyzed separately by sex using a nested mixed model ANOVA that contained dose as a fixed factor and birth dam nested within dose and residual error as random factors. This model adjusts for potential litter effects. Analyses were conducted using both nurse dam and birth dam as blocking factors, with no difference in results. A model using terminal body weight as a covariate was also used. Tests for linear and quadratic dose trends were conducted using contrasts, and 2-sided Dunnett's tests were used to compare treatment group means to the control group means. This approach was used to attempt to determine if any of the endpoints showed U-shaped dose-response curves. Markers of puberty were similarly analyzed by ANOVA. Birth weights were analyzed by ANCOVA with litter size as covariate. For AGD analyses, the litter means were used. The AGD of each pup was read 3 times, but the pups were measured in a random fashion so that the measurements could not be associated with a particular pup. Histopathology data were analyzed for dose effects on incidence and severity with a Jonckheere-Terpstra test; comparisons with control were made using Williams' modification of Shirley's test.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
22) Any historical control data relevant to the interpretation of experimental results should be provided.	The only data that the results can be compared to are those from the ethinyl estradiol and nonylphenol studies (Studies 2 and 3). Those studies were done following an identical protocol with the same control diet.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data collected for the endpoints requested has been provided.	

Barry Delclos

2. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary ethinyl estradiol exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The objective of the protocol was to obtain data on a range of endpoints to be used to select doses for a multigeneration protocol.	NCTR Protocol E-2129.01 **Data in file "Delclos-Study2 EE2.xls"**
2) Species, strain, and source of animals	Rat, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	NCTR Protocol E-2129.01
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content.	NCTR Protocol E-2129.01
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning; after weaning, pups were housed in same sex pairs	NCTR Protocol E-2129.01
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	NCTR Protocol E-2129.01
6) Bedding/source	Hardwood chips, Northeastern Products Corp., Caspian, MI	NCTR
7) Chemical analyses: Chemical(s)/source Purity of test agent	 Ethinyl Estradiol (Lot 57H1178), Sigma Chemical Co., St. Louis, MO >99%	NCTR Protocol E-2129.01

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Identified contaminants, %		
Stability of test agent	Stable > 6 months	
Analyses of dose Formulations	Wet extraction from chow, then analysis by GC/MS. Homogeneity and stability in chow demonstrated.	
Methods of analyses	Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR.	
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7.	NCTR Protocol E-2129.01
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. 8-10 vaginal plug positive dams were assigned to each dose group to ensure 5 litters for the study. The first five litters with an adequate pup number (see question 12) were continued on the experiment.	NCTR Protocol E-2129.01
10) Type of control groups? Concurrent with dosed groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	NCTR Protocol E-2129.01
11) Specifics of treatment regimens:		NCTR Protocol E-2129.01
Formulations/vehicle	Compound was blended into 5K96 chow and administered in the diet.	
Administration dates	Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 50.	
Route of exposure		
Dose levels	0, 0.1, 1, 5, 25, 100, and 200 ppb	
Frequency and duration of dosing	Chow was available ad libitum.	
Light/dark cycle	12 light/ 12 dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	4 males and 4 females from each litter were used. In the data spreadsheets, the birth mother and nursing mother of each pup are indicated. (While fostering was limited, there were some fostered pups, see below.) In life data was collected on all 4 pups of each sex. Three pups of each sex per litter were used for organ weights and histopath. The remaining pup of each sex per litter were used by Dr. Scallet (Study #9).	NCTR Protocol E-2129.01
13) Was there any "culling" of	Litters were culled on PND2 to 4 males and 4 females per litter.	NCTR Protocol E-2129.01

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Pups were selected randomly, using a random number chart. Fostering was done within dose groups only when necessary to maintain 4 males and 4 females per litter. The pups brought into litters were +/- 1 day of the litter date of birth. For each pup (row) in the spreadsheet, both a birth dam and a nurse dam are indicated. The fostered pups can be identified from these columns (i.e., different birth and nurse dams indicate a fostered pup).	
14) Survival information: were there any early deaths or notable "competing risks"?	One female in the 25 ppb group died early. No notable competing risks.	NCTR
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Dorsolateral prostate weight and preputial separation.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate.	NCTR Protocol E-2129.01
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A single technician read the anogenital distances. A separate team of technicians performed necropsies and collected organ weights. Necropsies were supervised by the Study Pathologist.	NCTR
18) Were animals and tissue samples examined in a blinded fashion?	The cages have color-coded stickers indicating dose groups to ensure that the proper dosed feed is placed on each cage. The technicians could thus be aware of the dose group of the animal being examined. Data/observations are entered under the unique animal identifier. Technicians collecting organ weights are not aware of the treatment groups.	NCTR
19) Were animals examined in a random order or were they examined in a systematic	The order was random. Necropsies take place over several days and the removals are stratified so that all groups are represented on each day.	NCTR

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Pup organ weights (absolute and body weight ratio) were analyzed separately by sex using a nested mixed model ANOVA that contained dose as a fixed factor and birth dam nested within dose and residual error as random factors. This model adjusts for potential litter effects. (Earlier studies had separately used nurse dam and birth dam in the analyses, with no difference in results.) A model using terminal body weight as a covariate was also used. Tests for linear and quadratic dose trends were conducted using contrasts, and 2-sided Dunnett's tests were used to compare treatment group means to the control group means. This approach was used to attempt to determine if any of the endpoints showed U-shaped dose-response curves. Markers of puberty were similarly analyzed by ANOVA. Birth weights were analyzed by ANCOVA with litter size as covariate. For AGD analyses, the litter averages were used. The AGD of each pup was read 3 times, but the pups were measured in a random fashion so that the measurements could not be associated with a particular pup.	NCTR
22) Any historical control data relevant to the interpretation of experimental results should be provided.	The only data that the results can be compared to are those from the genistein and nonylphenol studies (Studies 1 and 3). Those studies were done following an identical protocol with the same control diet.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data collected for the endpoints requested has been provided. Some litter weights were not recorded due to a computer malfunction that was not recognized until it was too late. This is the reason for the blank cells under litter/pup weights.	

Barry Delclos

3. Delclos, K. B., J. R. Latendresse, et al. (2000). "Effects of dietary p-nonylphenol exposure during development on male and female CD rats." In prep **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The objective of the protocol was to obtain data on a range of endpoints to be used to select doses for a multigeneration protocol.	NCTR Protocol E-2125.01 **Data in file "Delclos-Study4 NP.xls"**
2) Species, strain, and source of animals	Rat, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	NCTR Protocol E-2125.01
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content.	NCTR Protocol E-2125.01
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning; after weaning, pups were housed in same sex pairs	NCTR Protocol E-2125.01
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	NCTR Protocol E-2125.01
6) Bedding/source	Hardwood chips, Northeastern Products Corp., Caspian, MI	NCTR Protocol E-2125.01
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 p-nonylphenol, Schenectady International, Schenectady, NY Lot # 14081-001 >95% pure o-nonylphenol, 3.8%; dinonylphenol, 0.3% Stable in feed for up to 30 days (within 10% of target dose). Homogeneity demonstrated. HPLC. Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR.	NCTR
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7.	
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. 8-10 vaginal plug positive dams were assigned to each dose group to ensure 5 litters for the study. The first five litters with an adequate pup number (see question 12) were continued on the experiment.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
10) Type of control groups? Concurrent with dosed groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Compound was blended into 5K96 chow and administered in the diet. Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 50. 0, 5, 25, 200, 500, 1000, 2000 ppm Chow was available *ad libitum*. 12 light/ 12 dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	4 males and 4 females from each litter were used. In the data spreadsheets, the birth mother and nursing mother of each pup are indicated. (While fostering was limited, there were some fostered pups, see below.) In life data was collected on all 4 pups of each sex. Three pups of each sex per litter were used for organ weights and histopath. The remaining pup of each sex per litter were used by Dr. Scallet (Study #9).	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled on PND2 to 4 males and 4 females per litter. Pups were selected randomly, using a random number chart. Fostering was done within dose groups only when necessary to maintain 4 males and 4 females per litter. The pups brought into litters were +/- 1 day of the litter date of birth. For each pup (row) in the spreadsheet, both a birth dam and a nurse dam are indicated. The fostered pups can be identified from these columns (i.e., different birth and nurse dams indicate a fostered pup).	
14) Survival information: were there any early deaths or notable "competing risks"?	4 male pups died before the scheduled sacrifice date, one each in dose groups 25, 200, 1000, and 2000 ppm. Animals in the 2000 ppm dose group of both sexes had severe polycystic kidney disease and body weights that were significantly lower than controls. There was an increased incidence of moderate/mild polycystic kidney disease in the 1000 ppm dose group. (Background study #8)	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Male reproductive tract organ weights, epididymal sperm counts. Severe polycystic kidney disease and significantly depressed body weight in the 2000 ppm dose group. Also note that the animals were sacrificed on PND 50 (pubertal males) so epidiymal sperm are just beginning to appear.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A single technician read the anogenital distances. A separate team of technicians performed necropsies and collected organ weights. Necropsies were supervised by the Study Pathologist.	
18) Were animals and tissue samples examined in a blinded fashion?	The cages have color-coded stickers indicating dose groups to ensure that the proper dosed feed is placed on each cage. The technicians could thus be aware of the dose group of the animal being examined. Data/observations are entered under the unique animal identifier. Technicians collecting organ weights are not aware of the treatment groups.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	The order was random. Necropsies take place over several days and the removals are stratified so that all groups are represented on each day.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data	Pup organ weights (absolute and body weight ratio) were analyzed separately by sex using a nested mixed model ANOVA that contained dose as a fixed factor and birth dam nested within	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
and why?	dose and residual error as random factors. This model adjusts for potential litter effects. Analyses were conducted using both nurse dam and birth dam as blocking factors, with no difference in results. A model using terminal body weight as a covariate was also used. Tests for linear and quadratic dose trends were conducted using contrasts, and 2-sided Dunnett's tests were used to compare treatment group means to the control group means. This approach was used to attempt to determine if any of the endpoints showed U-shaped dose-response curves. Markers of puberty and epididymal sperm counts were similarly analyzed.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	The only data that the results can be compared to are those from the ethinyl estradiol and genistein studies (Studies 2 and 1). Those studies were done following an identical protocol with the same control diet. Control epididymal sperm counts from those studies (x 10^6): 5.19 +/- 1.36 (SEM, n=7); 2.50 +/- 0.83 (SEM, n=15)	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data collected for the endpoints requested has been provided.	

Barry Delclos

4. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of genistein in male and female Sprague Dawley rats." **(Unpublished Final Report)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The aim of this study was to establish the immunotoxicity of these compounds in F1 males and females following perinatal exposure	Germolec Zip Disk, Genistein Folder Genistein anti-CD3 data, Genistein bone marrow data
2) Species, strain, and source of animals	RAT, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content	
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning, after weaning, pups were housed in same sex pairs	
5) Assignment of treatment	Random, generated by ROW Sciences, Inc., NCTR	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
groups to cage location on racks		
6) Bedding/source	Hardwood chips, Northeastern Products, Corp., Caspian, MI	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 Genistein, synthesized by Toronto Research Chemicals > 99% Stable > 6 months Wet extraction from chow, then analysis by HPLC/UV. Homogeneity and stability in chow demonstrated. Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR	
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7 and was continued for 65 days. F1 generation male and female rats received the test article in utero for 14 days and continued post partum for 77 days. Terminal body weight data is included for only the AFC subset of animals. Body weight data for the anti-CD3 and bone marrow parameters are included in the data file from NCTR (GENIMMUNO.xls).	GENIMMUNO.xls
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. A total of 32 vaginal plug-postive dams were assigned to each dose group to ensure 20 pregnant dams per dose group. (10 dams per dose were used to assess humoral-mediated immunity, and 10 dams per dose were used to assess cell-mediated immunity, splenic cell quantities, and NK cell activity. These data were not requested.) The goal was to obtain 20 pups per sex per dose from these dams were used to assess 1) humoral-mediated immunity (10/sex/dose), 2) cell mediated immunity, splenic cell quantities, and NK cell activity (10/sex/dose). A separate set of dams were used to generate pups for assessment of bone marrow colony forming units (10/sex/dose). A requirement of the study was that for each of the 3 sets of assays, all animals needed to be sacrificed on a single day. In addition, the sacrifices needed to be spaced to allow for the schedule of the laboratory conducting the assays at the Medical College of Virginia. The priority was to get a sufficient number of animals per dose group per day; the litter of origin was not a primary consideration.	
10) Type of control groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Concurrent with dosed groups?		
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Compound was blended into 5K96 chow and administered in the diet. Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 77. 0, 25, 250 and 1250 ppm Chow was available ad libitum. 12 light/ 12 dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter identities for the CD3 and bone marrow CFU study animals are provided in the file GENIMMUNO.xls.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled to 8 pups per litter on PND 2. For the CD3 proliferation assay, litters were culled to 4 males and 4 females per dam if possible. For the bone marrow assays, litters were standardized to 8 males or 8 females per dam where possible. Random. Cross fostering was done within dose groups to bring the litter to 8 pups where necessary. No fostered pups were used in the CD3 proliferation or bone marrow assays.	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that	Cell proliferation in response to anti-CD3, CFU-GM, CFU-M. No confounders	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
should be considered in the data analysis		
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A separate team of technicians performed necropsies and collected tissues and organ weights. A third set of technicians performed the in vitro immune function assays at the Medical College of Virginia	
18) Were animals and tissue samples examined in a blinded fashion?	Tissue samples for in vitro assays were handled in a blinded fashion.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals were necropsied in the order: control, 25 ppm, 250 ppm, 1,250 ppm.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Data was first evaluated using Bartlett's Test for homogeneity. Homogeneous data were evaluated using ANOVA, and when significant, Dunnett's test was used to compare treated and control groups. Non-homogeneous data were evaluated using a non-parametric ANOVA and Wilcoxon's Rank Test when significant. Jonckheere's test was used to examine dose-related trends among the vehicle and treated groups	
22) Any historical control data relevant to the interpretation of experimental results should be		

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Aberrant values which were not used in the data analysis are flagged in the dataset. Tissue samples which were missing or not used in the in vitro portion of the immune studies are noted in the dataset	

Barry Delclos

5. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of methoxychlor in male and female Sprague Dawley rats." **(Unpublished Final Report)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The aim of this study was to establish the immunotoxicity of these compounds in F1 males and females following perinatal exposure	Data in File: Germolec Zip, Methoxychlor folder: Methoxychlor anti-CD3 data Methoxychlor Bone Marrow
2) Species, strain, and source of animals	RAT, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content	
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning, after weaning, pups were housed in same sex pairs	
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	
6) Bedding/source	Hardwood chips, Northeastern Products, Corp., Caspian, MI	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent	 Methoxychlor, Radian International, Austin, TX > 99% Demonstrated to be stable in diet > 6 months.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Analyses of dose Formulations Methods of analyses	Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR. Dose certifcation was conducted by GC with detection by electron capture.	
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7 and was continued for 65-75 days. F1 generation male and female rats received the test article in utero for 14 days and continued post partum for approximately 80 days. Terminal body weight data is included for only the AFC subset of animals. Body weight data for the anti-CD3 and bone marrow parameters are included in the data file from NCTR.	MXCIMMUNO.xls
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. A total of 32 vaginal plug-postive dams were assigned to each dose group to ensure 20 pregnant dams per dose group. (10 dams per dose were used to assess humoral-mediated immunity, and 10 dams per dose were used to assess cell-mediated immunity, splenic cell quantities, and NK cell activity. These data were not requested.) The goal was to obtain 20 pups per sex per dose from these dams were used to assess 1) humoral-mediated immunity (10/sex/dose), 2) cell mediated immunity, splenic cell quantities, and NK cell activity (10/sex/dose). A separate set of dams were used to generate pups for assessment of bone marrow colony forming units (10/sex/dose). A requirement of the study was that for each of the 3 sets of assays, all animals needed to be sacrificed on a single day. In addition, the sacrifices needed to be spaced to allow for the schedule of the laboratory conducting the assays at the Medical College of Virginia. The priority was to get a sufficient number of animals per dose group per day; the litter of origin was not a primary consideration.	
10) Type of control groups? Concurrent with dosed groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing	 Compound was blended into 5K96 chow and administered in the diet. Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 79-82. 0, 10, 100, 1000 ppm Chow was available ad libitum.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Light/dark cycle	12 light/ 12 dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter identities for the CD3 and bone marrow CFU study animals are provided in the file MXCIMMUNO.xls.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled to 8 pups per litter on PND 2. For the CD3 proliferation assay, litters were culled to 4 males and 4 females per dam if possible. For the bone marrow assays, litters were standardized to 8 males or 8 females per dam where possible. Random. Cross fostering was done within dose groups to bring the litter to 8 pups where necessary. No fostered pups were used in the CD3 proliferation assays. Fostered pups were used in the bone marrow assay as indicated in the file MXCIMMUNO.xls.	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Cell proliferation in response to anti-CD3, CFU-GM, CFU-M. No confounders	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A separate team of technicians performed necropsies and collected tissues and organ weights. A third set of technicians performed the in vitro immune function assays at the Medical College of Virginia	
18) Were animals and tissue	Tissue samples for in vitro assays were handled in a blinded fashion.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
samples examined in a blinded fashion?		
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals were necropsied in the order control, 10 ppm, 100 ppm, 1000 ppm.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	Data was first evaluated using Bartlett's Test for homogeneity. Homogeneous data were evaluated using ANOVA, and when significant, Dunnett's test was used to compare treated and control groups. Non-homogeneous data were evaluated using a non-parametric ANOVA and Wilcoxon's Rank Test when significant. Jonckheere's test was used to examine dose-related trends among the vehicle and treated groups	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Aberrant values which were not used in the data analysis are flagged in the dataset. Tissue samples which were missing or not used in the in vitro portion of the immune studies are noted in the dataset	

Barry Delclos

6. Germolec, D. R., J. A. Munson, et al. (2000). "Immunotoxicity of nonylphenol in male and female Sprague Dawley rats." **(Unpublished Final Report)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The aim of this study was to establish the immunotoxicity of these compounds in F1 males and females following perinatal exposure	Germolec File, Nonylphenol Folder Nonylphenol anti-CD3 data Nonylphenol bone marrow data
2) Species, strain, and source of animals	RAT, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content	
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning, after weaning, pups were housed in same sex pairs	
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	
6) Bedding/source	Hardwood chips, Northeastern Products, Corp., Caspian, MI	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 p-nonylphenol, Schenectady International, Schenectady, NY Lot #14081-001 > 95% o-nonylphenol, 3.8%; dinonylphenol, 0.3% Stable in feed for up to 30 days (within 10% of target dose). Homogeneity demonstrated HPLC Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR	
8) Age and weight of animals at start and end of study	Dams were bred when they were 70-90 days old, approximately 270 g at the start of the experiment. Dosing was begun on gestation day 7 and was continued for 65-72 days. F1 generation male and female rats received the test article in utero for 14 days and continued post partum for 77-82 days. Terminal body weight data is included for only the AFC subset of animals. Body weight data for the anti-CD3 and bone marrow parameters are included in the data file from NCTR (NPIMMUNO.xls).	NPIMMUNO.xls

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
9) Method of assigning animals to dosed and control groups	Random , allocation scheme generated by ROW Sciences, NCTR. A total of 32 vaginal plug-postive dams were assigned to each dose group to ensure 20 pregnant dams per dose group. (10 dams per dose were used to assess humoral-mediated immunity, and 10 dams per dose were used to assess cell-mediated immunity, splenic cell quantities, and NK cell activity. These data were not requested.) The goal was to obtain 20 pups per sex per dose from these dams were used to assess 1) humoral-mediated immunity (10/sex/dose), 2) cell mediated immunity, splenic cell quantities, and NK cell activity (10/sex/dose). A separate set of dams were used to generate pups for assessment of bone marrow colony forming units (10/sex/dose). A requirement of the study was that for each of the 3 sets of assays, all animals needed to be sacrificed on a single day. In addition, the sacrifices needed to be spaced to allow for the schedule of the laboratory conducting the assays at the Medical College of Virginia. The priority was to get a sufficient number of animals per dose group per day; the litter of origin was not a primary consideration.	
10) Type of control groups? Concurrent with dosed groups?	Control group was concurrent with dosed groups. Control animals were fed 5K96 control diet.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Compound was blended into 5K96 chow and administered in the diet. Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 77-82. 0, 25, 500 and 2000 ppm Chow was available ad libitum. 12 light/ 12 dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter identities for the CD3 and bone marrow CFU study animals are provided in the file NPIMMUNO.xls.	
13) Was there any "culling" of litters? If so, When? How much?	Litters were culled to 8 pups per litter on PND 2. For the CD3 proliferation assay, litters were culled to 4 males and 4 females per dam if possible. For the bone marrow assays, litters were standardized to 8 males or 8 females per dam where possible.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
What was the method of selection? Was any cross fostering done? If so, please provide details	Random. Cross fostering was done within dose groups to bring the litter to 8 pups where necessary. No fostered pups were used in the CD3 proliferation assays. One litter for the bone marrow data included fosters, as indicated in the file NPIMMUNO.xls.	
14) Survival information: were there any early deaths or notable "competing risks"?	No. (The parallel reproductive toxicity study, in which animals were killed at PND 50, showed that all animals in the high dose group (2000 ppm) had severe polycystic kidney disease.)	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Cell proliferation in response to anti-CD3, CFU-GM, CFU-M. No confounders	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single replicate	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A team of technicians handled the animals during the in-life phase of the experiments. A separate team of technicians performed necropsies and collected tissues and organ weights. A third set of technicians performed the in vitro immune function assays at the Medical College of Virginia	
18) Were animals and tissue samples examined in a blinded fashion?	Tissue samples for in vitro assays were handled in a blinded fashion.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Necropsies were carried out in the order : 0, 25, 500 and 2000 ppm	
20) Is there information on the	No	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.		
21) What statistical techniques were used to evaluate the data and why?	Data was first evaluated using Bartlett's Test for homogeneity. Homogeneous data were evaluated using ANOVA, and when significant, Dunnett's test was used to compare treated and control groups. Non-homogeneous data were evaluated using a non-parametric ANOVA and Wilcoxon's Rank Test when significant. Jonckheere's test was used to examine dose-related trends among the vehicle and treated groups	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Aberrant values which were not used in the data analysis are flagged in the dataset. Tissue samples which were missing or not used in the in vitro portion of the immune studies are noted in the dataset	

Barry Delclos

7. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of nonylphenol on serum testosterone levels and testicular steroidogenic enzyme activity in neonatal, pubertal, and adult rats." In prep **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To determine if nonylphenol effects on testosterone levels and steroidogenic enzymes (data not included) could contribute to the apparent feminization of hepatic testosterone metabolism and the sexually dimorphic nucleus of the preoptic area observed in the dose range finding study (studies 8 and 9).	NCTR Protocol E2135.11 **Data in file "Delclos-Study6 NPNeonatalTestosterone.xls"**
2) Species, strain, and source of animals	Rat, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	NCTR Protocol E2135.01/E135.l
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content.	NCTR Protocol E2135.01/E135.11
4) Caging protocols (single or	Dams were housed singly. Pups used for the experiment were PND 2 culls.	NCTR Protocol E2135.01/E135.11

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
multiple housing)		
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	NCTR Protocol E2135.01/E135.11
6) Bedding/source	Hardwood chips, P.J. Murphy, Montville, NJ	NCTR
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 p-nonylphenol, Schenectady International, Schenectady, NY Lot # 14081-001 >95% pure o-nonylphenol, 3.8%; dinonylphenol, 0.3% Stable in feed for up to 30 days (within 10% of target dose). Homogeneity demonstrated. HPLC. Purity determination of the test chemical and dose certification of dosed chow was conducted by the Division of Chemistry, NCTR.	 NCTR
8) Age and weight of animals at start and end of study	The animals used were male F1, F2, and F3 pups. Parents of the F1 generation were exposed to dietary nonyphenol from 28 days prior to breeding. Parents of F2 and F3 pups were exposed throughout their lives.	
9) Method of assigning animals to dosed and control groups	F0 generation was weight ranked and randomly assigned to dose groups. In each generation, 35 non-sibling breeding pairs were used and 25 litters randomly selected for continuation on study.	NCTR Protocol E2135.01/E135.11
10) Type of control groups? Concurrent with dosed groups?	The control group was concurrent and maintained on control chow.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels	 Compound was blended into 5K96 chow and administered in the diet. Dams were dosed from GD7, dosing was continued though weaning and pups after weaning were maintained on the same diet as the dam until sacrifice at PND 50. 0, 25, 200, 750 ppm	 NCTR Protocol E2135.01/E135.11

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Frequency and duration of dosing Light/dark cycle	Chow was available *ad libitum*. 12/12	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Pup trunk blood was pooled (2 pups per pool) for each sample. Litter mates and litters are indicated where this is known. Full information is available at this time only for the F1 pups.	NCTR
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	The animals used to generate the data provided were PND 2 culls from up to 12 different litters in each generation. Litters were randomly culled to 4 males and 4 females.	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Pup serum testosterone.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single.	NCTR Protocol E2135.01/E135.11
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used?	Blood was collected by the same team. One individual did the serum testosterone measurements.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Please give details.		
18) Were animals and tissue samples examined in a blinded fashion?	No.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Random order.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	For F1, where litter information was available, an ANOVA with dose as a fixed factor and litter as a random factor nested in dose was conducted. Comparisons to the control group were made with Dunnett's test. For the F2 and F3 generations, where litter information has not been located, one way ANOVA followed by Dunnett's test was used.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	None	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All data requested are provided.	

Barry Delclos

8. Laurenzana, E. M., C. C. Weis, et al. (2000). "Effect of dietarily administered endocrine active agents on hepatic testosterone metabolism, CYP450, and estrogen receptor alpha expression." In prep **(Unpublished Abstract)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To evaluate hepatic testosterone metabolism, CYP450 expression, and ERα levels in liver following exposure to the test compounds.	NCTR **Data in file "Delclos-Study5 Liver TandER.xls"**
2) Species, strain, and source of animals	Rat, CD (Sprague-Dawley), NCTR Strain Code 23, obtained from the NCTR breeding colony	NCTR
3) Diet/source	5K96, an irradiated soy- and alfalfa-free chow obtained from Purina Mills, St. Louis, MO. Chow assayed for genistein and daidzein content.	NCTR
4) Caging protocols (single or multiple housing)	Pregnant dams were housed singly with their litters until weaning; after weaning, pups were housed in same sex pairs	NCTR
5) Assignment of treatment groups to cage location on racks	Random, generated by ROW Sciences, Inc., NCTR	NCTR
6) Bedding/source	Hardwood chips, Northeastern Products Corp., Caspian, MI	NCTR
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
8) Age and weight of animals at start and end of study	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	NCTR
9) Method of assigning animals to dosed and control groups	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	NCTR
10) Type of control groups?	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED	NCTR

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Concurrent with dosed groups?	HERE WERE FROM THOSE STUDIES.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	NCTR
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	4 animals of each sex per dose group were used. Each animal was from a different litter, i.e. litter mates were not used.	NCTR
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Same information as for studies 1, 2, and 3. No fostered pups were used in this study.	NCTR
14) Survival information: were there any early deaths or notable "competing risks"?	Not applicable. See information from previous studies.	NCTR
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that	Testosterone metabolism, ER.	NCTR

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
should be considered in the data analysis		
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Single.	NCTR
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For in-life, same information as for studies 1,2, and 3. Assays on the tissues were conducted by two individuals.	NCTR
18) Were animals and tissue samples examined in a blinded fashion?	No.	NCTR
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Random order.	NCTR
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	One-way ANOVA with tests for linear and quadratic trends. Two-sided Dunnett's test for comparisons to control. Since each animal was from a different litter, no litter adjustment was necessary.	NCTR
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Not applicable.	
23) Are the endocrine response	Yes. There are tissues available from additional animals from these studies, but they were not	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	assayed.	

Barry Delclos

9. Meredith, J. M., C. Bennett, et al. (2000). "Ethinylestradiol and genistein, but not vinclozolin, decrease the volume of the SDN-POA in male rats." Society for Neuroscience Abstracts **(in press)**.

10. Scallet, A. C., C. Bennett, et al. (1999). "Decreased volume of the sexually dimorphic nucleus of the medial preoptic area (SDN-POA) in male rats after chronic nonylphenol exposure." Society for Neuroscience Abstracts **25**: 227.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
2) Species, strain, and source of animals	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
3) Diet/source	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
4) Caging protocols (single or multiple housing)	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
5) Assignment of treatment groups to cage location on racks	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
6) Bedding/source	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
8) Age and weight of animals at	S SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
start and end of study	HERE WERE FROM THOSE STUDIES.	
9) Method of assigning animals to dosed and control groups	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
10) Type of control groups? Concurrent with dosed groups?	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	One male and one female per Litter was randomly assigned to Neurohistology measurements from Delclos' study.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
14) Survival information: were there any early deaths or notable "competing risks"?	SAME INFORMATION AS PROVIDED FOR STUDIES 1, 2, AND 3. ANIMALS USED HERE WERE FROM THOSE STUDIES.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Note that there is a maximum volume and a minimum volume that the brain nucleus can achieve, which constrains the mean volumes that can be observed.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The animals were perfused in several batches according to the sacrifice schedule of Dr. Delclos.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	The same technician perfused and processed all the animals' brains; the same operator performed all the measurements and 3-D reconstruction work.	
18) Were animals and tissue samples examined in a blinded fashion?	Yes.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	A random order of measurement was employed.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No.	
21) What statistical techniques were used to evaluate the data and why?	2- way ANOVA, primarily	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	The manuscript on the 3-D reconstruction and measurement methodology has just been accepted for publication; it includes a review of the literature on hypothalamic measurement issues and historical control information.	
23) Are the endocrine response data provided to us complete in the sense of including all	Some animals were excluded if they failed to meet quality criteria to make them suitable for reliable measurement…these animals are identified in the data tables.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	**Requirements for inclusion of the measurements of an animal's** SDN-POA into the final data set. 1. The embedded paraffin block must produce a set of serial sections spanning the region of interest. 2. The block must produce sections that are consistent in thickness (20 μm). 3. For each animal to be included in the final data set, the tissue section immediately prior to the first section containing its SDN-POA must be present. It must not be wrinkled or damaged in the vicinity of the medial preoptic area, and it must contain no visible evidence of the SDN-POA. 4. For each animal to be included in the final data set, the tissue section immediately following the last section containing its SDN-POA must be present. It must not be wrinkled or damaged in the vicinity of the medial preoptic area, and it must contain no visible evidence of the SDN-POA. 5. The data from animals missing two or more sequential sections that would otherwise be expected (by comparison to intact neighboring sections) to contain portions of the SDN-POA will be discarded. 6. For each animal, if only a single SDN-POA section is missing or otherwise unanalyzable in some way (such as a wrinkle or staining bubble obscuring the SDN-POA), its contribution to the volume will be determined by extrapolation.	

John O'Connor

1. Biegel, L. B., J. C. Cook, et al. (1998). "Effects of 17ß-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats." Toxicological Sciences **44**: 143-154.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The purpose of the 90-day/one-generation reproduction study with 17β-estradiol was to evaluate the significance of a range of responses in order to determine their utility for identifying estrogenic compounds, as well as to provide benchmark data for a risk assessment for chemicals with estrogen-like activities. Ultimately, the data would be used to set dose levels for a future multigeneration reproduction and/or combined chronic toxicity/oncogenicity study in order to further provide a benchmark for a risk assessment of chemicals with estrogen-like effects.	Abstract (page 143) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
2) Species, strain, and source of animals	Crl:CD BR Rats, Charles River Laboratories, Inc. (Raleigh, North Carolina).	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
		one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
3) Diet/source	PMI Feeds, Inc. Certified Rodent Diet 5002.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
4) Caging protocols (single or multiple housing)	Single housing.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
5) Assignment of treatment groups to cage location on racks	After assignment to groups and identified with individual animal identification, they were placed on the rack in numerical order within each treatment group. Animals were rotated on each rack on a bi-weekly basis during the premating phase for the P1 rats and for the entire treatment period for the F1 rats (after weaning). During mating, male and female rats were cohoused in the same cage until mating was confirmed. On fay 14 of gestation, dams were placed into polycarbonate pans and maintained there until the end of lactation.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
6) Bedding/source	While in the stainless steel cages – no bedding. Gestation/lactation – Bed-O-Cobs (Anderson's Company, Maumee, Ohio)	Not in manuscript.
7) Chemical analyses: Chemical(s)/source	17β-Estradiol (catalog #E-8875), 100% purity, was purchased from Sigma Chemical Company (St. Louis, MO).	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
Purity of test agent Identified contaminants, %	The concentration of 17β-estradiol in the diets was determined by HPLC with fluorescence detection (dose levels of 0, 0.05, and 2.5 ppm) or HPLC with UV detection (10 and 50 ppm) three times during the study – study start, during the 90-day phase, and during the F1 generation feeding phase.	
Stability of test agent	Stability was determined at the beginning and at the end of the study.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Analyses of dose Formulations Methods of analyses		
8) Age and weight of animals at start and end of study	Animals were approximately 42 days at study start. Mean body weights (± SD) upon arrival (35 days of age) were 123.9 ± 18.03g for males and 117.9 ± 11.28g for females. Body weights from the end of the study are available in the manuscripts and the raw data that is included.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
9) Method of assigning animals to dosed and control groups	Rats were divided by computerized, stratified randomization into male and female treatment groups so that there were no statistically significant differences among group body weight means within each sex.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
10) Type of control groups? Concurrent with dosed groups?	The concurrent control groups (male and female) were fed diet prepared in the same manner as the treatment groups. Control diets were prepared with 0.005% acetone only (see below).	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of	17β-estradiol was added to rodent chow and thoroughly mixed for approximately 6 minutes in a high-speed Hobart mixer to assure homogeneous distribution in the diet. 17β-estradiol was dissolved in acetone (0.005% acetone) prior to diet preparation to aid in homogeneous distribution of the test substance. Control diets were prepared with 0.005% acetone only. The route of administration was feeding (ad libitum). The diet was administered continuously (i.e., 7 days week) from July 30, 1996 – February 2, 1997 at dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm.	Materials and Methods (page 144) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
dosing Light/dark cycle	 Animal rooms were maintained on a 12 hour light/dark cycle (light approximately 0600-1800 hr).	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used for some of the analyses; however, these data were not easily accessible).	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled randomly to 8 pups (4/sex/litter where possible) on day 4 postpartum. No cross-fostering was performed.	Materials and Methods (page 119) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
14) Survival information: were there any early deaths or notable "competing risks"?	No.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Within the context of this study (which included this publication and the other two manuscripts under consideration that referenced the same study), the data were characteristic of a typical dose-response effect of 17β-estradiol. The only real issues with these studies (i.e., the 90-day/1-generation reproduction study) was determining the LOEL/NOEL. In the males, the LOEL was 2.5 ppm (NOEL = 0.05 ppm) based on multiple findings. In the females the LOEL/NOEL was not clearly established due to a slight hastening of vaginal patency in the 0.05 ppm group. The authors concluded that this was most likely not a compound-related effect.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The study was performed using a single shipment of animals.	
17) Did the same technician examine and measure dosed and control animals, or were	For all aspects of the study (i.e., collection of body weights, clinical signs, food consumption data, necropsy, etc) multiple technicians were used to collect and/or analyze the data. All technicians that participated on particular aspect of a study were trained to perform that specific	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
multiple technicians used? Please give details.	task according to Haskell training requirements. For any one day of data collection or assay, one technician would have collected the data for all treatment groups.	
18) Were animals and tissue samples examined in a blinded fashion?	Histopathology evaluation was performed in a blind manner. All other parameters were performed with knowledge of the treatment level.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	All animals/samples that were examined with knowledge of the treatment groups were examined in a systematic fashion. For example, for the organ weights collected at necropsy, animals were euthanized so that time of euthanization would not contribute to additional variability. In the case of organ weights, animals were euthanized across treatment groups (i.e., the order of euthanization would be in the following order assuming increasing numbers designate increased treatment levels – 101, 201, 301, 401, 501, 102, 202, 302, 402, 502, etc.) to minimize any effects from time of euthanization. Similarly, for staining for cell proliferation, all slides from 1 particular timepoint inclusive of all treatment groups (i.e., 1 week timepoint) would be prepared/stained together. Anogenital distance, vaginal patency, and preputial separation measurements were collected for each particular treatment group in sequential order (i.e., group I, followed by group II, followed by group III, etc.). For hormonal analyses, all treatment groups would be analyzed simultaneously for each particular hormone analyzed.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	This information was not available on this study due to the study design.	
21) What statistical techniques were used to evaluate the data and why?	The statistical analyses that were used for the data are in accordance with those utilized by Haskell laboratory. They were as follows: Hormonal data – Jonckheere's trend test Estrous cycle data – Jonckheere's trend test	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Historical control data for most of these endpoints are routinely collected at Haskell. However, since the historical control data were not used in the interpretation of the results from the current study, these historical control data were not provided. The one area where the historical control data would have been useful, namely, for the vaginal patency and preputial separation data, we did not have historical control data.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If	All animals for which data was collected were reported.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
some selectivity was involved, please provide the details		

John O'Connor

2. Biegel, L. B., J. A. Flaws, et al. (1998). "90-day feeding and one-generation reproduction study in Crl:CD BR rats with 17ß-estradiol." Toxicological Sciences **44**: 116-142.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The purpose of the 90-day/one-generation reproduction study with 17β-estradiol was to evaluate the significance of a range of responses in order to determine their utility for identifying estrogenic compounds, as well as to provide benchmark data for a risk assessment for chemicals with estrogen-like activities. Ultimately, the data would be used to set dose levels for a future multigeneration reproduction and/or combined chronic toxicity/oncogenicity study in order to further provide a benchmark for a risk assessment of chemicals with estrogen-like effects.	Abstract (page 116)
2) Species, strain, and source of animals	Crl:CD BR Rats, Charles River Laboratories, Inc. (Raleigh, North Carolina).	Materials and Methods (page 117)
3) Diet/source	PMI Feeds, Inc. Certified Rodent Diet 5002.	Materials and Methods (page 117)
4) Caging protocols (single or multiple housing)	Single housing.	Materials and Methods (page 118)
5) Assignment of treatment groups to cage location on racks	After assignment to groups and identified with individual animal identification, they were placed on the rack in numerical order within each treatment group. Animals were rotated on each rack on a bi-weekly basis during the premating phase for the P1 rats and for the entire treatment period for the F1 rats (after weaning). During mating, male and female rats were co-housed in the same cage until mating was confirmed. On day 14 of gestation, dams were placed into polycarbonate pans with bedding and maintained there until the end of lactation.	Materials and Methods (page 117 & 118)
6) Bedding/source	While in the stainless steel cages – no bedding. Gestation/lactation – Bed-O-Cobs (Anderson's Company, Maumee, Ohio)	Not in manuscript.
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, %	17β-Estradiol (catalog #E-8875), 100% purity, was purchased from Sigma Chemical Company (St. Louis, MO). The concentration of 17β-estradiol in the diets was determined by HPLC with fluorescence detection (dose levels of 0, 0.05, and 2.5 ppm) or HPLC with UV detection (10 and 50 ppm) three times during the study – study start, during the 90-day phase, and during the F1 generation feeding phase.	Materials and Methods (page 117)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Stability of test agent Analyses of dose Formulations Methods of analyses	Stability was determined at the beginning and at the end of the study.	
8) Age and weight of animals at start and end of study	Animals were approximately 42 days at study start. Mean body weights (± SD) upon arrival (35 days of age) were 123.9 ± 18.03g for males and 117.9 ± 11.28g for females. Body weights from the end of the study are available in the manuscripts and the raw data that is included.	Materials and Methods (page 117)
9) Method of assigning animals to dosed and control groups	Rats were divided by computerized, stratified randomization into male and female treatment groups so that there were no statistically significant differences among group body weight means within each sex.	Materials and Methods (page 117 & 118)
10) Type of control groups? Concurrent with dosed groups?	The concurrent control groups (male and female) were fed diet prepared in the same manner as the treatment groups. Control diets were prepared with 0.005% acetone only (see below).	Materials and Methods (page 117)
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	17β-estradiol was added to rodent chow and thoroughly mixed for approximately 6 minutes in a high-speed Hobart mixer to assure homogeneous distribution in the diet. 17β-estradiol was dissolved in acetone (0.005% acetone) prior to diet preparation to aid in homogeneous distribution of the test substance. Control diets were prepared with 0.005% acetone only. The route of administration was feeding (ad libitum). The diet was administered continuously (i.e., 7 days week) from July 30, 1996 – February 2, 1997 at dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm. Animal rooms were maintained on a 12 hour light/dark cycle (light approximately 0600-1800 hr).	Materials and Methods (page 117)
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used for some of the analyses; however, these data were not easily accessible.	
13) Was there any "culling" of litters? If so, When?	Litters were culled randomly to 8 pups (4/sex/litter where possible) on day 4 postpartum.	Materials and Methods (page 119)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
How much? What was the method of selection? Was any cross fostering done? If so, please provide details	No cross-fostering was performed.	
14) Survival information: were there any early deaths or notable "competing risks"?	No.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Within the context of this study (which included this publication and the other two manuscripts under consideration that referenced the same study), the data were characteristic of a typical dose-response effect of 17β-estradiol. The only real issues with these studies (i.e., the 90-day/1-generation reproduction study) was determining the LOEL/NOEL. In the males, the LOEL was 2.5 ppm (NOEL = 0.05 ppm) based on multiple findings. In the females the LOEL/NOEL was not clearly established due to a slight hastening of vaginal patency in the 0.05 ppm group. The authors concluded that this was most likely not a compound-related effect.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The study was performed using a single shipment of animals.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For all aspects of the study (i.e., collection of body weights, clinical signs, food consumption data, necropsy, etc) multiple technicians were used to collect and/or analyze the data. All technicians that participated on particular aspect of a study were trained to perform that specific task according to Haskell training requirements. For any one day of data collection or assay, one technician would have collected the data for all treatment groups.	
18) Were animals and tissue samples examined in a blinded fashion?	Histopathology evaluation was performed in a blind manner. All other parameters were performed with knowledge of the treatment level.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	All animals/samples that were examined with knowledge of the treatment groups were examined in a systematic fashion. For example, for the organ weights collected at necropsy, animals were euthanized so that time of euthanization would not contribute to additional variability. In the case of organ weights, animals were euthanized across treatment groups (i.e., the order of euthanization would be in the following order assuming increasing numbers designate increased treatment levels – 101, 201, 301, 401, 501, 102, 202, 302, 402, 502, etc.) to minimize any effects from time of euthanization. Similarly, for staining for cell proliferation,	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	all slides from 1 particular timepoint inclusive of all treatment groups (i.e., 1 week timepoint) would be prepared/stained together. Anogenital distance, vaginal patency, and preputial separation measurements were collected for each particular treatment group in sequential order (i.e., group I, followed by group II, followed by group III, etc.). For hormonal analyses, all treatment groups would be analyzed simultaneously for each particular hormone analyzed.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	This information was not available on this study due to the study design.	
21) What statistical techniques were used to evaluate the data and why?	The statistical analyses that were used for the data are in accordance with those utilized by Haskell laboratory. They were as follows: Organ weight – ANOVA/pairwise comparisons Cell proliferation – ANOVA/pairwise comparisons Pup numbers/pup survival – Jonckheere's trend test Reproductive indices – Cochran-armitage trend test Developmental landmarks – ANOVA/pairwise comparisons	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Historical control data for most of these endpoints are routinely collected at Haskell. However, since the historical control data were not used in the interpretation of the results from the current study, these historical control data were not provided. The one area where the historical control data would have been useful, namely, for the vaginal patency and preputial separation data, we did not have historical control data.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All animals for which data was collected were reported.	

John O'Connor

3. Cook, J. C., L. Johnson, et al. (1998). "Effects of dietary 17ß-estradiol exposure on serum hormone concentrations and testicular parameters in male Crl:CD BR rats." Toxicological Sciences **44**: 155-168.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The purpose of the 90-day/one-generation reproduction study with 17β-estradiol was to evaluate the significance of a range of responses in order to determine their utility for identifying estrogenic compounds, as well as to provide benchmark data for a risk assessment for chemicals with estrogen-like activities. Ultimately, the data would be used to set dose levels for a future multigeneration reproduction and/or combined chronic toxicity/oncogenicity study in order to further provide a benchmark for a risk assessment of chemicals with estrogen-like effects.	Abstract (page 155) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
2) Species, strain, and source of animals	Crl:CD BR Rats, Charles River Laboratories, Inc. (Raleigh, North Carolina).	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
3) Diet/source	PMI Feeds, Inc. Certified Rodent Diet 5002.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
4) Caging protocols (single or multiple housing)	Single housing.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
5) Assignment of treatment groups to cage location on racks	After assignment to groups and identified with individual animal identification, they were placed on the rack in numerical order within each treatment group. Animals were rotated on each rack on a bi-weekly basis during the premating phase for the P1 rats and for the entire treatment period for the F1 rats (after weaning). During mating, male and female rats were cohoused in the same cage until mating was confirmed. On day 14 of gestation, dams were	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	placed into polycarbonate pans and maintained there until the end of lactation.	estradiol. Toxicological Sciences 44, 116-142]
6) Bedding/source	While in the stainless steel cages – no bedding. Gestation/lactation – Bed-O-Cobs (Anderson's Company, Maumee, Ohio)	Not in manuscript.
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	17β-Estradiol (catalog #E-8875), 100% purity, was purchased from Sigma Chemical Company (St. Louis, MO). The concentration of 17β-estradiol in the diets was determined by HPLC with fluorescence detection (dose levels of 0, 0.05, and 2.5 ppm) or HPLC with UV detection (10 and 50 ppm) three times during the study – study start, during the 90-day phase, and during the F1 generation feeding phase. Stability was determined at the beginning and at the end of the study.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
8) Age and weight of animals at start and end of study	Animals were approximately 42 days at study start. Mean body weights (± SD) upon arrival (35 days of age) were 123.9 ± 18.03g for males and 117.9 ± 11.28g for females. Body weights from the end of the study are available in the manuscripts and the raw data that is included.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
9) Method of assigning animals to dosed and control groups	Rats were divided by computerized, stratified randomization into male and female treatment groups so that there were no statistically significant differences among group body weight means within each sex.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
10) Type of control groups? Concurrent with dosed groups?	The concurrent control groups (male and female) were fed diet prepared in the same manner as the treatment groups. Control diets were prepared with 0.005% acetone only (see below).	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
		estradiol. Toxicological Sciences 44, 116-142]
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	17β-estradiol was added to rodent chow and thoroughly mixed for approximately 6 minutes in a high-speed Hobart mixer to assure homogeneous distribution in the diet. 17β-estradiol was dissolved in acetone (0.005% acetone) prior to diet preparation to aid in homogeneous distribution of the test substance. Control diets were prepared with 0.005% acetone only. The route of administration was feeding (ad libitum). The diet was administered continuously (i.e., 7 days week) from July 30, 1996 – February 2, 1997 at dietary concentrations of 0, 0.05, 2.5, 10, or 50 ppm. Animal rooms were maintained on a 12 hour light/dark cycle (light approximately 0600-1800 hr).	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used for some of the analyses; however, these data were not easily accessible.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled randomly to 8 pups (4/sex/litter where possible) on day 4 postpartum. No cross-fostering was performed.	Materials and Methods (page 157) and primary manuscript [Biegel et al., (1998) 90-Day Feeding and one-generation reproduction study in Crl:CD BR rats with 17β-estradiol. Toxicological Sciences 44, 116-142]
14) Survival information: were there any early deaths or notable "competing risks"?	No.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially	Within the context of this study (which included this publication and the other two manuscripts under consideration that referenced the same study), the data were characteristic of a typical dose-response effect of 17β-estradiol. The only real issues with these studies (i.e., the 90-day/1-generation reproduction study) was determining the LOEL/NOEL. In the males, the LOEL was	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
"confounding variables" that should be considered in the data analysis	2.5 ppm (NOEL = 0.05 ppm) based on multiple findings. In the females the LOEL/NOEL was not clearly established due to a slight hastening of vaginal patency in the 0.05 ppm group. The authors concluded that this was most likely not a compound-related effect.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The study was performed using a single shipment of animals.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For all aspects of the study (i.e., collection of body weights, clinical signs, food consumption data, necropsy, etc) multiple technicians were used to collect and/or analyze the data. All technicians that participated on particular aspect of a study were trained to perform that specific task according to Haskell training requirements. For any one day of data collection or assay, one technician would have collected the data for all treatment groups.	
18) Were animals and tissue samples examined in a blinded fashion?	Histopathology evaluation was performed in a blind manner. All other parameters were performed with knowledge of the treatment level.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	All animals/samples that were examined with knowledge of the treatment groups were examined in a systematic fashion. For example, for the organ weights collected at necropsy, animals were euthanized so that time of euthanization would not contribute to additional variability. In the case of organ weights, animals were euthanized across treatment groups (i.e., the order of euthanization would be in the following order assuming increasing numbers designate increased treatment levels – 101, 201, 301, 401, 501, 102, 202, 302, 402, 502, etc.) to minimize any effects from time of euthanization. Similarly, for staining for cell proliferation, all slides from 1 particular timepoint inclusive of all treatment groups (i.e., 1 week timepoint) would be prepared/stained together. Anogenital distance, vaginal patency, and preputial separation measurements were collected for each particular treatment group in sequential order (i.e., group I, followed by group II, followed by group III, etc.). For hormonal analyses, all treatment groups would be analyzed simultaneously for each particular hormone analyzed.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	This information was not available on this study due to the study design.	
21) What statistical techniques were used to evaluate the data	The statistical analyses that were used for the data are in accordance with those utilized by Haskell laboratory. They were as follows:	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
and why?	Organ weight – ANOVA/pairwise comparisons Sperm parameters - Jonckheere's trend test Hormonal data – Jonckheere's trend test Estrous cycle data – Jonckheere's trend test	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Historical control data for most of these endpoints are routinely collected at Haskell. However, since the historical control data were not used in the interpretation of the results from the current study, these historical control data were not provided. The one area where the historical control data would have been useful, namely, for the vaginal patency and preputial separation data, we did not have historical control data.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All animals for which data was collected were reported.	

John O'Connor

4. O'Connor, J. C., S. R. Frame, et al. (1998). "Sensitivity of a tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17ß-estradiol." Toxicological Sciences **44**: 169-184.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	This manuscript is one piece in a project to develop short-term models to screen compounds for potential endocrine activity. 17β-Estradiol was one compound that was used to evaluate the proposed Tier I screening battery from identifying compounds with estrogenic activity.	Abstract (page 169)
2) Species, strain, and source of animals	Crl:CD BR Rats, Charles River Laboratories, Inc. (Raleigh, North Carolina).	Materials and Methods (page 171)
3) Diet/source	PMI Feeds, Inc. Certified Rodent Diet 5002.	Materials and Methods (page 171)
4) Caging protocols (single or multiple housing)	Single housing.	Materials and Methods (page 171)
5) Assignment of treatment groups to cage location on racks	After assignment to groups and identified with individual animal identification, they were placed on the rack in numerical order within each treatment group. Due to the short duration (2 weeks) of treatment, animals were not rotated on each rack during the the study.	Materials and Methods (page 171)
6) Bedding/source	While in the stainless steel cages – no bedding.	Materials and Methods (page 171)
7) Chemical analyses:	17β-Estradiol (catalog #E-8875), 100% purity, was purchased from Sigma Chemical Company	Not discussed in manuscript.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	(St. Louis, MO). The concentration/stability of 17β-estradiol in the dosing solutions was not analyzed.	
8) Age and weight of animals at start and end of study	Male rats were approximately 70 days old at study start and were requested to be between 260 and 300 grams body weight upon receipt (63 days of age). Females were approximately 49 days old at study start. Body weights from the end of the study are available in the manuscripts and the raw data that is included.	Materials and Methods (page 171)
9) Method of assigning animals to dosed and control groups	Rats were divided by computerized, stratified randomization into male and female treatment groups so that there were no statistically significant differences among group body weight means within each sex.	Materials and Methods (page 171)
10) Type of control groups? Concurrent with dosed groups?	The concurrent control groups (male and female) were administered dose solutions of the vehicle (0.25% methylcellulose) alone at the same dose volume as that received by the treated animals.	Materials and Methods (page 171)
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	17β-estradiol was prepared by mixing in 0.25% methylcellulose vehicle and was stirred constantly during dosing to ensure homogeneity. Control groups received 0.25% methylcellulose vehicle only. The route of administration was intraperitoneal injection (dose volume 2 ml/kg). The dose solutions were administered continuously (i.e., 7 days week) from September 17, 1996 – October 1, 1996 at dose concentrations of 0, 0.001, 0.0025, 0.0075, or 0.05 mg/kg/day. Animal rooms were maintained on a 12 hour light/dark cycle (light approximately 0600-1800 hr).	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	No litters were produced in this study.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection?	No litters were produced in this study.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Was any cross fostering done? If so, please provide details		
14) Survival information: were there any early deaths or notable "competing risks"?	No.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	For this study, the data show that the *in vivo* screening batteries can detect 17β-estradiol-induced changes at lower levels of exposure than were observed in the 90-day/1-generation reproduction study. Similar to the previous study, the data from the current *in vivo* screening studies were indicative of a typical dose-response effect.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The study was performed using a single shipment of animals.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	For all aspects of the study (i.e., collection of body weights, clinical signs, food consumption data, necropsy, etc) multiple technicians were used to collect and/or analyze the data. All technicians that participated on particular aspect of a study were trained to perform that specific task according to Haskell training requirements. For any one day of data collection or assay, one technician would have collected the data for all treatment groups.	
18) Were animals and tissue samples examined in a blinded fashion?	Histopathology evaluation was performed in a blind manner. All other parameters were performed with knowledge of the treatment level.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	All animals/samples that were examined with knowledge of the treatment groups were examined in a systematic fashion. For example, for the organ weights collected at necropsy, animals were euthanized so that time of euthanization would not contribute to additional variability. In the case of organ weights, animals were euthanized across treatment groups (i.e., the order of euthanization would be in the following order assuming increasing numbers designate increased treatment levels – 101, 201, 301, 401, 501, 102, 202, 302, 402, 502, etc.) to minimize any effects from time of euthanization. Similarly, for staining for cell proliferation, all slides from 1 particular timepoint inclusive of all treatment groups (i.e., 1 week timepoint) would be prepared/stained together. For hormonal analyses, all treatment groups would be analyzed simultaneously for each particular hormone analyzed.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	This information was not available on this study due to the study design.	
21) What statistical techniques were used to evaluate the data and why?	The statistical analyses that were used for the data are in accordance with those utilized by Haskell laboratory. They were as follows: Organ weight – ANOVA/pairwise comparisons Hormonal data – Jonckheere's trend test Uterine cell prolif./cell height - Jonckheere's trend test Estrogen receptor data – Jonckheere's trend test	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Historical control data for some of the endpoints was available (means and standard deviations) and will be provided in the spreadsheet.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All animals for which data was collected were reported.	

Frederick vom Saal

1. Alworth, L. C., K. L. Howdeshell, et al. (1999). Uterine response to estradiol: low-dose facilitation and high-dose inhibition due to fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Paper presented at the Environmental Hormones meeting, Tulane University, New Orleans, October.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	We examined whether the low-dose stimulating effect and high-dose inhibitory effect of DES and methoxychlor (MXC) previously observed in the prostate of male offspring also occurred in the uterus of female offspring.	ALL DATA WILL BE IN THE MANUSCRIPT
2) Species, strain, and source of animals	CD-1 Swiss mice, initially purchased from charles river in 1998 and outbred in my lab	
3) Diet/source	pregnancy&lactation - Purina 5008 post weaning - purina 5001	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
4) Caging protocols (single or multiple housing)	At both 7 and 8 months of age, fenakes housed in same-sex litter groups were randomly selected to be ovariectomized and implanted with estradiol, at which time they were individually housed for one week.	
5) Assignment of treatment groups to cage location on racks	we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	We did not check the purity of DES or MXC used in this experiment DES = Sigma MXC - technical grade, Kincaid Laboratories	
8) Age and weight of animals at start and end of study	Experiment 1 =7 months old Experiment 2 = 8 months old Body weights of the animals in the different groups were recorded at the time of death	
9) Method of assigning animals to dosed and control groups	random	
10) Type of control groups? Concurrent with dosed groups?	unhandled dams oil (vehicle) fed dams concurrent with dosed groups for MXC oil injected (s.c.) for the DES group Controls were concurrent with dosed groups	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure	Plug positive mice were either left unhandled or they were administered orally (in 30 µl oil) an average dose of 0, 10 or 10,000 µg/kg methoxychlor from gestation day 12-18 (Plug = day 0). Oil vehicle: ICN Biomedicals, Inc. Aurora, OH. Tocopherol, stripped. 8001-30-7,Lot 95315, cat n 901415 Other pregnant mice were injected s.c. with an average dose of 0.1 and 100 µg/kg DES dissolved in 40 µl of tocopherol-stripped corn oil (Cat# 901415, ICN, Aurora, OH), which was administered once a day during gestational days 12-18. ADMINISTRATION DATES - August, 1998 FED WITH A PIPETTER - . The corn oil was administered by electronic micorpipetter	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Dose levels Frequency and duration of dosing Light/dark cycle	(Rainen Instruments) because mice readily consume corn oil that is pipetted into their mouth, and this procedure is not as stressful as gavage (force feeding by stomach tube). CF-1 mice (Mus musculus domesticus) were obtained from Charles River Laboratories (Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond, IN) and, after weaning, females were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{o}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	7-month old experiment = one female per litter was used to control for litter effects. 8-month old experiment = all of the remaining females were used. They were randomly assigned to groups, but some groups had more than one female from a litter, thus requiring statistical correction for litter effects.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no to all questions	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	uterine weight liver weight (very interesting findings with low doses of DES) no confounding variables	
16) Was the study done in a single "replicate" with a single	one group of animals all produced and examined together	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
shipment of animals, or whether multiple replicates or shipments were used. Please give details		
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	one graduate student (a DVM working on a masters associated with a lab animal medicine postdoc) did this experiment by herself. But, at tissue collection, the whole lab participates. The student (Allwort) removed the organs, but animals were passed to her so she would do the dissection blind to the group of the animal. Another student weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	yes - see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in random order without knowledge of experimental group	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	NO, natural delivery	
21) What statistical techniques were used to evaluate the data and why?	ANOVA and ANCOVA was used depending on whether body weight was significantly related to the organ being analyzed.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have historical control data from animals from females administered a wider range of doses of estradiol to determine the dose that maximally stimulates the uterus. These data will be presented at the meeting.	
23) Are the endocrine response data provided to us complete in the sense of including all	There was absolutely no selectivity - all animals used are included in the data set to increase power	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details		

Frederick vom Saal

4. Nagel, S. C., F. S. vom Saal, et al. (1997). "Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol." Environmental Health Perspectives **105**(1): 70-76.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	We examined the low-dose effect of prenatal exposure to bisphenol A and octlphenol on the prostate and other reprudictive organs of male offspring..	ALL DATA ARE IN THE MANUSCRIPT
2) Species, strain, and source of animals	CF-1 Swiss mice, initially purchased from charles river in 1979 and outbred in vom Saal's lab	
3) Diet/source	pregnancy&lactation - Purina 5008 post weaning - purina 5001	
4) Caging protocols (single or multiple housing)	Until 5 - 5.5 months of age, nakes housed in same-sex litter groups and then randomly selected to be individually housed for one month and then killed.	
5) Assignment of treatment groups to cage location on racks	we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	We did not check the purity of bisphenol A and octylphenol used in this experiment Bisphenol A (research grade) – Aldrich Octylphenol (technical grade) – Chem Services	
8) Age and weight of animals at start and end of study	Body weights of the animals in the different groups were recorded at the time of death at 6 months old	
9) Method of assigning animals to dosed and control groups	Random	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
10) Type of control groups? Concurrent with dosed groups?	Vehicle control group - unhandled oil (vehicle) fed dams concurrent with dosed groups for MXC Controls were concurrent with dosed groups	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	Plug positive mice were either left unhandled or they were administered orally (in 30 µl oil) an average dose of 0, 2 or 20 µg/kg bisphenol A or octylphenol from gestation day 11-17 (Plug = day 0). Oil vehicle: ICN Biomedicals, Inc. Aurora, OH. Tocopherol, stripped. 8001-30-7,Lot 95315, cat n 901415 ADMINISTRATION DATES – december - January, 1994 FED WITH A PIPETTER - . The corn oil was administered by electronic micorpipetter (Rainen Instruments) because mice readily consume corn oil that is pipetted into their mouth, and this procedure is not as stressful as gavage (force feeding by stomach tube). CF-1 mice (Mus musculus domesticus) were obtained from Charles River Laboratories (Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond, IN) and, after weaning, females were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{o}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	one male per litter was used to control for litter effects.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no to all questions	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Prostate weight daily sperm production per g testis weight epididymal weight no confounding variables	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	one group of animals all produced and examined together	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician removed the organs. But, at tissue collection, the whole lab participates. The technician removed the organs, but animals were passed to her so she would do the dissection blind to the group of the animal. Another person weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	yes - see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in random order without knowledge of experimental group	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available	NO, natural delivery	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
information.		
21) What statistical techniques were used to evaluate the data and why?	ANCOVA was used when body weight was significantly related to the organs being analyzed, otherwise ANOVA was used. Posthoc analysis was with Lsmeans in SAS	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have historical control data from animals from males. These data will be presented at the meeting.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	There was no selectivity - all animals used are included in the data set to increase power	

Frederick vom Saal

6. Thayer, K. A., R. L. Ruhlen, et al. (2000). "Altered reproductive organs in male mice exposed prenatally to sub-clinical doses of 17α-ethinyl estradiol." **(in press).**

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	EE is used by millions of women who forget to take pills (OCs) and get pregnant, exposing fetuses to this drug. No data exist about effects of clinically relevant doses in animal studies - amazing!. We examined effects of fetal exposure to doses hundreds of times lower to 10-times higher than doses used in OCs on the male reproductive system.	DATA ARE IN THE MANUSCRIPT SUBMITTED FOR PUBLICATION
2) Species, strain, and source of animals	CF-1 Swiss mice, initially purchased from charles river in 1979 and outbred in my lab	
3) Diet/source	pregnancy& lactation - Ralston Purina 5008 post weaning - Purina 5001	
4) Caging protocols (single or multiple housing)	At both weaning (postnatal day 21) and 4 months of age, one male was randomly selected from each litter as the litter is considered the unit of analysis in developmental studies involving maternal dosing. Following weaning litter identity was preserved by continuing to house males from the same litter together. A randomly selected male from a litter was individually housed for one month. For males killed when 50 days old, individual housing began at weaning, and for males killed when 5 month old, individual housing began at four months of age. The reason that males were individually housed is that when male mice are housed in groups, a non-linear	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	hierarchy is often observed where there is one dominant male and the remaining males are subordinate, which can have marked effects on reproductive organs and behavior in CF-1 mice. However, individual housing for one month eliminates the prior effects of subordination in CF-1 mice (unpublished observation).	
5) Assignment of treatment groups to cage location on racks	this is not in the paper, but we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	We did not check the purity of EE used in this experiment Sigma	
8) Age and weight of animals at start and end of study	2 mos and 5 mos old when killed. Body weights of animals in each group in results	
9) Method of assigning animals to dosed and control groups	random	
10) Type of control groups? Concurrent with dosed groups?	unhandled dams oil (vehicle) fed dams concurrent with dosed groups	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing	Plug positive mice were either left unhandled or they were administered orally an average dose of 0, 0.002, 0.02, 0.2, 2, 20 or 200 µg/kg EE_2 dissolved in 30 µl of tocopherol-stripped corn oil (Cat# 901415, ICN, Aurora, OH) once a day during gestational days 0-17. ADMINISTRATION DATES FED WITH A PIPETTER - . The corn oil was administered by electronic micorpipetter (Rainen Instruments) because mice readily consume corn oil that is pipetted into their mouth, and this procedure is not as stressful as gavage (force feeding by stomach tube). CF-1 mice (Mus musculus domesticus) were obtained from Charles River Laboratories	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Light/dark cycle	(Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond, IN) and, after weaning, males were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{o}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	one male per litter was used to control for litter effects	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no	
14) Survival information: were there any early deaths or notable "competing risks"?	One entire group was lost - 20 microgram gp since high estradiol inhibits implantation. No other animals were lost.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	prostate variables and testicular sperm production no confounding variables	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please	one group of animals all produced at tested together	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
give details		
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	one graduate student did this experiment by herself. But, at tissue collection, the whole lab participates. The student (Thayer) removed the organs, but animals were passed to her so she would do the dissection blind to the group of the animal. Another student weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in random order without knowledge of who is who	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	NO, natural delivery	
21) What statistical techniques were used to evaluate the data and why?	We had help here from a renouned statistician. ANOVA and ANCOVA was used with transformations due to heterogeneous variance when required.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have LOTS of historical control data from animals from young adulthood to old age. JOE, I previously sent you these data to discuss the issue of prostate - body weight non-relationship. Do you still have the data set? I will discuss these control data at the conference.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved,	There was no selectivity - all animals used are included in the data set	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
please provide the details		

Frederick vom Saal

9. vom Saal, F. S., P. S. Cooke, et al. (1998). "A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior." Toxicology and Industrial Health **14** (1/2): 239-260.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	We examined the low-dose effect of prenatal exposure to bisphenol A and octlphenol on the prostate and other reprudictive organs of male offspring..	ALL DATA ARE IN THE MANUSCRIPT
2) Species, strain, and source of animals	CF-1 Swiss mice, initially purchased from charles river in 1979 and outbred in vom Saal's lab	
3) Diet/source	pregnancy&lactation - Purina 5008 post weaning - purina 5001	
4) Caging protocols (single or multiple housing)	Until 5 - 5.5 months of age, nakes housed in same-sex litter groups and then randomly selected to be individually housed for one month and then killed.	
5) Assignment of treatment groups to cage location on racks	we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	We did not check the purity of bisphenol A and octylphenol used in this experiment Bisphenol A (research grade) – Aldrich Octylphenol (technical grade) – Chem Services	
8) Age and weight of animals at start and end of study	Body weights of the animals in the different groups were recorded at the time of death at 6 months old	
9) Method of assigning animals to dosed and control groups	Random	
10) Type of control groups?	Vehicle control group - unhandled oil (vehicle) fed dams concurrent with dosed groups for MXC	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Concurrent with dosed groups?	Controls were concurrent with dosed groups	
11) Specifics of treatment regimens: Formulations/vehicle	Plug positive mice were either left unhandled or they were administered orally (in 30 µl oil) an average dose of 0, 2 or 20 µg/kg bisphenol A or octylphenol from gestation day 11-17 (Plug = day 0). Oil vehicle: ICN Biomedicals, Inc. Aurora, OH. Tocopherol, stripped. 8001-30-7,Lot 95315, cat n 901415	
Administration dates	ADMINISTRATION DATES – december - January, 1994	
Route of exposure Dose levels	FED WITH A PIPETTER - . The corn oil was administered by electronic micorpipetter (Rainen Instruments) because mice readily consume corn oil that is pipetted into their mouth, and this procedure is not as stressful as gavage (force feeding by stomach tube).	
Frequency and duration of dosing Light/dark cycle	CF-1 mice (Mus musculus domesticus) were obtained from Charles River Laboratories (Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond, IN) and, after weaning, females were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{o}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	one male per litter was used to control for litter effects.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no to all questions	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Prostate weight daily sperm production per g testis weight epididymal weight no confounding variables	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	one group of animals all produced and examined together	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician removed the organs. But, at tissue collection, the whole lab participates. The technician removed the organs, but animals were passed to her so she would do the dissection blind to the group of the animal. Another person weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	yes - see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in random order without knowledge of experimental group	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available	NO, natural delivery	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
information.		
21) What statistical techniques were used to evaluate the data and why?	ANCOVA was used when body weight was significantly related to the organs being analyzed, otherwise ANOVA was used. Posthoc analysis was with Lsmeans in SAS	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have historical control data from animals from males. These data will be presented at the meeting.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	There was no selectivity - all animals used are included in the data set to increase power	

Frederick vom Saal

10. vom Saal, F.S., K.L. Howdeshell, et al. (2000). High sensitivity of the fetal prostate to endogenous and environmental estrogens. Paper to be presented at the Bisphenol A: low dose effects-high dose effects meeting, Freie Universitat, Berlin, November.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	We examined the relationship between endogenous estradiol due to intrauterine position and prostate size in male offspring in adulthood while exposed to testosterone (T) or dihydrotestosterone (DHT). A similar outcome with T and DHT would rule out that differences are due to an effect of IUP on 5α-reductase.	IUP data sent to Joe Haseman
2) Species, strain, and source of animals	CF-1 Swiss mice, initially purchased from charles river in 1979 and outbred in vom Saal's lab	
3) Diet/source	Pregnancy&lactation - Purina 5008 post weaning - purina 5001	
4) Caging protocols (single or multiple housing)	Until 3 months of age, nakes housed in same-sex litter groups and then randomly selected to be individually housed for one month.	
5) Assignment of treatment groups to cage location on racks	we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses		
8) Age and weight of animals at start and end of study	Body weights of the animals in the different groups were recorded at the time of death at 4 months old	
9) Method of assigning animals to dosed and control groups	Random	
10) Type of control groups? Concurrent with dosed groups?	IUP comparison of 0M and 2M males, so no control group	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	Plug positive mice were either left unhandled until Cesarean delivery just prior to normal parturition on gestation day 19. One week after being isolated at 3 months of age, the animals were castrated and implanted s.c. with a 10 mm-long Silastic capsule (0.62 in ID, 1.25 in OD) containing 0.5 mg testosterone (T) or 0.5 mg 5α-dihydrotestosterone (DHT) dissolved in 2 µl. corn oil. Three weeks later the body weights and prostate weights were recorded. Mice (Mus musculus domesticus) were obtained from Charles River Laboratories (Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond, IN) and, after weaning, females were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{O}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	One male per litter was used to control for litter effects within groups.	
13) Was there any "culling" of	No to all questions	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details		
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Prostate weight no confounding variables	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	one group of animals all produced and examined together	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One technician removed the organs. But, at tissue collection, the whole lab participates. The technician removed the organs, but animals were passed to her so she would do the dissection blind to the group of the animal. Another person weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	yes - see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls	Animals examined in random order without knowledge of experimental group	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	Yes, this was a comparison of 2M and 0M males	
21) What statistical techniques were used to evaluate the data and why?	ANOVA was used. Posthoc analysis was with Lsmeans in SAS	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have historical control data from animals from males. These data will be presented at the meeting.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	There was no selectivity	

Frederick vom Saal

11. Welshons, W. V., S. C. Nagel, et al. (1999). "Low-dose bioactivity of xenoestrogens in animals: fetal exposure to low doses of methoxychlor and other xenoestrogens increases adult prostate size in mice." Toxicology and Industrial Health **15**: 12-25.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	We examined the low-dose effect of prenatal exposure to methoxychlor (MXC) on the prostate and other organs of male offspring..	ALL DATA ARE IN THE MANUSCRIPT
2) Species, strain, and source of animals	CF-1 Swiss mice, initially purchased from charles river in 1979 and outbred in vom Saal's lab	
3) Diet/source	pregnancy&lactation - Purina 5008 post weaning - purina 5001	
4) Caging protocols (single or	Until 8.5 months of age, nakes housed in same-sex litter groups and then randomly selected to	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
multiple housing)	be individually housed for one month and then killed.	
5) Assignment of treatment groups to cage location on racks	we routinely scatter groups across rack from high to low. F.H. Bronson and I published an article in Biol Reprod on the effects of light on LH in CF-1 mice, and light intensity varies from top to bottom of racks	
6) Bedding/source	corn cob	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	We did not check the purity of MXC used in this experiment MXC - technical grade, Kincaid Laboratories	
8) Age and weight of animals at start and end of study	Body weights of the animals in the different groups were recorded at the time of death at 9.5-9.8 months old	
9) Method of assigning animals to dosed and control groups	Random	
10) Type of control groups? Concurrent with dosed groups?	oil (vehicle) fed dams concurrent with dosed groups for MXC Controls were concurrent with dosed groups	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	Plug positive mice were either left unhandled or they were administered orally (in 30 µl oil) an average dose of 0, 20 or 2,000 µg/kg methoxychlor from gestation day 11-17 (Plug = day 0). Oil vehicle: ICN Biomedicals, Inc. Aurora, OH. Tocopherol, stripped. 8001-30-7,Lot 95315, cat n 901415 ADMINISTRATION DATES – March 30 – April 4, 1994 FED WITH A PIPETTER - . The corn oil was administered by electronic micorpipetter (Rainen Instruments) because mice readily consume corn oil that is pipetted into their mouth, and this procedure is not as stressful as gavage (force feeding by stomach tube). CF-1 mice (Mus musculus domesticus) were obtained from Charles River Laboratories (Wilmington, MA) in 1979 and were maintained as an outbred colony in a closed facility since that time. All animals were housed in standard (11.5 x 7.5 x 5 in) polypropylene cages on corn cob bedding. Pregnant and lactating mice were fed Purina breeder chow (#5008) (Richmond,	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	IN) and, after weaning, females were maintained on Purina #5001 standard chow. Rooms were maintained at 23^{o}C on a 12:12 light:dark cycle, with lights on at 1000 h, so that timed-mating occurred at the end of the dark phase and examination for seminal plugs occurred after the onset of the light phase of the light cycle.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	one male per litter was used to control for litter effects.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no to all questions	
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	prostate weight liver weight (very interesting findings with low doses of MXC) no confounding variables	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	one group of animals all produced and examined together	
17) Did the same technician examine and measure dosed	One technician removed the organs. But, at tissue collection, the whole lab participates. The technician removed the organs, but	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
and control animals, or were multiple technicians used? Please give details.	animals were passed to her so she would do the dissection blind to the group of the animal. Another person weighs the organs and records the data. Animal cages are just numbered, so at the time of tissue collection there is no way to identify the individual animals.	
18) Were animals and tissue samples examined in a blinded fashion?	yes - see above	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Animals examined in random order without knowledge of experimental group	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	NO, natural delivery	
21) What statistical techniques were used to evaluate the data and why?	ANCOVA was used since body weight was significantly related to the organs being analyzed (prostate weight was related to body weight at P = 0.06)	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Yes, we have historical control data from animals from males. These data will be presented at the meeting.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	There was no selectivity - all animals used are included in the data set to increase power	

Ibrahim Chahoud

1. Chahoud, I. "Studies on the reproductive effects of in utero exposure to bisphenol A and ethinyl estradiol of male and female Sprague Dawley rat offspring." **(3 Abstracts**).

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	There is concern that current risk assessment strategies for compounds released into the environment are not adequate to evaluate possible hormonal disruption capabilities. It is postulated that endocrine disrupters may have variable dose-dependent effects which are not manifested according to the classic linear dose-response curve commonly seen in toxicological studies. We examined the influence of bisphenol A (BPA) and a reference estrogen, 17-α ethinyl estradiol, on several reproductive endpoints during early development and adulthood at low and high doses to address the questions of (1) whether *in utero* exposure interferes with the reproductive system of the offspring and (2) whether differential effects occur between the low and high doses. Finally, several endpoints were evaluated to examine which endpoints are more sensitive to endocrine disruption. This work was supported by a grant from the Bundesministerium für Umwelt Naturschutz und Reaktorsicherheit # 07HOR01-7 (Federal Ministry for Environmental Protection and Radiation Security).	
2) Species, strain, and source of animals	Rat, Sprague Dawley (Bor: spf, TNO; Fa. Harlan-Winkelmann (Borchen, Germany)	
3) Diet/source	*ad libitum* Type: altromin 1324 FORTII Supplier: Altromin GmbH, Langestr. 42, 32791 Lage, Germany Details: Contents: Raw protein 19%; Raw fat 4.0%; Raw fiber 6.0%; Raw Ash 7.0%; Calcium 0.9%; Phosphorus 0.7% Additives / kg: Vitamin A 25000 IU; Vitamin D3 1000 IU; Vitamin E 125mg; Copper 5mg **Water** Tap water was supplied *ad libitum* via bottles on each cage.	
4) Caging protocols (single or multiple housing)	Gravid females singly housed in Type III macrolon cages with stainless steel covers Male and female offspring separated on approximately day 30 and housed with same gender litter mates.	
5) Assignment of treatment groups to cage location on racks	Not done	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
6) Bedding/source	Type: wood chip Supplier: Altromin GmbH, Germany Details: low dust	
7) Chemical analyses: Chemical(s)/source	Bisphenol A/Sigma-Aldrich Chemicals GmbH (Steinheim, Germany) 4, 4'-isopropylidenediphenol	
Purity of test agent	not given	
Identified contaminants, %	not given	
Stability of test agent	Test solutions prepared shortly before use	
Chemical(s)/source	17-alpha ethinyl estradiol (EE) / Aldrich Chemical Company (Milwaukee, WI, USA) 17alpha-ethinyl-1,3,5-oestratrien-3, 17ß diol....	
Purity of test agent	98%	
Identified contaminants, %	not given	
Stability of test agent	Test solutions prepared shortly before use	
Chemical(s)/source	Mondamin (corn starch) / Maizena Markenartikel (Heilbronn, Germany)	
Purity of test agent	Not given	
Identified contaminants, %	Not given	
Analyses of dose Formulations	Not done	
Methods of analyses	Not done	
8) Age and weight of animals at start and end of study Day 0 = the day sperm was detected in the vaginal smear	**BW (g) on Day 0** / **BW (g) Gestation Day 21** Dams Mondamin: 249 ± 31, N=22 / 382 ± 40, N=21 Dams 50 mg/kg/d BPA: 269 ± 13, N=18 / 398 ± 22, N=18 Dams 0.1 mg/kg/d BPA: 216 ± 15, N=20 / 358 ± 36, N=20 Dams 0.02 mg/kg/d BPA 224 ± 8, N=13 / 392 ± 28, N=13	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Mean weights ± SD represent only those dams who actually were gravid. Age of dams not available *pups are not weighed individually The whole litter is weighed and then the weight is divided by the number of offspring	Dams 0.2 mg/kg/d EE 230 ± 12, N=11 311 ± 25, N=9 Dams 0.02 mg/kg/d EE 231 ± 14, N= 13 337 ± 20, N= 13 **<u>*Mean pup Body Weight g (birth)</u>** Offspring Mondamin: 6.65 ± 0.68, N=20 litters (Mean litter size = 8.4 ± 3.12) Offspring 50 mg/kg/d BPA: 6.65 ± 0.75, N=18 litters (Mean litter size = 9.5 ± 3.97) Offspring 0.1 mg/kg/d BPA: 6.62 ± 0.71, N= 20 litters (Mean litter size = 11.1 ± 3.8) Offspring 0.02 mg/kg/d BPA: 5.96 ± 0.52, N=12 litters (Mean litter size = 12.6 ±3.55) Offspring 0.2 mg/kg/d EE: 5.36 ± 0.5, N=11 litters (Mean litter size = 8.09 ± 4.53) Offspring 0.02 mg/kg/d EE: 5.74 ± 0.53, N=12 litters (Mean litter size = 12.5 ± 2.2) **<u>Necropsy Mean BW (g) PND 70 (males)</u>** Offspring Mondamin: 299.8 ± 29.4 (N=20 male offspring) Offspring 50 mg/kg/d BPA: 310.3 ± 24.9 (N=20) Offspring 0.1 mg/kg/d BPA: 273.4 ± 18.6 (N=20) Offspring 0.02 mg/kg/d BPA: 217.3 ± 40.7 (N=20) Offspring 0.2 mg/kg/d EE: 314.1 ± 28.8 (N=20) Offspring 0.02 mg/kg/d EE: 195.7 ± 9.3 (N=20) **<u>Necropsy Mean BW g PND 170 (males)</u>** Offspring Mondamin: 447.3 ± 27.1 (N=17) Offspring 50 mg/kg/d BPA: 443.6 ± 35.3 (N=20) Offspring 0.1 mg/kg/d BPA: 400.1 ± 35.6 (N=20) Offspring 0.02 mg/kg/d BPA: 351.4 ± 34.7 (N=20) Offspring 0.2 mg/kg/d EE: 420.4 ± 34.5 (N=20) Offspring 0.02 mg/kg/d EE: 349.5 ± 30.3 (N=20) **<u>Necropsy Estrus Females</u>** **<u>Mean Age (d)</u>** **<u>Mean Body Weight (g)</u>** Offspring Mondamin: 124.2 ± 7.6 246.3 ± 13.2 (N=20) Offspring 50 mg/kg/d BPA: 120.0 ± 2.4 242.3 ± 23.4 (N=20) Offspring 0.1 mg/kg/d BPA: 111.3 ± 1.3 226.8 ± 14.5 (N=20) Offspring 0.02 mg/kg/d BPA: 117.1 ± 0.9 217.2 ± 21.2 (N=19) Offspring 0.2 mg/kg/d EE: 120.2 ± 4.1 253.9 ± 23.3 (N=20) Offspring 0.02 mg/kg/d EE: 117.6 ± 0.9 205.6 ± 14.3 (N=20) **Necropsy Diestrus Females**	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	Mean Age (d) / **Mean Body Weight (g)** Offspring Mondamin: 131.3 ± 5.9 / 261.5 ± 11.0 (N=10) Offspring 50 mg/kg/d BPA: 128.2 ± 2.6 / 256.4 ± 17.3 (N=9) Offspring 0.1 mg/kg/d BPA: 108.8 ± 1.2 / 221.4 ± 8.9 (N=8) Offspring 0.02 mg/kg/d BPA: 127 ± 3.4 / 219 ± 12.8 (N=18) Offspring 0.2 mg/kg/d EE: 131.0 ± 3.0 / 273.4 ± 31.8 (N=5) Offspring 0.02 mg/kg/d EE: 126.6 ± 4.1 / 223.6 ± 21.3 (N=19)	
9) Method of assigning animals to dosed and control groups	The treatment groups are performed one after the other and, therefore, no selection of the animals takes place. (see administration dates below)	
10) Type of control groups? Concurrent with dosed groups?	Mondamin Vehicle for bisphenol A Not concurrent	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Formulations/vehicle Administration dates Formulations/vehicle Administration dates Formulations/vehicle	 50 mg/kg/d Bisphenol A / 2% Mondamin (10ml/kg) 150 mg BPA /30 ml 2% Mondamin given at 10 ml/kg 9/21/98 – 10/6/98 9/22/98 – 10/7/98 9/23/98 – 10/8/98 2% Mondamin (10 ml/kg) (w:v) 2g Mondamin / 100 ml tap water 10/12/98-10/27/98 11/2/98-11/17/98 11/3/98-11/18/98 11/4/98-11/19/98 12/14/98-12/29/98 0.1 mg/kg/d Bisphenol A / 2% Mondamin (10 ml/kg) 1 mg BPA / 100 ml 2% Mondamin given at 10ml/kg 12/15/98-12/30/98 12/16/98-12/31/98 12/17/98-1/1/99 0.2 mg/kg/d 17-alpha EE / peanut oil (5 ml/kg) 2 mg EE / 50 ml peanut oil given at 5ml/kg	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Administration dates	2/4/99-2/19/99 2/7/99-2/22/99 2/8/99-2/23/99 2/9/99-2/24/99 2/10/99-2/25/99	
Formulations/vehicle	0.02 mg/kg/d Bisphenol A / 2% Mondamin (1 ml/kg) 1mg BPA / 50 ml 2% Mondamin given at 1 ml/kg	
Administration dates	9/14/99-9/29/99 9/15/99-9/30/99 9/16/99-10/1/99	
Formulations/vehicle	0.02 mg/kg/d 17-alpha EE / peanut oil (1 ml/kg) 1 mg EE / 50 ml peanut oil given at 1 ml/kg	
Administration dates	9/20/99 – 10/5/99 9/21/99 – 10/6/99 9/22/99 – 10/7/99	
Dose levels	50 mg/kg/d, 0.1 mg/kg/d, 0.02 mg/kg/d BPA 0.2 mg/kg/d, 0.02 mg/kg/d EE 2% Mondamin (10ml/kg)	
Route of exposure	Oral per gavage	
Frequency and duration of dosing	*s.i.d*, gestation days 6 through 21	
Light/dark cycle	12 hours light/ 12 hours dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were used. All data is identified by the day of conception and a number ID for the mother and her offspring. For example, in one treatment group, there will be more than one mother with the number 1, but the "day 0" will be different.	
13) Was there any "culling" of litters? If so,	Not done	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Not done	
14) Survival information: were there any early deaths or notable "competing risks"?	Mondamin: no resorptions, no maternal deaths, no dead offspring 50 mg/kg/d BPA: 1 dam with dead offspring, 1 maternal death, no cannibalism of the pups, no resorptions 0.1 mg/kg/d BPA: no resorptions, no maternal deaths, no dead offspring 0.02 mg/kg/d BPA: no resorptions, no maternal deaths, no dead offspring 0.2 mg/kg/d EE: 1 dam with dead offspring, 1 maternal death, 7 dams cannibalism of the pups, 6 resorptions 0.02 mg/kg/d EE: 1 dam with dead offspring, 1 dam cannibalism of the pups, no resorptions	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Body weight at the end of lactation was reduced following exposure to 0.1 BPA, 0.02 BPA and 0.02 EE. Body weight remained low compared to the control for these groups at the time of vaginal opening. The question remains whether the delayed vaginal opening is substance-related or secondary to the decreased body weight resulting from the treatment. However, vaginal opening occurred somewhat earlier in the 50 mg/kg group although the mean body weight at vaginal opening was lower in this group than control.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	There were multiple shipments of animals.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used?	The same technician examined dosed and control animals for all pre-mortem parameters. Post mortem parameters were carried out by different individuals, but always the same individual for the dosed and control animals for each parameter.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Please give details.		
18) Were animals and tissue samples examined in a blinded fashion?	No	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Systematic BPA 50 mg/kg/d (treated September 98-Oct 98) Mondamin (treated (treated Oct 98-Dec 98) 0.1 BPA (treated Dec 98-Jan 99) 0.2 EE (treated Feb 99 – Feb 99) 0.02 BPA (treated Sept 99-Oct 99) 0.02 EE (treated Sept 99- Oct 99)	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	Not available	
21) What statistical techniques were used to evaluate the data and why?	Prostate Weight: ANOVA followed by Dunnett's test Anogenital distance: Mann Whitney Vaginal opening: Mann Whitney test and Fisher's exact test for different types of data presentations Preputial separation: Mann Whitney test and Fisher's for different types of data presentations Daily sperm production: ANOVA followed by Dunnett's test Testosterone concentration: ANOVA followed by Dunnett's test Estrous cycle: Fisher's exact test	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	We usually use Wistar rats for our studies and, therefore, all of our historical data is for Wistar rats. However, the conditions of the grant for this study stipulated that we use Sprague Dawley rats. We have just competed studies on *in utero* exposure to 5mg Daidzein using the same protocol and will be analyzed by mid September.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	<u>Parameter: Estrous Cycle evaluation</u>: Mondamin: 1 female (27-10-98 3-2) was excluded from the estrous cycle analysis due to repetitive metestrus. 0.02 BPA and 0.02 EE groups are not completed for full analysis <u>Parameter: Testosterone levels</u>: The testosterone levels for Mondamin male offspring on PND 70 have not yet been measured. <u>Parameter: Necropsy body weight for females</u>:	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	Estrus 0.02 BPA: 1 female (9-9-99 4-2) not used because uterus did not look like estrus at necropsy Diestrus 0.02 EE: 1 female (16-9-99 1-2) not used because uterus did not look like diestrus at necropsy Diestrus 0.1 BPA: 2 females (11-12-98 1-2 and 4-1) not included because body weight was not recorded at necropsy and 1 female (11-12-98 4-3) not used because uterus looked more like estrus at necropsy. Diestrus 0.02 BPA: 2 females (9-9-99 4-1 and 10-9-99 3-5)not used because uterus did not like it was in the diestrus phase.	

Mokoto Ema

1. Ema, M. (2000). "Two-generation reproduction study of bisphenol A in rats." **(Unpublished Study Report)**.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To examine the effects of bisphenol A (BPA) on the reproductive capability of the parent animals and their offspring in two-generation reproduction study in rats	p. 1, first paragraph
2) Species, strain, and source of animals	SPF Crj:CD (SD) IGS rats obtained from Atsugi Breeding Center of Charles River Corporation, Japan	p. 8, Methods (1)Test animals
3) Diet/source	Gamma-ray irradiated pellets (CRF-1; Oriental Yeast Co., Ltd., Japan)	p. 10, 3)Basic feed
4) Caging protocols (single or multiple housing)	One rat/cage (excluding the acclimatization period, the mating period and the period from the day of parturition to the day of weaning)	p. 10, 2)Materials and Methods
5) Assignment of treatment groups to cage location on racks	Cages of each group were placed lengthwise	
6) Bedding/source	Whiteflake (Charles River Corporation, Japan): from day 17 of pregnacy to the day of weaning	p. 10, 2)Materials and Methods
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations	 Bisphenol A (Lot No. AT 1022; Mitui Chemicalc Co., Ltd., Japan) 99.9% 21 days < (at 0.1 and 100 ug/ml, 24 hours at room temperature, 21 days in cool and dark place) 90.9-105.3%	 p. 7, 1. Test substance p. 7, 1. Test substance p. 8, Stability and concentration tests

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Methods of analyses	HPLC	
8) Age and weight of animals at start and end of study	Start: F0 male: at 5 wk age (167-168 g) End: F2 male at 7 wk age (315-326 g) Start: F0 female at 10 wk age (253-255 g) End: F2 female at 14 wk age (311-323 g)	p. 9, (3) selection of animals to start……. Tables 3 and 4
9) Method of assigning animals to dosed and control groups	Random sampling	p. 9, (3) Selection of animals to start…..
10) Type of control groups? Concurrent with dosed groups?	Yes	p. 7, 2. Control
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Distilled water (for HPLC, Kanto Kagaku Co., Japan) Male: from 5 wk age (F0) to 7 wk age (F2), Female: 10 wk age (F)) to 14 wk age (F2) p.o. using stomach tube 0, 0.2 2, 20, 200 micrograms/2 ml/kg/day Once a day 12hr/12 hr: Light (8:00 AM-8:00 PM), Dark (8:00 PM-8:00 AM)	p. 12, (9) Doses of the test substance p. 9, (5)Breeding of animals 1)Environment
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study		
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide	 Day 4 after birth (parturition = day 0 after birth) 8 littermate (4 males and 4 females, if possible) Random sampling No	p. 14, (4) Number of weanlings

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
details		
14) Survival information: were there any early deaths or notable "competing risks"?	No	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	F1 rats were not administered on day 22 after birth. Administration to F1 rats was started from day 23 after birth. Administration to F2 rats was started from day 22 after birth.	p. 15, C) Premating administration period
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	As the F0 generation rats, 145 each of male and female rats were purchased. Male rats at 4 wk age on 1/22/'99 Female rats at 8 wk age on 3/10/'99	p. 8, last paragraph
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	2-3 technicians examined and measured dosed and control animals.	
18) Were animals and tissue samples examined in a blinded fashion?	No.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Systematic fashion (examine one rat each in order of control, 0.2 ug/kg, 2 ug/kg, 20 ug.kg, and 200 ug/kg, repeated this procedure)	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available	No We used offspring after birth.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
information.		
21) What statistical techniques were used to evaluate the data and why?	Bartlett's equivalence assay, ANOVA, Dunnett's test, Kruskal-Wallis test, Mann-Whitney's U test, Wilcoxon's Historical data rank sam test, Fisher's exact test	p. 31, 5. Statistical methods
22) Any historical control data relevant to the interpretation of experimental results should be provided.	Historical data of Chemical Compound Safety Research Institute	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details		

Earl Gray

1. Gray, L. E., J. Ostby, et al. (1999). "Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat." Toxicology and Industrial Health **15**: 48-64.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	Determine if relatively low doses (3 and 6) of vinclozolin administered to the dam during GD 14 to pnd 3 produced any functional or morphological alterations of androgen-dependent endpoints in the rat.	My office, (OPPTS public docket, BASF, NIEHS and NRDC have some of the data too)
2) Species, strain, and source of animals	Rattus norvegicus, the Norway Rat, LE Hooded, Chalres river (see methods from submitted paper)	Data books, in paper as above
3) Diet/source	See M and Ms	
4) Caging protocols (single or multiple housing)	See M and Ms, varies with life cycle	
5) Assignment of treatment groups to cage location on racks	Randomized complete blocks	
6) Bedding/source	See M and Ms	
7) Chemical analyses:	Provided with purchase	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	See M and Ms See M and Ms Assumed to be stable in oil from work done on a coop with RTI Done once for a single sample See Kelce et al., 1994	 RTI report on vinclozolin Data Book Literature
8) Age and weight of animals at start and end of study	See M and Ms	
9) Method of assigning animals to dosed and control groups	See M and Ms	
10) Type of control groups? Concurrent with dosed groups?	See M and Ms, vehicle-treated Yes	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	See M and Ms	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter mates were identified throughout the study by cage number and an ID unique within the litter.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	no	
14) Survival information: were there any early deaths or notable "competing risks"?	no	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis		
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	See M and Ms	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	One person did AGD on all animals, one did all the areola determination (both with help), one examined nipples, generally, one weighs reproductive tissues, while another does non reproductive tissues and etc.	Some of this is noted on original data sheets
18) Were animals and tissue samples examined in a blinded fashion?	AGD, areolas are done very carefully in a completely blinded fashion. For some other endpoints, like determination of serum testosterone (by RIA) sperm numbers bu machine (coulter couter), or identification of age at puberty this may be unwarranted and/or inefficient, and was not done in a blinded fashsion.	Some in paper and in data books
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls	Randomized complete block design, to the degree possible using scheme set up at the beginning of the study for the dams.	See necropsy sheets in books

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No, not necessary in the rat (Hotchkiss et al., 2000)	None
21) What statistical techniques were used to evaluate the data and why?	See M and Ms	See paper and computer files, data books
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	The only "endocrine" data evaluated were serum testosterone values. These were collected carefully, as described in the M and Ms. All animals examined	

Ping Lee

1. Lee, P. C. (1998). "Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male new born rats." Endocrine **9**(1): 105-111.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To evaluate the consequence of neonatal exposure of male rat pups to nonylphenols in their subsequent reproductive development.	
2) Species, strain, and source of animals	Sprague-Dawley, in-bred colony for the past 11 years in my laboratory.	
3) Diet/source	Purina Rat Chow	
4) Caging protocols (single or multiple housing)	Single housing for each mating pair during pregnancy.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	After giving birth, dam with her own litter were housed together until sacrifice.	
5) Assignment of treatment groups to cage location on racks	Random	
6) Bedding/source	Sani-Chips, P.J. Murphy Forest Products, Montville, NJ 07045	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	Sample of nonylphenols was from Dr. John Lech of the Pharmacology and Toxicology Department of the Medical College of Wisconsin. I was told that the sample of nonylphenols was from American Cyanamide Company. Analysis data unknown	
8) Age and weight of animals at start and end of study	Start at birth, about 5.5-7 gm/animal at 1 day of age. Final weights varied depending on litter and treatment.	
9) Method of assigning animals to dosed and control groups	Random	
10) Type of control groups? Concurrent with dosed groups?	Only littermates were used.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Nonylphenols dispersed in dimethyl sulfoxide (adjusted to 10 ul/10gm body weight) March, May and September 1997 i.p. Injection 0.08, 0.8, & 8.0 mg/Kg daily (A.M.) from 1 to 15 days of age. Daily 12 on/12 off, automatic setting	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Litter-mates were used. Three (3) litters were used for this set of Expts. Sacrificed dates 3/21/97 5/30/97 9/13/97 Control #3-6 #1-2 #7-9	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	0.08mg/Kg NPs --- #10-13 #14-16 0.8mg/Kg NPs --- #17-19 ---- 8.0mg/Kg NPs #20-22 --- --- #refers to the sample numbers in the attached Excel file.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Sometimes At birth Depends on the litter. All male pups were kept. Some female pups were killed to keep the size of the litter as close to 12 as possible (ma.x. 12/litter). None for this set.	
14) Survival information: were there any early deaths or notable "competing risks"?	None for this set.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis		
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	N/A (see #2). Rats were in-bred in house.	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used?	One person.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Please give details.		
18) Were animals and tissue samples examined in a blinded fashion?	No	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Tissues were harvested, placed in tubes and frozen immediately in a box filled with crushed dry ice. Samples were randomly withdrawn for weight determination subsequently.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	N/A; only male pups were used.	
21) What statistical techniques were used to evaluate the data and why?	ANOVA, post-hoc t-test (K-Stat).	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	There are other experiments done at 8.0 mg and heigher /Kg body weight but not included in this set of calculations.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details		

Jimmy Spearow

1. Spearow, J. L., P. Doemeny, et al. (1999). "Genetic variation in susceptibility to endocrine disruption by estrogen in mice." Science **285**: 1259-1261.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To determine the magnitude and nature of genetic variation in sensitivity to endocrine disruption of male reproductive development by estradiol. Treated mice with increasing doses of estradiol in silastic implants followed by measuring testes weight, accessory gland weights, and histological evaluation of spermatogenesis.	Files including: Spearow.testes Spearow.Spermatids
2) Species, strain, and source of animals	Mice (Mus musculus) CD-1 from Charles River; C57Bl/6J from Jackson Laboratory; all bred in Dr. Spearow's laboratory for 1 to 6 generations; C17/Jls and S15/Jls bred and maintained in Dr. Spearow's laboratory.	Listed in data files
3) Diet/source	AgWay /Prolab RMH 2000	
4) Caging protocols (single or multiple housing)	Multiple housing in Stainless Steel cages	
5) Assignment of treatment groups to cage location on racks	No regional pattern used. Mice were put on rack with first implanted cage on upper left and subsequent cages to the right and on lower shelves.	
6) Bedding/source	Pine Shavings	
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses	 Estradiol 17 Beta Sigma Chemical Company	
8) Age and weight of animals at start and end of study	Started at 3 weeks (22-23 days) of age) finished at 6 weeks (43 days) of age. Weights listed in data file "Spearow. Testes"	Listed in File "Spearow.testes"
9) Method of assigning animals to dosed and control groups	Mice were assigned to dose-treatment groups as they became available. An effort was made to assign dose-treatments for each strain in an balanced and overlapping manner, according to the available cage space. Depending on the number of mice available per litter they were assigned to 1 to 3 treatment groups. Note that each strain X dose treatment was from mice from several	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	different litters.	
10) Type of control groups? Concurrent with dosed groups?	Concurrent in an overlapping manner.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	 Estradiol 17Beta was disolved in ethanol, then mixed with Silgard 184 Elastomer and catalyist, and pumped into silastic tubing (0.062 in ID X 0.125 in OD), allowed to polmerize, cut into 2.5 to 10 mm lengths, the ends sealed with medical grade Silastic Type A adhesive. Implants were soaked in PBS one day before implanting. Zero dose received Silgard 184 (no estradiol or ethanol). Subcutaneous Silastic Implants 0, 2.5, 10, 20 or 40 µg estradiol per animal a single implant inserted at 22 - 23 days of age 14 hours light: 10 hours dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Not conducted as litter mate controls.	
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were generally culled down to Max of 10 pups soon after birth. No cross fostering in this study.	
14) Survival information: were there any early deaths or notable "competing risks"?	Some mice died under anesthesia, so we reduced the dose of avertin. Any sick animals were culled from the study.	
15) Specific variables that are considered to be most critical in	1) Strain; 2) Dose of Estradiol; 3) Strain X Dose	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Dependent Variables A) testes weight; B) total Accessory Gland weight; C) Percentage of seminiferous tubules with elongated spermatids. Confounding variables: Potentially Body weight.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	This study was conducted with animals produced by our breeding colony. Animals were dosed and assayed as they became available. This study was also repeated with a single shipment of CD-1 mice from Charles River and will as mice of other strains produced by our laboratory. (These confirmatory data are not shown).	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Multiple technicians. But all administered both control (0) and estradiol doses and measured testes weight and accessory gland weights on control as well as on treated animals. Dr. Spearow trained and supervised technicians in administering treatments and collection of data. Spermatogenesis data was collected by one technician and confirmed by Dr. Spearow. Additional histological analyses on the remainder of critical strain X treatment groups were conducted by Dr. Spearow.	
18) Were animals and tissue samples examined in a blinded fashion?	Dissections were not collected blindly. Histological analyses were conducted "blindly" without knowledge of the strain X treatment group.	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Within a given day of putting in implants, the 0 and low doses were implanted first. For collection, data were collected as animals reached 43 days of age, randomly in regard to dose treatment.	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	1 way ANOVA , 2 way ANOVA and 2 way ANOVA with interaction. Tukey -Kramer Multiple range tests. 4 parameter logistic curve fits to estimate ID 50.	
22) Any historical control data	None	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Animals which lost implant were excluded from analysis since we did know the date of losing their implant. A very small number of sick animals, or animals that were examined on the wrong end date were excluded from the analysis. All other testes and accessory gland weight data are provided. Since this project was conducted with out any direct funding, we could only afford to conduct a histological analysis of spermatogenesis on about 6 animals per strain X treatment group. Nevertheless, we did repeat it with all of the animals with available fixed testes for critical Strain X dose treatment groups, for example the 2.5 µg E2 treatment in B6 strain mice, and found essentially the same results.	

Jimmy Spearow

2. Spearow, J. L., T. Sofos, et al. (2000). Genetic variation in sensitivity to endocrine disruption by estrogenic agents. Paper modified from a poster presented at the Second Annual UC Davis Conference for Environmental Health Scientists, Napa, California, August.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To determine the magnitude and nature of genetic variation in sensitivity to estrogenic agents in a uterotrophic assay.	Files including: Spearow.Uterotrophic
2) Species, strain, and source of animals	Mice (Mus musculus) including CD-1 from Charles River; C57Bl/6J from Jackson Laboratory; CD-9 (a partially inbred strain developed from CD-1), C17/Jls and mus Spretus/RP. All bred and maintained in Dr. Spearow's laboratory.	Listed in data files
3) Diet/source	Prolab RMH 2000 until weaning. Teklad 2016 - Soy and Alfalfa free diet while on test.	
4) Caging protocols (single or multiple housing)	Multiple housing in polypropylene cages with Stainless Steel lids.	
5) Assignment of treatment groups to cage location on racks	Mice housed on same shelf, or in some cases one shelf above. Controls were kept on the left so that there were injected with EB first.	
6) Bedding/source	Pine Shavings	
7) Chemical analyses: Chemical(s)/source Purity of test agent	 Estradiol 3 Benzoate Sigma Chemical Company	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Identified contaminants, % Stability of test agent Analyses of dose Formulations Methods of analyses		
8) Age and weight of animals at start and end of study	CD-1 started at 3 weeks (21-23 days) of age) finished at 6 weeks (43 days) of age. Weights listed in data file "Spearow. Uterotrophic"	Listed in File "Spearow.Uterotrophic"
9) Method of assigning animals to dosed and control groups	Mice were assigned to dose-treatment groups as they became available. An effort was made to assign dose-treatments for each strain in an balanced manner, according to the available cage space. Depending on the number of mice available per litter they were assigned to 1 to 4 treatment groups. Note that each strain X dose treatment was from mice from several different litters.	
10) Type of control groups? Concurrent with dosed groups?	Concurrent in an overlapping manner.	
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of dosing Light/dark cycle	Estradiol 3 Benzoate was dissolved in Organic Canola Oil. Zero dose (Control) received Canola Oil. Canola Oil First 3 days after weaning. Administered Daily. Late June and July, 2000 SC Estradiol -3 -Benzoate 0, 0.1 µg, 1.0 µg, or 10 µg/Kg body weight. Administered Subcutaneously Daily for 3 days: assay on fourth day. 14 hours light: 10 hours dark	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal	Not conducted as litter mate control study.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
in the study		
13) Was there any "culling" of litters? If so, When? How much? What was the method of selection? Was any cross fostering done? If so, please provide details	Litters were culled down to Max of 10 pups soon after birth. No cross fostering of animals in this study.	
14) Survival information: were there any early deaths or notable "competing risks"?	No early deaths, except for one mouse injured during restraint/ injection procedure.	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	1) Strain; 2) Dose of Estradiol 3 Benzoate per Kg Body Weight; 2) Strain X Dose means. Dependent Variable: Uterine Weight Confounding variables: Potentially Body weight and Vaginal Opening.	
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	This study was conducted with animals produced by our breeding colony. Animals were dosed and assayed as they became available	
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Multiple technicians. But all administered both control (0) and estradiol 3 Benzoate doses and measured uterine weight on control as well as on treated animals. Dr. Spearow trained and supervised technicians in administering treatments and collection of data.	
18) Were animals and tissue samples examined in a blinded fashion?	Dissections and measurement of uterine weight were not collected blindly.	
19) Were animals examined in a random order or were they examined in a systematic	Within a given day of injecting Estradiol Benzoate, the 0 dose was injected first, then the 0.1 µg EB/ Kg dose, then the 1 µg/ Kg dose, then the 10 µg / Kg dose.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?		
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No	
21) What statistical techniques were used to evaluate the data and why?	1 way ANOVA , 2 way ANOVA and 2 way ANOVA with interaction. Tukey -Kramer Multiple range tests.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	None	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	All Animals of these strains which received complete treatments according to the protocol were reported. I.e., no selectivity was practiced.	

Rochelle Tyl

1. Tyl, R. W., C. B. Myers, et al. (1999). "Two-generation reproduction study with para-tert-octylphenol in rats." Regulatory Toxicology and Pharmacology **30**: 81-95.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To evaluate the potential of octylphenol (OP) administered in the feed at 0, 0.2, 20, 200, and 2000 ppm to CD® rats to produce alterations in parental fertility, maternal pregnancy and lactation, and growth and development of offspring for two generations, one litter per generation	Final report, protocol in RTI CLS archives and in published paper
2) Species, strain, and source of animals	Rat, outbred albino Crl:CD® (SD)Br, known as CD® rat from Charles River Breeding Laboratories, Raleigh, NC (area R12)	Final report, protocol in RTI CLS archives and in published paper

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
3) Diet/source	Purina certified ground rodent chow (No. 5002), PMI Feeds, Inc., St. Louis, MO	Final report, protocol in RTI CLS archives and in published paper
4) Caging protocols (single or multiple housing)	Singly housed for prebreed, gestation, housed in breeding pairs (one male:one female) during mating, dam housed with litter during lactation, all in polycarbonate solid-bottom cages, bedding, and stainless steel lids	Final report, protocol in RTI CLS archives and in published paper
5) Assignment of treatment groups to cage location on racks	housed by treatment group (identified by Rx code and color code only); then by day of gd 0 by group, then by day of pnd 0 by group	Final report, protocol in RTI CLS archives and in published paper
6) Bedding/source	Ab-Sorb-Dri® cage litter, Laboratory Products, Rochelle Park, NJ	Final report, protocol in RTI CLS archives and in published paper
7) Chemical analyses: Chemical(s)/source	Commercial grade OP (containing primarily para-tert OP) from Union Carbide from South Charleston, WV	In final report (Appendix I) and study records in RTI CLS archives
Purity of test agent	Supplier initial: 91.050% para-tert OP; subsequent analysis: 90.2%	
Identified contaminants, %	Branched OP isomer (0.94%), para-tert OP (90.19%), branched OP isomer (5.96%), branched OP isomer (0.29%), branched OP isomer (0.31%), alkyl (C12) phenol, MW 262 (0.14%), alkyl (C12) phenol, MW 262 (0.31%), dioctylphenol, MW 318 (1.23%)	
Stability of test agent	Pre- and post-study analyses by Sponsor indicated no change in OP purity (at least 10 months)	
Analyses of dose Formulations	Prospectively all doses per formulation date for first four formulation dates and then prospectively for all doses per formulation date for at least every fourth formulation	
Methods of analyses	Standards for acceptable accuracy of mixing were: the mean of the analyzed samples were within ± 15% of nominal, and the % RSD (Relative Standard Deviation) for triplicate samples did not exceed 10%. If one or more of these standards were not met, the dosed diets were not administered to the animals until the problem was resolved by analysis of the archived sample of the dosed diet and/or reformulation and reanalysis. Initially, dosed feed formulations at 0.2 and 2000 ppm, encompassing the range of dosed feed concentrations employed, were evaluated for homogeneity and stability at freezer and ambient temperatures. Formulations were homogeneous and stable for at least 54 days in sealed amber bottles at freezer temperatures (-15 to -20ºC). Formulations were stable for at least nine days under conditions which simulated presentation to animals (at room temperature, in open containers, exposed to light). Dosed feed formulations were made approximately every five weeks and stored frozen. Feed jars were changed weekly. Verification of dosage concentrations was performed prospectively for all	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	doses for the first four formulation dates, and then prospectively for all doses per formulation date for at least every fourth formulation. For analyses of dosing formulations, triplicate 10 g aliquots of each dosage formulation were weighed into scintillation vials, then transferred to separate 125 ml amber glass screw-top bottles, and extracting solvent (20 ml isopropanol/hexane, 5/95) was added to each bottle. The bottles were then sonicated for five minutes and then shaken on a platform shaker for 60 minutes. After the feed had settled for five minutes, approximately 10-13 ml of each sample was transferred to a separate silanized scintillation vial and centrifuged for five minutes. Six ml of each supernatant was transferred to a clean silanized scintillation vial containing 9 ml of hexane. Each extract was applied to a Bond Elut NH2 solid phase extraction column (10 cc, Varian Associates) prerinsed with 2 ml of isopropanol/hexane (1/99). The extracts were allowed to run through the columns by gravity. Each column was then rinsed by 2x2 ml isopropanol/hexane (5/95). A sample (100 µl) of internal standard solution (1.0 mg 2,6-di-tert-butyl-4-methylphenol/ml of isopropanol/ hexane [5/95]) was added to each sample. The samples were mixed, and a 1.5 ml sample was transferred to a silanized amber glass autosampler vial and analyzed by gas chromatography as described below. Six standard samples, spanning the range of formulations assayed, were prepared and analyzed each time samples were analyzed. Samples and standards were analyzed by Hewlett Packard 5890A gas chromatograph with a flame ionization detector (FID), using a J&W DB-225 (15 m X 0.25 mm ID) 0.25 micron film capillary column. (Details of analytical methods and results are presented in Appendix IB.) All dosed feed formulations used in the study had analytical values of 90.0 to 110% of target concentrations. Vehicle control feed formulations contained no octylphenol, with an estimated detection limit of 0.022 ppm.	
8) Age and weight of animals at start and end of study	42 days old (DOB 11/25/96; date of arrival at RTI 1/6/97); F0 males 222.1-272.2g, F0 females 163.5-199.2 g at study start (at necropsy of F0 animals, males 590.04-532.74 g; females 292.05-286.43 g)	In final report and study records in RTI CLS archives
9) Method of assigning animals to dosed and control groups	Assignment by randomization, stratified by body weight (by sex), so body weights of all groups by sex were homogenous and equivalent at study start; range per sex did not exceed 20% of mean weight per sex	In final report and study records in RTI CLS archives
10) Type of control groups? Concurrent with dosed groups?	30/sex, same as treated groups; all same parameters evaluated (except calculation for OP intake) Yes; concurrent with dosed groups	In final report and study records in RTI CLS archives
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure	OP dissolved in acetone, added to premix for each dose level, premix air dried under hood to evaporate acetone, then mixed with blank feed in "sandwich" technique; vehicle is blank feed with acetone as described above January 15, 1997 (F0) - December 19, 1997 (F2 retained male sacrifice) In diet ad libitum 24 hours/day, seven days/week	In final report and study records in RTI CLS archives

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Dose levels	0, 0.2, 20, 200, and 2000 ppm	
Frequency and duration of dosing	Ad libitum in feed jars	
Light/dark cycle	12:12	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	If I understand the question, litter mates were <u>not</u> used for mating inter se; litter mates were selected for retention (F1, F2, F2 postwean exposures), usually 1/sex/litter on pnd 21, occasional "double dipping," all retained animals with unique ear tag (selected animals tracked from lactation records by dam and litter to eartag number)	Study records in RTI CLS archives
13) Was there any "culling" of litters? If so,	Litters were standardized on pnd 4 to 10, with as equal a sex ratio as possible	Protocol and final report, RTI CLS archives
When?	pnd 4	
How much?	10 (as equal sex ratio as possible)	
What was the method of selection?	Random selection	
Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	All F0 males survived to scheduled necropsy; one F0 female at 0 ppm died on sd 109, pnd 13. Rest survived to selected necropsy. All F1 males survived to scheduled necropsy, one F1 female at 0.2 ppm developed a lump on forelimb, euthanized on gd 84; the rest survived to scheduled necropsy. All retained F2 males and females survived to scheduled necropsy. No "competing risks."	Study records in RTI CLS archives
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Specific variables: absolute and relative reproductive organ weights; F0, F1 parents; F1, F2 weanlings; F2 retained males. Histopathology of reproductive organs; F0, F1 parents; F2 retained males. Acquisition of vaginal patency and preputial separation (F1 and F2 offspring). Anogenital distance in F2 pups on pnd 0. F0, F1, F2 male andrology: sperm number, motility, morphology, testicular homogenization resistant, spermatid head counts, daily sperm production (DSP), efficiency of DSP, F0 and F1 estrous cyclicity and estrous stage at necropsy; F0 and F1 ovarian primodial follicle counts -- high dose and control. No "confounding variables"	Data in final report, study records in RTI CLS archives, and in publication
16) Was the study done in a	Single "replicate" with single shipment of animals, arriving at RTI on 1/6/97	Protocol and final report in RTI

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details		CLS archives
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A study team of two technicians (with other staff on weekends, as necessary) did all the in-life observations and measurements. All work was done "blind for dose." Necropsies were done with a group of trained technicians, specific people designated as prosectors and as weighers.	Study records in RTI CLS archives
18) Were animals and tissue samples examined in a blinded fashion?	Yes - always (identified by Rx code, color code, eartag number, and study number)	Study records in RTI CLS archives
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	In-life observations and measurements were performed by group (Rx and color code, "blind for dose"). Necropsy of males was by male study number (lowest to highest), so each "set" of 5 males represented one from each group, within "sets" identity of dose group represented varied. Necropsy of parental females and weanlings on pnd 21 by group (Rx and color code, "blind for dose") on each pnd 21 date	Study records in RTI CLS archives
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No information. F1 and F2 offspring were born (uterine nidation scars were counted at necropsy in F0 and F1 dams)	NA
21) What statistical techniques were used to evaluate the data and why?	The unit of comparison was the male, the female, the pregnant female, or the litter, as appropriate. Quantitative continuous data (*e.g.*, parental and pup body weights, organ weights, feed consumption, food efficiency, etc.) were compared among the four treatment groups and the one vehicle control group by the use of Bartlett's test for homogeneity of variances. If Bartlett's test indicated lack of homogeneity of variances (*i.e.*, $p<0.001$), then nonparametric statistical tests were employed for the continuous variables (see below; Winer, 1962). If Bartlett's test indicated homogeneous variances (*i.e.*, $p>0.001$), then parametric statistical tests were employed for the continuous variables as follows. Appropriate General Linear Models (GLM) procedures (SAS Institute Inc., 1989a,b, 1990a,b,c, 1996, 1997) for the proposed Analyses of Variance (ANOVA) are available at Research Triangle Institute (RTI). Prior to GLM analysis, an arcsine-square root transformation was performed on all litter-derived	Protocol and final report in RTI CLS archives and in publication Text Table D (p.30-32) in final report in RTI CLS archives

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	percentage data (Snedecor and Cochran, 1967) to allow use of parametric methods. For these litter-derived percentage data (*e.g.*, periodic pup survival indices), the ANOVA was weighted according to litter size. GLM analysis was used to determine the significance of the dose-response relationship (Test for Linear Trend) and to determine whether significant dosage effects had occurred for selected measures (ANOVA). When a significant ($p<0.05$) main effect for dosage occurred, Dunnett's Multiple Comparison Test (Dunnett, 1955; 1964) was used to compare each treatment group to the vehicle control group for that measure. A one-tailed test (*i.e.*, Dunnett's Test) was used for all pairwise comparisons to the vehicle control group, except that a two-tailed test was used for parental and pup body weight parameters, feed consumption, food efficiency, and percent males per litter. Nonparametric tests, used for continuous data which did not have homogeneous variances, included the Kruskal-Wallis Test (Siegel, 1956) to determine if significant differences were present among the groups, followed by the Mann-Whitney U test for pairwise comparisons to the vehicle control group, if the Kruskal-Wallis test was significant. Jonckheere's test for k independent samples (Jonckheere, 1954) was used to identify significant dose-response trends for nonparametric continuous data. Frequency data such as reproductive indices (*e.g.*, mating and fertility indices) were not transformed. All indices were analyzed by Chi-Square Test for Independence for differences among treatment groups (Snedecor and Cochran, 1967) and by the Cochran-Armitage Test for Linear Trend on Proportions (Cochran, 1954; Armitage, 1955; Agresti, 1990). When Chi-Square revealed significant ($p<0.05$) differences among groups, then a Fisher's Exact Probability Test, with appropriate adjustments for multiple comparisons, was used for pairwise comparisons between each treatment group and the control group. For developmental landmarks (*e.g.*, vaginal patency and preputial separation), each treatment percent or mean was compared to the control percent or mean by Mann-Whitney U test (Siegel, 1956). In addition, acquisition of reproductive landmarks was analyzed by analysis of covariance (ANCOVA), with body weight as the covariate (the actual body weight on the day of acquisition for selected F1 and retained F2 offspring), and the Least Squares Means Test for pairwise comparisons to the control group value. For all statistical tests, the significance limit of 0.05 (one- or two-tailed) was used as the criterion for significance. Analysis of linear trend and for overall and pairwise comparisons of correlated data (*i.e.*, body weights and absolute and relative organ weight data from weanling necropsies) were performed using SUDAANâ Software (Shah *et al*., 1997). A test for statistical outliers (SAS, 1990b) was performed on male and female body weights. If examination of pertinent study data did not provide a plausible biologically-sound reason for inclusion of the data flagged as "outlier," the data were excluded from summarization and analysis and were designated as outliers. If body weight data for a given observational interval (*e.g.*, sd 0-7, or sd 28-35, or gd 0-7 during the evaluation period) were designated outliers or unrealistic, then summarized data encompassing this period (*e.g.*, sd 0-70 or gd 0-20) also did	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	not include this value.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	They are complete: (1) all high dose and control F0 and F1 parents and retained F2 males subjected to histopathology of selected organs; also gross lesions and reproductive organs from unsuccessful breeders in intermediate dose groups; (2) all males in all groups in F0, F1, and F2 generations had andrology assessment; (3) all F0 and F1 females in all groups evaluated for estrous cyclicity and stage of estrus at demise; and (4) up to three F1 and F2 weanlings/sex/litter were necropsied (selected randomly)	NA

Rochelle Tyl

2. Tyl, R. W., C. B. Myers, et al. (2000). "Three-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats." RTI Study No 65C-07036-000 **(Draft Final Report).**

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To evaluate the potential of bisphenol A (BPA) administered in the feed at 0, 0.015, 0.3, 4.5, 75, 750, and 7500 ppm to CD® rats to produce alterations in parental fertility, maternal pregnancy and lactation, and growth and development of offspring for three generations, one litter per generation	Final report, protocol, and study records in RTI CLS archives
2) Species, strain, and source of animals	Rat, outbred albino Crl:CD® (SD)Br, known as CD® rat from Charles River Breeding Laboratories, Raleigh, NC (area R12 for males, area R04 for females)	Final report, protocol, and study records in RTI CLS archives
3) Diet/source	Purina certified ground rodent chow (No. 5002), PMI Feeds, Inc., St. Louis, MO	Final report, protocol in RTI CLS archives
4) Caging protocols (single or multiple housing)	Housing: Singly housed for prebreed and gestation, housed in breeding pairs (one male:one female) during mating; dam housed with litter during lactation. In stainless steel wire-mesh hanging cages in stainless steel cage racks with automatic watering system (stainless steel piping) for prebreed/postwean exposures; polypropylene solid-bottom cages, bedding, and stainless steel lids (with glass water bottles, with polypropylene screw-cap lids and stainless steel sipper tubes) during mating, gestation, and lactation	Final report, protocol, and study records in RTI CLS archives
5) Assignment of treatment groups to cage location on racks	Housed by treatment group (identified by Rx code and color code only), then by day of gd 0 by group, then by day of pnd 0 by group (cohort 1 and 2, each with 50% of animals/group in different rooms)	Final report, protocol, and study records in RTI CLS archives
6) Bedding/source	Ab-Sorb-Dri® cage litter (Laboratory Products, Rochelle Park, NJ) when in solid-bottom	Final report, protocol, and study

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	polypropylene cages; noncontact absorbent paper under cages when in hanging cages	records in RTI CLS archives
7) Chemical analyses: Chemical(s)/source	BPA (CAS No. 80-05-7) from Fisher Scientific, Pittsburgh, PA, from Acros Organics, Fairlawn, NJ	In final report and study records in RTI CLS archives
Purity of test agent	Supplier: 99.9%; initial RTI analysis: 99.75%; poststudy analysis at RTI: 99.52%	
Identified contaminants, %	At initial analysis, 10 impurities were identified, labeled only as A-D, F-K; impurity B 0.13%, J 0.03%, K 0.04%, all the rest 0.01% or less. At end of study analysis, 11 impurities were identified, labeled only as A-K; impurity B 0.13%, J 0.16%, K 0.07%, F 0.03%, impurities G and H each at 0.02%, remaining impurities (A, C, D, E, and I) at 0.01% each.	
Stability of test agent	Pre- and post-study analyses by RTI indicated no change in BPA purity (at least 12 months; >99.5% purity); see above.	
Analyses of dose Formulations	Prospectively, all doses per formulation date for all formulation dates	
Methods of analyses	Homogeneity and stability of the test material in the animal diets, at the concentrations used in this study, were determined by RTI, using negative ion CI (chemical ionization) GCMS analysis, as follows. The mass spectrometer (MS) was Hewlett Packard No. 5989A, the gas chromatograph (GC) was a Hewlett-Packard No. 5890 Series 2 with Hewlett-Packard Software ChemStation® Version B.07.00 (Unix) in use until September 19, 1999, and a Hewlett Packard HP 6890 Series GC system with a Hewlett Packard No. 5973 Mass Selective Detector, with Hewlett Packard Software Enhanced ChemStation® G1701BA Version B.01.00 (Windows NT), in use from September 19, 1999, until completion of the study. Prior to the performance of this study, aliquots of treated diets, encompassing the range of dietary concentrations to be used in this study (15 ppb and 7500 ppm and 10,000 ppm), were used to assess the homogeneity and stability of the prepared diets. Homogeneity (one sample each, assayed in triplicate, from the left, right, and center of the V-shell blender) were established for the high and low dietary concentrations. Stability analyses were also performed on dosed feed samples at 15 ppb and 10,000 and 7500 ppm at room temperature in open containers (to simulate cageside exposures) for 0, 4, 7, 9, and 14 days for the low dose and at 0 and 9 days for the high doses, and on dosed feed samples stored in sealed containers frozen (approximately -20°C) for 0, 7-9, 14-16, 33-35, and 55-57 days for the low dose; for 0, 9-11, 28-30, and 50-52 days for 7500 ppm; and for 0, 9-11, 29-31, and 48-50 for 10,000 ppm. The cageside stability was at least nine days, and the frozen (approximately -20°C) stability was at least 50 days. The dosed feed for this study was formulated approximately monthly (to allow for formulation, analysis	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	and, if necessary, reformulation and reanalysis), based on the storage stability of the frozen (-20°C) feed. Aliquots from all dosed feed levels for each formulation date were analyzed for BPA concentration. Feed formulations were stored frozen, and feed in the feed jars was changed at least every seven days. Standards for acceptable stability and dose level verification were: the mean of the analyzed samples for each dosed feed level and timepoint was within ±15% of the time 0 mean value, and the % RSD (relative standard deviation) for the triplicate analyses did not exceed 10%. Study animals were administered the control feed or dosed diets *ad libitum*, seven days per week, 24 hours per day, throughout the study.	
8) Age and weight of animals at start and end of study	42 days old (Cohort 1: DOB 8/17/98; date of arrival at RTI 9/28/98; for Cohort 2: DOB 8/24/98; date of arrival at RTI 10/5/98). Body weights at study start (F0): Cohort 1: males 139.8 - 254.4 g females126.6 - 194.8 g Cohort 2: males 206.2 - 264.7 g females154.8 - 194.6 g Body weights at end of study (F3): Cohort 1: males 312.73 - 584.70 g females199.45 - 371.80 g Cohort 2: males 333.29 - 599.80 g females198.16 - 377.98 g (lowest end-study weights from high dose animals)	In final report and study records in RTI CLS archives
9) Method of assigning animals to dosed and control groups	Assignment by randomization within cohorts, stratified by body weight (by sex), so body weights of all groups by sex were homogenous and equivalent at study start; range per sex did not exceed 20% of mean weight per sex	In final report and study records in RTI CLS archives
10) Type of control groups? Concurrent with dosed groups?	30/sex (15/sex/cohort), same as treated groups; all same parameters evaluated (except calculation for BPA intake) Yes; concurrent with dosed groups	In final report and study records in RTI CLS archives
11) Specifics of treatment regimens: Formulations/vehicle Administration dates	 BPA dissolved in acetone, added to premix for each dose level, premix air dried under hood in pans (with "raking" with stainless steel rakers) to evaporate acetone, then mixed with blank feed in "sandwich" technique; vehicle is blank feed with acetone as described above Cohort 1: October 6, 1998 (F0) - January 13, 2000 (F3 retained females) Cohort 2: October 14, 1998 (F0) - January 20, 2000 (F3 retained females)	In protocol, final report, and study records in RTI CLS archives

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Route of exposure	In diet ad libitum 24 hours/day, seven days/week	
Dose levels	0, 0.015, 0.3, 4.5, 75, 750, and 7500 ppm	
Frequency and duration of dosing	Ad libitum in feed jars	
Light/dark cycle	12:12	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	If I understand the question, litter mates were <u>not</u> used for mating inter se; litter mates were selected for retention (F1, F2, F3 postwean exposures), usually 1/sex/litter on pnd 21, occasional "double dipping," all retained animals with unique ear tag (selected animals tracked from lactation records by dam and litter to eartag number)	Study records in RTI CLS archives
13) Was there any "culling" of litters? If so,	Litters were standardized on pnd 4 to 10, with as equal a sex ratio as possible	Protocol, study records, and final report, RTI CLS archives
When?	pnd 4	
How much?	10 (as equal sex ratio as possible)	
What was the method of selection?	Random selection by sex	
Was any cross fostering done? If so, please provide details	No	
14) Survival information: were there any early deaths or notable "competing risks"?	One F0 male at 0.3 ppm and no F0 females died or were sacrificed moribund. All remaining F0 parental animals (out of 30/sex/group) survived for scheduled necropsy. Two F1 males, one each at 0 ppm and 4.5 ppm, and one F1 female at 0 ppm died or was sacrificed moribund (one additional F1 animal, at 7500 ppm, identified in-life as a female, was determined to be an internal male at necropsy, so its in-life data were not included). All remaining F1 parental animals (out of 30/sex/group) survived to scheduled sacrifice. No F2 males and two F2 females (both at 7500 ppm) died or were sacrificed moribund. All remaining F2 parental animals (out of 30/sex/group) survived to scheduled sacrifice. One F3 male at 0.3 ppm and no F3 females died or were sacrificed moribund. All remaining retained adult F3 animals (out of 30/sex/group) survived to scheduled necropsy. There were no "competing risks."	Study records in RTI CLS archives

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Specific variables: absolute and relative reproductive organ weights; F0, F1, F2 parents; F1, F2, F3 weanlings; F3 retained adults. Histopathology of reproductive organs; F0, F1, F2 parents; F3 retained adults. Acquisition of vaginal patency and preputial separation (F1, F2, F3 offspring). Anogenital distance in F2 and F3 pups on pnd 0. F0, F1, F2, F3 male andrology: sperm number, motility, morphology, testicular homogenization resistant, spermatid head counts, daily sperm production (DSP), efficiency of DSP; F0, F1, F2, and F3 estrous cyclicity and estrous stage at necropsy; F0, F1, F2, and F3 ovarian primodial follicle counts -- high dose and control. No "confounding variables"	Data in final report and study records in RTI CLS archives
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	Two replicates: 120/sex/replicate from same supplier, location, and animal room for both replicates Replicate 1: Animals arrived at RTI on 9/28/98, DOB 8/17/98, 42 days old on arrival Replicate 2: Animals arrived at RTI on 10/5/98, DOB 8/24/98, 42 days old on arrival 30/sex/group, 15/sex/cohort; bred within cohorts; for histopathology 10/sex/treatment group (used 5/sex/group/cohort); separate sentinels and QC animals; cohorts 1 and 2 housed in adjacent animal rooms	Protocol, study records, and final report in RTI CLS archives
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	A study team of two to four technicians (with other staff on weekends, as necessary) did all the in-life observations and measurements. All work was done "blind for dose." Necropsies were done with a group of trained technicians, specific people designated as euthanizers, prosectors, and as weighers; also "blind for dose".	Study records in RTI CLS archives
18) Were animals and tissue samples examined in a blinded fashion?	Yes - always (identified by Rx code, color code, eartag number, and study number)	Study records in RTI CLS archives
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	In-life observations and measurements were performed by group (Rx and color code, "blind for dose"). Necropsy of males was by male study number (lowest to highest), so each "set" of 7 males represented one from each group, within "sets", identity of dose group represented varied. Necropsy of parental females and weanlings on pnd 21 by group (Rx and color code, "blind for dose") on each pnd 21 date	Study records in RTI CLS archives
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No information. F1, F2, and F3 offspring were born (uterine nidation scars were counted at necropsy in F0, F1, and F2 dams)	NA

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
21) What statistical techniques were used to evaluate the data and why?	The unit of comparison was the male, female, pregnant female, or the litter, as appropriate. Data from the cohorts were combined for summarization and statistical analyses. Quantitative continuous data (e.g., parental and pup body weights, organ weights, feed consumption, anogenital distance, etc.) were compared among the six treatment groups and the one vehicle control group by the use of Bartlett's test for homogeneity of variances. If Bartlett's test indicated lack of homogeneity of variances (i.e., $p<0.001$), then nonparametric statistical tests were employed for the continuous variables (see below; Winer, 1962). If Bartlett's test indicated homogeneous variances (i.e., $p>0.001$), then parametric statistical tests were employed for the continuous variables as follows. Appropriate General Linear Models (GLM) procedures (SAS Institute Inc., 1989a,b, 1990a,b,c, 1996, 1997) for the proposed Analyses of Variance (ANOVA) are available at Research Triangle Institute (RTI). Prior to GLM analysis, an arcsine-square root transformation was performed on all litter-derived percentage data (Snedecor and Cochran, 1967) to allow use of parametric methods. For these litter-derived percentage data (e.g., periodic pup survival indices), the ANOVA was weighted according to litter size. GLM analysis was used to determine the significance of the dose-response relationship (Test for Linear Trend) and to determine whether significant dosage effects had occurred for selected measures (ANOVA). When a significant ($p<0.05$) main effect for dosage occurred, Dunnett's Multiple Comparison Test (Dunnett, 1955; 1964) was used to compare each treatment group to the vehicle control group for that measure. A one-tailed test (i.e., Dunnett's Test) was used for all pairwise comparisons to the vehicle control group, except that a two-tailed test was used for parental and pup body weight and organ weight parameters, feed consumption, percent males per litter, and anogenital distance, if measured. Nonparametric tests used for continuous data, which did not have homogeneous variances, included the Kruskal-Wallis Test (Siegel, 1956) to determine if significant differences were present among the groups, followed by the Mann-Whitney U test for pairwise comparisons to the vehicle control group, if the Kruskal-Wallis test was significant. Jonckheere's test for k independent samples (Jonckheere, 1954) was used to identify significant dose-response trends for nonparametric continuous data. Frequency data, such as reproductive indices (e.g., mating and fertility indices) was not transformed. All indices were analyzed by Chi-Square Test for Independence for differences among treatment groups (Snedecor and Cochran, 1967) and by the Cochran-Armitage Test for Linear Trend on Proportions (Cochran, 1954; Armitage, 1955; Agresti, 1990). When Chi-Square revealed significant ($p<0.05$) differences among groups, then a Fisher's Exact Probability Test, with appropriate adjustments for multiple comparisons, was used for pairwise comparisons between each treatment group and the control group. For acquisition of developmental landmarks (e.g., vaginal patency and preputial separation) and anogenital distance, each treatment percent or mean was compared to the control percent or mean by Mann-Whitney U test (Siegel, 1956). Acquisition of these	Protocol and final report in RTI CLS archives and in publication Text Table D (p.30-32) in final report in RTI CLS archives (see attached graphic)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	developmental landmarks (e.g., vaginal patency and preputial separation), as well as anogenital distance, was also analyzed by Analysis of Covariance (ANCOVA) using body weight at acquisition or measurement as the covariate. For correlated data (e.g., body and organ weights at necropsy of weanlings, with more than one pup/sex/litter), SUDAAN® software (Shah et al., 1997) was used for analysis of overall significance, presence of trend, and pairwise comparisons to the control group values. For all statistical tests, the significance limit of 0.05 (one- or two-tailed) was used as the criterion for significance. A test for statistical outliers (SAS, 1990b) was performed on parental body weights and feed consumption (in g/day). If examination of pertinent study data did not provide a plausible biologically sound reason for inclusion of the data flagged as "outlier," the data were excluded from summarization and analysis and were designated as outliers. If feed consumption data for a given animal for a given observational interval (e.g., study day [sd] 0-7, 7-14, 14-28, 28-35, etc., during the prebreed exposure period) were designated outliers or unrealistic, then summarized data encompassing this period (e.g., sd 0-70 for the prebreed exposure period) did not include this value.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	They are complete: (1) all control and ten/sex for all treatment groups F0, F1, and F2 parents and retained F3 adults subjected to histopathology of selected organs; also gross lesions and reproductive organs from unsuccessful breeders in intermediate dose groups; (2) all males in all groups in F0, F1, F2, and F3 generations had andrology assessment; (3) all F0, F1, F2, and F3 females in all groups evaluated for estrous cyclicity (and stage of estrus at demise for F0, F1, and F2 parental females; F3 females not bred); (4) up to three F1, F2, and F3 weanlings/sex/litter were necropsied (selected randomly); and (5) ovarian primordial follicle counts in ten F0, F1, F2, and F3 females (five/cohort) for high dose and control.	NA

John Waechter

1. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A." Toxicological Sciences **50**: 36-44.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	The objective of this study was to investigate and further elucidate the potential effects of the test article (Bisphenol A, BPA) on sexual development, as measured by sex organ weights, daily sperm production (DSP), epididymal sperm count and testis histopathology in the male	Report titled "Evaluation of Reproductive Organ Development in CF-1 Mice Following Prenatal

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	offspring of female mice exposed to the test article by deposition in the mouth on days 11 to 17 of gestation.	Exposure to Bisphenol A" submitted by MPI Research, Mattawan, MI to The Society of the Plastics Industry, Inc., Study number 328-046, October 9, 1998; page 10, section 3.1.
2) Species, strain, and source of animals	Species: Mouse Strain: CF-1 Source: Charles River Laboratories, Portage, MI	Page 10, section 3.2 and page 11, section 4.1.1, paragraph 1.
3) Diet/source	Diet: Certified Rodent Chow #5002 Source: PMI Feeds, Inc., St. Louis, Missouri The same lots of diet were provided to animals from all groups at the same time during the course of the study to control across groups for possible variation in the content of the diet.	Page 11, section 4.1.1, paragraph 3. Page 11, section 4.1.1, paragraph 4.
4) Caging protocols (single or multiple housing)	Time mated female mice were individually housed upon receipt. Offspring were co-housed with their respective dams through lactation. F1 males retained for 90 days were singly housed after weaning.	Page 11, section 4.1.1, paragraphs 1 & 3.
5) Assignment of treatment groups to cage location on racks	Animals were housed in consecutive numerical order according to the assigned permanent animal number. Groups of animals were housed consecutively, beginning with male animals. The negative control groups were housed first, followed by the positive control group, the low-dose treatment group, and so on until all groups were housed.	MPI Research Laboratory SOP
6) Bedding/source	Adult females and weanling males were individually housed in polypropylene tubs with stainless steel lids and corn cob bedding. Males retained for 90 days were individually housed following weaning in suspended, stainless steel, wire-mesh type cages. Source of Bedding: Alpha-Dri™ Corncob bedding, Shepherd Specialty Papers, Inc.	Page11, section 4.1.1, paragraph 3.
7) Chemical analyses: Chemical(s)/source Purity of test agent Identified contaminants, %	 4,4'-isopropylidene-2-diphenol (Bisphenol A, BPA, CAS # 80-05-7) from The Dow Chemical Company, Midland, MI and Diethylstilbestrol (DES) from Sigma Chemical Co., St. Louis, MO. Tocopherol-stripped corn oil from ICN Biomedicals Inc., Aurora Ohio 99.7 % for BPA 99 % for DES. Greater than 99 % for corn oil o,p-bisphenol or isomer present at 0.2% by area; other impurities present at less than 0.1% by area.	 Page 12, section 4.2.1; page 109 Page 13, section 4.2.3; page 109. Page 13, section 4.2.2 Page 12, section 4.2.1; page 109 and Page 13, section 4.2.3; page 109. Page 13, section 4.2.2 Page 113.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Stability of test agent	The test agent in the dose solution was found to be stable for at least 7 days.	Page 24, section 5.1.2; Page 116 (Appendix B)
Analyses of dose Formulations	Weekly analyses of dose formulations: The concentrations of BPA and DES in dose solutions were within 10% of the nominal concentrations.	Page 24, section 5.1.3; Page 116 (Appendix B)
Methods of analyses	HPLC	Page 116 (Appendix B)
8) Age and weight of animals at start and end of study	Age of mated females: The females were received time-mated and were specified to be 8-10 weeks of age at receipt. The F1 males were on study from day of birth to 90 days of age.	
	Gestation and Lactation Body weight: No statistically significant differences in body weights or body weight changes were observed between the 2 control groups or between the combined controls and the DES or BPA groups.	Page 25, section 5.2.3; Pages 52-63 (summary tables); Individual data on pages 201-216 (Appendices E &F & G), 225-253 (Appendix I), 399-445 (Appendix L)
	Post-weaning Body Weights: No treatment-related effects were found	Page 27, section 5.2.6.3; Pages 81-83 (summary tables); Individual data on pages 399-445
9) Method of assigning animals to dosed and control groups	Animals gaining greater than or equal to 4.5 grams in body weight during the gestational day 0 to 10 pre-exposure interval were randomized into seven groups on Day 10 of gestation using a stratified, by weight, block randomization procedure until 28 mice/treatment group were assigned to test.	Page 12, section 4.1.2.
10) Type of control groups? Concurrent with dosed groups?	Two negative control groups (vehicle only) each with 28 time mated female mice and a diethylstilbestrol (DES, purity of 99%) group of 28 mice were used. All control groups were studied concurrent with the BPA dosed groups.	Page 12, section 4.1.2.
11) Specifics of treatment regimens:		
Formulations/vehicle	The dosing solutions consisted of BPA or DES dissolved in tocopherol-stripped corn oil (vehicle).	Page 13, sections 4.2.2, 4.2.3, 4.2.4 (with subsections), 4.2.5 (with subsections).
Administration dates	October 31, 1997 to November 22, 1997. (The treatment of mice was stagger-started over a period of 17 days in 13 subgroups for each treatment group to facilitate animal treatment and handling during gestation, delivery and necropsy)	Page 10, section 3.3; page 12, section 4.1.2.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Route of exposure	The vehicle, DES in vehicle and test article (BPA) in vehicle were administered by deposition into the mouth using a micropipetter.	Page 16, section 4.2.9.
Dose levels	All doses were adjusted daily, based on body weight, to provide constant dose levels of 0, 0.2, 2.0, 20 and 200 micrograms/kg/day of BPA or 0.2 microgram/kg/day of DES. The control animals received the vehicle at a dosage volume comparable to that received by the test animals.	Page17, first paragraph.
Frequency and duration of dosing	Once daily on gestation days 11 – 17.	Page 12, section 4.1.2
Light/dark cycle	12 hours of light alternating with 12 hours of darkness. Lighting levels were maintained below 18 ft-candles, as measured from 1 meter off the floor, approximately 1-6 inches in front of the cages on each side of the rack.	Page 11, section 4.1.1, paragraph 3.
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	No, litter mates were not used in this study. A maximum of four F1 male litter mates were retained for 90 days and examined.	Page 18, section 4.3.5.
13) Was there any "culling" of litters? If so, When? How much?	Litters with a total number of pups exceeding 8 were culled on Day 4 postpartum. Culled litters were reduced to a total of 8 pups (8 males when possible). If fewer than 8 males were available, the appropriate number of females was retained to achieve a total of 8 pups. Litters with 8 or fewer pups were not culled. Preferential culling of runts was not performed.	Page 18, section 4.3.5.
What was the method of selection?	Pups to be culled were selected using a computer-generated randomization procedure.	
Was any cross fostering done? If so, please provide details	Cross fostering was not a part of this study design and was not conducted.	
14) Survival information: were there any early deaths or notable "competing risks"?	A total of 8 dams died (2 in the controls, 1 in the DES group, 3 in the 0.2 microgram/kg/day BPA group) and 1 each of the 2.0 and 20 microgram/kg/day BPA groups) and one was euthanized in extremis (0.2 microgram/kg/day BPA group) during the lactation period. The cause of death or moribundity of these animals could not be determined.	Page 25, first paragraph and page 27, section 5.3.1.
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially	Body weight and weight change for dams	Page 25, section 5.2.3; pages 52-63; Pages 201-216 (Appendices F & G)
	Fertility data	Page 25, section 5.2.4; pages 64-65

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
"confounding variables" that should be considered in the data analysis	Litter and pre-weaning pup body weights	Page 26, section 5.2.5; pages 66-73; pages 217-253 (Appendices H & I)
	Post-weaning body weights	Page 27, section 5.2.6.3; page 74; pages 399-445 (Appendix L)
	Organ weights, expression of fluid from selected organs, and necropsy prosection technique	Page 28, section 5.3.2; pages 90-101; pages 447-534 (Appendix M)
	Cauda epididymal sperm concentration and daily sperm production	Page 28, section 5.3.3; pages 102-104; pages 535-579 (Appendix N)
	Microscopic pathology The were no know confounding variables.	Page 29, sections 5.3.4.1 & 5.3.4.2; pages 105-106; pages 580-770 (Appendix O)
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The treatment of mice was stagger-started over a period of 17 days in 13 subgroups for each treatment group to facilitate animal treatment and handling during gestation, delivery and necropsy	Page 12, section 4.1.2.
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Multiple trained technicians were used to perform dose preparation and analysis, in-life clinical observations, organ weights, and pathology. Training was standardized and utilized videotaped demonstrations of proper techniques as well as individual technician performance.	Page 4
18) Were animals and tissue samples examined in a blinded fashion?	Yes, during the complete necropsy at 90 +/- 2 days of age for retained male offspring, dissection of male reproductive organs, and daily sperm production and epididymal sperm counts were conducted with the laboratory technicians blind to the treatment group of each animal.	Page 19, section 4.5.
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Necropsy schedules were prepared placing the animals in a replicate order. On the day of necropsy, animals were randomly placed on an animal rack for delivery to necropsy. All group identification markings were removed from the cage tags on the transfer rack of animals prior to delivery to necropsy.	MPI Research Laboratory SOP

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No, the design and objectives of this study did not allow for the collection of information regarding uterine location of each F1 animal. This study was designed to duplicate the procedures detailed in studies by other investigators as closely as possible, and the collection of information on uterine location would have required surgical intervention during the gestation phase adding a significant new variable to this study (Nagel, S.C., et al, (1997) Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ. Health Perspect. 105: 70-76; vom Saal, F.S., et al. A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production, and behavior. Toxicol. Ind. Health 14: 1/2): 239-260.).	
21) What statistical techniques were used to evaluate the data and why?	Two full groups (each with n, 28) of vehicle-exposed mice (control groups) were used in this study. Prior to statistical evaluation of potential treatment-related effects, statistical analyses were performed on these 2 groups. Because the 2 groups were not statistically different, potential treatment-related effects were evaluated against the combined control values. Descriptive statistics (mean and standard deviations) were reported for food consumption. Body weights, gestation/lactation body weight gains, organ weights, sperm counts, sperm production, and litter size were analyzed using the litter as the experimental unit (Haseman, J.K. and L.L. Kupper (1979). Analysis of dichotomous response data from certain toxicological experiments. Biometrics 35: 281-293). These data were first analyzed by Levine's test (Milliken, G.A. and D.E. Johnson (1992). Analysis of Messy Data. Chapman and Hall, London). If Levene's test was not significant ($p > 0.01$), an analysis of variance (ANOVA) was conducted. If the ANOVA was significant ($p < 0.05$), a Dunnet's test (Dunnet, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Stat. Assoc. 56: 52-64) was performed. If the Levene's test was significant ($p < 0.01$), a rank transformation was performed on these data and an ANOVA was conducted. If the ANOVA was significant ($p < 0.05$), the Wilcoxon Rank-Sum Test (Conover, W.J. (1980). Practical Nonparametric Statistics, 2nd Ed. John Wiley & Sons, New York) with Bonferroni's correction was performed. Fertility indicies were analyzed by the Fisher exact probability test (Agresti, A. (1990). Categorical Data Analysis. John Wiley and Sons, New York) and Bonferroni's correction was used for multiple testing of groups in comparison to a single control. Evaluation of the neonatal sex ratio was performed by the binomial distribution test (Gill, J.L. (1978). Design and Analysis of Experiments in the Animal and Medical Sciences. The Iowa State University Press, Ames, Iowa). Survival indices and other incidence data among neonates were analyzed, using the litter as the experimental unit, by a non-parametric ANOVA. If the ANOVA was significant, the Wilcoxon Rank-Sum Test with Bonferroni's correction was performed. The nominal alpha level used was 0.05.	Page 22, section 4.10.
22) Any historical control data	Charles River Data for Average Litter Size	Pages 813-814 (Appendix R).

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
relevant to the interpretation of experimental results should be provided.		
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Yes, all data that were part of this research effort are included in this report.	

John Waechter

2. Cagen, S. Z., J. M. Waechter, et al. (1999). "Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water." Regulatory Toxicology and Pharmacology **30**: 130-139.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study objectives	To investigate and further elucidate potential effects of BPA on sexual development, as reported by Sharpe and colleagues (*10th Int. Congress. Endocrinology*. S23-4. 1996). Sexual development was measured by determining sex organ weights, daily sperm production, epididymal sperm count, and testes histopathology in the male offspring of female rats exposed to low concentrations of BPA in drinking water during the premating, mating, gestation, and lactation periods.	Report titled "Normal reproductive organ development in Wistar rats exposed bisphenol A in the drinking water." submitted by MPI Research, Mattawan, MI to The Society of the Plastics Industry, Inc., Study number 328-045, December 5, 1998 ; page 9 (summary) and 11(section 3.1.)
2) Species, strain, and source of animals	Species: Rat Strain: Han-Wistar albino Source: Taconic Farms, Inc., Germantown, NY	Page 12, section 4.1.1, paragraph 1
3) Diet/source	Diet: Certified Rodent Chow #5002 Source: PMI Feeds, Inc., St. Louis, MO. The same lots of diet were provided to animals from all groups at the same time during the course of the study to control across groups for possible variation in the content of the diet.	Page 12, Section 4.1.1, paragraph 2
4) Caging protocols (single or multiple housing)	Each female was housed with 1 male from the same treatment group in the male's cage until evidence of mating occurred or 2 weeks had elapsed. After mating, each female was returned to individual housing throughout the study except near parturition and during lactation when they	Page 12, section 4.1.1, paragraph 2

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	were housed with their litters. Adult females and weanling males were individually housed in polypropylene tubs with stainless-steel lids and corncob bedding. Male offspring were individually housed following weaning in suspended, stainless-steel, wire-mesh cages.	
5) Assignment of treatment groups to cage location on racks	**Animals were housed in consecutive numerical order according to the assigned permanent animal number. Groups of animals were housed consecutively, beginning with male animals. The negative control groups were housed first, followed by the positive control group, the low-dose treatment group, and so on until all groups were housed.** Rats were rotated into clean cages every 2 weeks and placed in the same area on the cage rack throughout the study. Cages were rotated during the study, with the exception of the gestation period.	MPI Research Laboratory SOP Page 12, section 4.1.1., paragraph 2
6) Bedding/source	Adult females and weanling males were individually housed in polypropylene tubs with stainless steel lids and corn cob bedding. Males retained for 90 days were individually housed following weaning in suspended, stainless steel, wire-mesh type cages Source: Alpha-dri™ corncob bedding/Shepherd Specialty Papers, Inc., Kalamazoo, MI	Page 12, section 4.1.1, paragraph 2
7) Chemical analyses: Chemical(s)/source	4,4'-isopropylidene-2-diphenol (Bisphenol A, BPA, CAS # 80-05-7) from The Dow Chemical Company, Midland, MI and Diethylstilbestrol (DES) from Sigma Chemical Co., St. Louis, MO. Tocopherol-stripped corn oil from ICN Biomedicals Inc., Aurora Ohio	Page 13, section 4.2.1; paragraph 1 Page 14, section 4.2.3
Purity of test agent	99.7 % for BPA 99 % for DES.	Page 13, section 4.2.1 and page 137 (Appendix A) Page 14, section 4.2.3
Identified contaminants, %	o,p-bisphenol or isomer present at 0.2% by area; other impurities present at less than 0.1% by area.	Page 13, section 4.2.1; page 141 (Appendix A)
Stability of test agent	The test agent in the dose solution was found to be stable for at least 10 days.	Page 15, section 4.2.6, paragraph 2, page 22, section 5.1.2, and pages 151-155 (Appendix B)
Analyses of dose Formulations	The concentrations of BPA and DES in dose solutions were within 10% of the nominal concentrations.	Page 15, section 4.2.6, paragraph 3, page23, section 5.1.3, and pages 148-150 (Appendix B)
Methods of analyses	HPLC	Pages 145-147 (Appendix B)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
8) Age and weight of animals at start and end of study	For the breeding pairs, male rats weighed between 213 and 414 g and females weighed between 141 and 234 g at the start of the study. The males were approximately 10 weeks and females were 8 weeks of age upon arrival. For the females, there was no significant treatment-related effects on premating, gestation, or lactation body weights for any dose level of BPA when compared to controls. For the retained male offspring no effects on body weight were observed at PND 1, the mean body weights ranged from 6.48 to 6.81 grams. At study start for post-weaning males (PND21) the body weights ranged from 49.82 to 52.28 grams. Again there was no significant difference between BPA treated and the control groups At the end of the study the range of body weight means for each treatment group was 322 to 332 grams and the male offspring were 90 days of age.	Page 13, section 4.1.2; pages 80-88 (Tables 11 and 12); and pages 239-262 (Appendices I, J, K) Page 25, section 5.2.5, paragraph 2; pages 83-88 (Table 12); and Individual data on pages 239-262 (Appendices I, J, K) Page 26, section 5.2.8, paragraph 4; pages 103-104 (Table 17); and Individual data on pages 279-307 (Appendix N) Page 26, section 5.2.8, paragraph 4; pages 118-129 (Table 21); and Individual data on pages 479-592 (Appendix Q)
9) Method of assigning animals to dosed and control groups	After pretest body weights were obtained, animals were randomized using a stratified (by weight) block randomization procedure into 7 groups of 28 females each. Any rat whose absolute body weight was outside +/- 20% of the population mean for each sex at the time of randomization was excluded from this study.	Pages 13, section 4.1.2, paragraph 1
10) Type of control groups? Concurrent with dosed groups?	Two negative control groups (drinking water only) were used and were concurrent with the dosed groups. A concurrent DES exposure group was also used and served as a positive control.	Page 13, section 4.1.2
11) Specifics of treatment regimens: Formulations/vehicle Administration dates Route of exposure Dose levels Frequency and duration of	Appropriate amounts of test material were mixed with tap water (Village of Mattawan, MI) to achieve the desired concentrations. Fresh solutions were prepared weekly for each concentration and stored in glass containers. Test material was administered orally via a dissolution in the drinking water. The test article was available in glass water bottles *ad libitum.* Teflon liners were placed over the rubber stoppers used to seal the water bottles to prevent contact of the water with the rubber stopper. The test article was administered to females for 2 weeks prior to mating and during mating (2 weeks), gestation (21-22 days) and lactation periods (22 days). Animals in the treatment group received the test article in concentrations of 0, 0.01, 0.1, 1.0 and 10 ppm BPA in the drinking water. The concentration of the positive control article also remained constant at 0.1 ppm DES	Page 14, section 4.2.4 Page 16, section 4.2.7

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
dosing	in the drinking water. The negative control received only tap water.	
Light/dark cycle	Throughout the study, all rats were kept in an environmentally controlled room. Temperature and relative humidity were monitored and recorded daily and maintained between 67° F (19°C) and 73°F (22°C) and 46% and 78%, respectively. Fluorescent lighting provided illumination 12 hours per day via an automatic timer. Lighting levels were maintained below 18 ft-candles, as measured from 1 meter off the floor, approximately 1-6 inches in front of the cages on each side of the rack. Light intensity was monitored and recorded 6 times throughout the study. Low volume music was played in the animal rooms to provide background noise.	Page 12, section 4.1.1, paragraph 2
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	No, litter mates were not used in this study.	
13) Was there any "culling" of litters? If so, When? How much?	To reduce the variation in the growth of pups, the litters with a total number of pups exceeding 8 were culled on Day 4 postpartum. Culled litters were reduced to a total of 8 pups (8 males when possible). If fewer than 8 males were available, the appropriate number of females were retained to achieve a total of 8 pups. Pups to be culled were selected using a computer-generated randomization procedure. Litters with 8 or fewer pups were not culled. Preferential culling of runts was not performed. Culled pups were examined externally for abnormalities and euthanized by the deposition of a pentobarbital solution into the oral cavity.	Page 17, section 4.3.5
What was the method of selection?	Weaning of all litters were performed 22 days after delivery. A maximum of 4 male weanlings per litter were randomly selected, using a computer-generated procedure, to continue on test to 90 days of age.	
Was any cross fostering done? If so, please provide details	Cross fostering was no a part of this study design and was not conducted.	
14) Survival information: were there any early deaths or notable "competing risks"?	For the breeding pairs, no deaths or clinical alterations were observed in dams from any study group.	Page 23, section 5.2.1
	For retained male offspring, no significant differences in the gestation index, duration of gestation, total pups delivered, live born index, viability index (pups surviving to day 4), lactation index (pups surviving to day 22), or pup sex ratio were observed at any exposure level of BPA.	Page 26, section 5.2.8, paragraphs 1 and 2; pages 97-100 (Table 16); and Individual data on pages 271-271 (Appendix M)
15) Specific variables that are considered to be most critical in	Body weight and weight change for dams	Page 25, section 5.2.5; pages 80-88 (Tables 11 and 12); and p239-262

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis		(Appendices I, J, K))
	Natural delivery data and pre-weaning body weights	Pages 25-26, sections 5.2.6-7 and; pages 95-100 (Tables 15-16); and Individual animal data on pages 271-278 (Appendix M)
	Post-weaning body weights	Page 26, section 5.2.8; pages 111-113 (Table 19); and Individual animal data on pages 421-477 (Appendix P)
	Organ weights, expression of fluid from selected organs, and necropsy prosection technique	Page 27, section 5.3.2; pages 118-129 (Table 21); and Individual animal data on 479-592 (Appendix Q)
	Cauda epididymal sperm concentration and daily sperm production	Pages 27-28, section 5.3.3; pages 130-132 (Table 22) and Individual animal data on pages 593-649 (Appendix R)
	Microscopic pathology	Page 28, section 5.3.5; pages 133-134 (Table 23); and Individual animal data on pages 650-800 (Appendix S)
16) Was the study done in a single "replicate" with a single shipment of animals, or whether multiple replicates or shipments were used. Please give details	The study was done in a single replicate with a single shipment of animals. The treatment of rats by groups was stagger-started over a period of approximately 4 weeks in 5 subgroups for each treatment group to facilitate animal treatment and handling during mating, gestation, delivery, and necropsy.	Page 12, section 4.1.1, paragraph 1
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Multiple trained technicians were used to perform dose preparation and analysis, in-life clinical observations, organ weights, and pathology. Training was standardized and utilized videotaped demonstrations of proper techniques as well as individual technician performance.	Key personnel are listed on pages 833-835 (Appendix W)

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
18) Were animals and tissue samples examined in a blinded fashion?	Yes, during the complete necropsy at 90 +/- 2 days of age for retained male offspring, dissection of male reproductive organs, and daily sperm production and epididymal sperm counts were conducted with the laboratory technicians blind to the treatment group of each animal.	Page 18, section 4.5
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	Necropsy schedules were prepared placing the animals in a replicate order. On the day of necropsy, animals were randomly placed on an animal rack for delivery to necropsy. All group identification markings were removed from the cage tags on the transfer rack of animals prior to delivery to necropsy.	MPI Research Laboratory SOP
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	No, the design and objectives of this study did not allow for the collection of information regarding uterine location of each F1 animal. This study was designed to duplicate the procedures detailed in studies by other investigators as closely as possible, and the collection of information on uterine location would have required surgical intervention during the gestation phase adding a significant new variable to this study (Sharpe et al, (1995) Gestational and lactational exposure of rats to xenoestrogens results in reduced testicular size and sperm production. Environ. Health Perspect. 103, 1136-1143; Sharpe et al. (1996) Effects on testicular development and function. 10th International Congress of Endocrinology, June 12-15, San Francisco. S23-4 (Abstract); Nagel, S.C., et al, (1997) Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ. Health Perspect. 105: 70-76; and vom Saal, F.S., et al. A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production, and behavior. Toxicol. Ind. Health 14: 1/2): 239-260.).	
21) What statistical techniques were used to evaluate the data and why?	Two full groups (each with n=28) of tap water-only exposed rats (negative control groups) were used in this study. Prior to statistical evaluation of potential treatment-related effects, statistical analyses were performed on these two groups. Because the two groups were not statistically different, potential treatment-related effects were evaluated against the combined control values (n=56). Descriptive statistics (means and standard deviations) were reported for food and water consumption. Body weights, gestation/lactation body weight gains, organ weights, sperm counts and production, and litter size were first analyzed by a Levene's test. If Levene's test was not significant ($p \geq 0.01$), an analysis of variance (ANOVA) was run. If the ANOVA was significant, a Dunnett's test was performed. If the Levene's test was significant ($p < 0.01$), a rank transformation was performed on these data and an ANOVA was run. If the ANOVA was significant ($p \leq 0.05$), the Wilcoxon Rank-Sum Test with Bonferroni's Correction was	Pages 20-21, section 4.10

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
	performed. Fertility indices were analyzed by the Fisher exact probability test and Bonferroni's correction was used for multiple testing of groups in comparison to a single control. Evaluation of the neonatal sex ratio was performed by the binomial distribution test. Survival indices and other incidence data among neonates were analyzed using the litter as the experimental unit by a non-parametric ANOVA. If the ANOVA was significant, the Wilcoxon Rank-Sum Test with Bonferroni's Correction was performed. The nominal alpha level used was 0.05. Because numerous measurements were statistically compared in the same group of animals, the overall false positive rate (Type I errors) was much greater than the cited alpha levels would suggest. Thus, the final interpretation of numerical data would consider statistical analyses along with other factors such as dose-response relationships and whether the results were significant in the light of other biologic and pathologic findings.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	No historical control data are available for this strain of rat at this laboratory.	
23) Are the endocrine response data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	Yes, all data that were part of this research effort are included in this report.	

Frank Welsch

1. Elswick, B. A., F. Welsch, et al. (2000). "Effect of different sampling designs on outcome of endocrine disruptor studies." Reproductive Toxicology **(in press).**
2. Elswick, B. A., D. B. Janszen, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." Toxicological Sciences **54**(Supplement): 256A.
3. Welsch, F., B. A. Elswick, et al. (2000). "Effects of perinatal exposure to low doses of bisphenol A on female offspring of Sprague-Dawley rats." Toxicological Sciences **54**(Supplement): 256A.

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
1) Specific overall study	The purpose of this study was to look for permanent effects from in utero and perinatal	All data are stored in notebooks and

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
objectives	exposure to BPA in offspring whose mothers had consumed drinking water containing BPA at five concentration levels ranging from 0.005 up to 50 mg BPA/l from GD 2 to PND 21.	CD roms that are archived in QA facilities at CIIT.
2) Species, strain, and source of animals	Rats; Crl:CD® BR VAF/plus (Sprague-Dawley); Charles River Laboratories, Raleigh, NC.	
3) Diet/source	NIH-07, Zeigler Brothers, Gardners, PA	
4) Caging protocols (single or multiple housing)	Dams were housed singly in plastic cages with their litters. At weaning, several male sex offspring were housed per cage.	
5) Assignment of treatment groups to cage location on racks	After randomization, cages were arranged on racks with animals in numerical order according to treatments.	
6) Bedding/source	Alpha Dri®, Sheperd Specialty Papers, Kalamazoo, MI.	
7) Chemical analyses:		
Chemical(s)/source	Sigma, St. Louis, MO	
Purity of test agent	99+%	
Identified contaminants, %	not available	
Stability of test agent	very stable. We found through repeated analyses that BPA in drinking water solutions that BPA at room temperature was stable for at least 6 weeks.	
Analyses of dose Formulations	Yes, dosing solutions were analyzed.	
Methods of analyses	HPLC with UV detection	
8) Age and weight of animals at start and end of study	The pregnant dams destined for BPA exposures arrived on GD 0 from Charles River. The male offspring were retained for up to 6 months of age.	
9) Method of assigning animals to dosed and control groups	Randomized so that weight distributions were similar across all groups.	
10) Type of control groups?	Pregnant dams consuming deionized water, the vehicle for BPA.	
Concurrent with dosed groups?	yes	
11) Specifics of treatment regimens:		
Formulations/vehicle		
Administration dates	GD 2 to PND 21.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
Route of exposure	Oral via drinking water	
Dose levels	The water contained either 0, 0.005, 0.05, 0.5, 5, or 50 mg BPA/L.	
Frequency and duration of dosing	The water bottles were changed every 3 days and the volume consumed was determined.	
Light/dark cycle	12 hours light/dark cycle—7 am to 7 pm.	
12) Whether litter mates were used, if so, specify the precise litter identities of each animal in the study	Littermates were used. The dam provided the liter identity, i.e. litter identity stays with a given pup.	
13) Was there any "culling" of litters? If so, When?	There was no culling, but at various times (PND 21, 41, 100 and 177) littermates that had been randomly assigned to numerous different experimental end points were removed for various studies.	
How much?	1 or 2 males were removed per time point.	
What was the method of selection?	Males within a litter were randomly tatooed on PND 2. Allocation of the pups to particular end points had been predetermined in the protocol, e.g. pup 6 would be necropsied on PND 21 etc.	
Was any cross fostering done? If so, please provide details	No cross fostering was done.	
14) Survival information: were there any early deaths or notable "competing risks"?	no	
15) Specific variables that are considered to be most critical in the overall evaluation of the study; any potentially "confounding variables" that should be considered in the data analysis	Only one or two males per litter were selected per end point for organ weights. We found a lot of interindividual variability in ventral prostate weights. From another study where all males in the litter were retained, we learned there was a lot of interindividual variability within a given litter in this end point. This variability could lead to misinterpretation of the results if only one or two offspring are examined. Additionally, in our BPA drinking water study, more than one prosector dissected the prostates. Individual dissection technique differences may have contributed to some to the variability.	
16) Was the study done in a single "replicate" with a single	This study was done in 2 replicates. Two groups of dams with 56 animals each were received approximately 4 months apart.	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
shipment of animals, or whether multiple replicates or shipments were used. Please give details		
17) Did the same technician examine and measure dosed and control animals, or were multiple technicians used? Please give details.	Multiple technicians were used because there was more work than could be done by one person. The same team worked on both replicates.	
18) Were animals and tissue samples examined in a blinded fashion?	Yes	
19) Were animals examined in a random order or were they examined in a systematic fashion (e.g. all controls measured first)? If non-random, what was the order of examination used?	random	
20) Is there information on the uterine location of each animal (e.g., between 2 males, between 2 females, between one of each sex)? Please provide available information.	no	
21) What statistical techniques were used to evaluate the data and why?	JMP software was used. A nested ANOVA where dose was considered a fixed effect and litter within dose a random effect was used for organ weight analysis since there was not a significant relationship between body weight and organ weights. A nested ANOVA where dose was considered a fixed effect and litter within dose a random effect with body weight as covariate was used for anogenital distance analysis for each sex. Litter averages were taken and an ANOVA performed on PPS and VO data. A *t*-test was used for DES and control comparisons.	
22) Any historical control data relevant to the interpretation of experimental results should be provided.	We have included the data from 19 control litters of another study conducted at CIIT where prostate weights from all males in the litter were available. These animals were 120 ± 10 days old at the time of prostate dissection. We also have control weights from another BPA study in which all males in the litter were retained until 6 months old.	
23) Are the endocrine response	One control animal VP weight of 0.07 g was excluded. We believe that this weight was the	

ISSUE	INVESTIGATOR'S RESPONSE	LOCATION OF THESE DATA
data provided to us complete in the sense of including all animals that were evaluated as part of this research effort? If some selectivity was involved, please provide the details	result of either a weighing error or recording error, but we were unable to confirm our suspicion.	

Appendix C:
Public Comments
Index

From: Tony Tweedale
Sent: Wednesday, May 23, 2001 3:57 PM
To: liaison@starbase.niehs.nih.gov
Subject: ED Low-Dose Peer Review rpt.-- public comment

NTP's & NIEHS' Endocrine Disruption (ED) Low-Dose Peer Review Rpt.-- public comment

Office of Liaison & Scientific Review, Ntl. Toxicology Program
RTP, NC (by email)

Dear NTP & NIEHS:

Though I have not read the report's detailed appendices, and I am not familiar with the detailed discussions of your peer review of this work, it seems evident you had more dispute in Chapter 1 (estrogenic low-dose effects)than other parts of the review. You admit having insufficient time to carefully review all the available estrogenic low-dose effects data, and perhaps as a result significantly distort the current state of knowledge.

For example, you do not discuss the relevant data of Dr. Gupta (you only mention its existence, in response to question 4 in this chapter). Yet Dr. vom Saal, in a speech last year at a workshop (19-21 Apr. '01) at the Health Effects of Env. Contaminants of the U. of Iowa, says that both DES and BPA stimulate androgen receptor increase in fetal prostate at 0.1 ppt and 50 ppt, respectively (apparently, this is the delivered dose to the fetus, but in either case, an very low dose indeed). While not a frank adverse effect, it is obviously a mechanistically *highly relevant* effect for the work that you did spend so much time on, and it obviously merits your review, discussion and conclusion. Please do so.

Further evidence of your unbalanced and/or incomplete criticism of the existing low-dose estrogenic effect data (i.e., Ch. 1) surfaces in your response to Ch. 1's Q. #7 ("..what is the subpanel's overall conclusion..?"). While you are applying an appropriately rigorous standard of proof, it is also important that your discussion and conclusion be complete. You conclude that you are not persuaded that low-dose effects have been conclusively established for estrogenic chemicals. Yet you summarize precious little detailed critique (other than in Q. 3, factors

that may account for different results) of the studies that "consistently" do not replicate positive results, and (possibly) mention only one, by an investigator with an economic conflict of interest (Ashby, of Zeneca's labs).

Given your conclusion of insufficient evidence to make a positive conclusion, it is fair that you mention the lack of mechanistic knowledge, and the "unclear biological relevance" of the work you reviewed. Yet, as with other parts of Ch. 1, I found these statements unbalanced. They badly need statements of the countervailing general truth: that biologic organisms are highly balanced systems dependent on signal transduction, whose balance has frequently been shown to be significantly changed by very low input chemical (endogenous or exogenous) and other signals.

Specifically regarding the alleged lack of mechanistic data, again I rely on the same speech of Dr. vom Saal. He concludes with four data points on the mechanistic steps of adverse effects, each of which lowers the previously accepted NOAEL for BPA. Together, he indicates that the steps (and there could be more that act similarly) bring the bioactivity of BPA down to some 20,000 times less than the previous NOAEL (which is based on an utterly useless (dangerous)high-dose testing paradigm). These four factors are bioaccumulation, non-binding to inactivating blood plasma proteins, the beta ER's affinity for BPA, and the induced increase in epidermal growth factor (EGF).

Next, I protest the failure to even note the economic conflict of interest of a good percentage of your reviewers. First, it is not a "potential" conflict: their involvement alone creates the conflict (so many bodies make this mistake that I begin to wonder if it is sometimes deliberate). More important, fairness and the best science demands you add explicit statement/s that this economic conflict inevitably has some influence on their work. This is especially critical in this report, dealing with such a critical area of health. You need not (nor did you) *discount* their work in any way--it is simply a matter of complete, not superficial, disclosure. By no means will all readers (including the media) be sufficiently aware of the natural, if unquantifiable, influence their employment has on their work. In an ideal world, the agents of a potential problem would not be involved in the judgment of the problem. Say so.

In sum, I find your standard for conclusions appropriate, but find your discussion and conclusions somewhat but critically biased in favor of the dangerous concept that biological systems are resilient to insult.

Certainly not all changes in signal transduction are detrimental to an organism. But just because we are not dropping dead or contracting disease after three centuries of the industrial age does not mean we are not severely being affected (by the same). As you know, the world is enormously complex. So while assigning causation is difficult, it is equally important to guard against the 'no evidence = no problem' fallacy.

The solution, especially in Ch. 1, is not to change what you write, but to add to it. Otherwise, 'lack of knowledge' is inevitably conflated with 'no problem', if only subconsciously. Given the potential vulnerability of the signaling systems that organisms depend on for health; and given the significant potential for harm of xenobiotics to health and ecosystems (what it is *all* about), we all deserve this change in emphasis in this important report, before you transmit it to the EPA.

I appreciate, both the enormous amount of work you did in reviewing the low dose ED data set; and the general openness (disclosure at all levels) of the report.

Tony Tweedale
224 E. Pine (2)
Missoula, MT 59802
406-542-1709 ttweed@wildrockies.org

Causality is a concept not subject to empirical demonstration. -David Hume

'Letting the Grandchildren Do It: Environmental Planning in the Age of Oil' [ie, Petrochemicals] -Joseph A. Pratt

Temperate but endangered planet. Enjoys weather, northern lights, continental drift. Seeks caring relationship with intelligent life form. -Friends of the Earth

PEOPLE FOR THE ETHICAL TREATMENT OF ANIMALS
501 FRONT STREET
NORFOLK, VA 23510
TEL 757-622-PETA
FAX 757-622-0457
www.peta-online.org
info@peta-online.org

July 6, 2001

NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive
P.O. Box 12233, MD A3-01
Research Triangle Park, NC 27709

VIA REGISTERED MAIL AND FACSIMILE

Re: **Public Comments on the NTP Final Report from the Endocrine Disruptors Low-Dose Peer Review**

These comments are submitted on behalf of People for the Ethical Treatment of Animals (PETA) and our 700,000 members in response to a ***Federal Register*** notice of May 8, 2001 soliciting public comment on the final report of the National Toxicology Program (NTP) Endocrine Disruptors Low-Dose Peer Review.

The findings of the NTP peer review amply demonstrate the many shortcomings of using animals to screen or test for endocrine effects, including obvious and unpredictable differences between species, strains and individuals. These problems, coupled with the glaring lack of reproducibility in study results and the numerous inadequacies in the experimental design of existing animal testing protocols (including uncontrolled variables such as phytoestrogen levels in the diet, dosing methods and regimens, housing conditions, differing laboratory techniques, etc.), further confound the situation. These findings clearly underscore the huge problems associated with attempting to extrapolate animal-derived data to human situations, in which our sensitivity to hormonal disturbance and our endogenous levels of hormones will vary widely from those of non-human animals.

We therefore question the panel's recommendations that additional animal-based experiments be performed to further study the low-dose hypothesis. A more appropriate course of action would be to undertake detailed mechanistic studies *in vitro* (involving ligand binding to different receptor subtypes) to investigate kinetics and affinities of binding for both receptors and chemicals at a wide range of doses.

We also would like to remind the panel that there is currently no scientific consensus regarding what an "endocrine disruptor" is or what constitutes an "adverse effect." Moreover, no screening or testing method has undergone rigorous scientific validation according to internationally accepted criteria to demonstrate its reliability and relevance to humans. Until such basic scientific issues are resolved, it is premature and highly inappropriate for any regulatory agency to begin to assess chemicals for endocrine effects <u>at any dose level</u>.

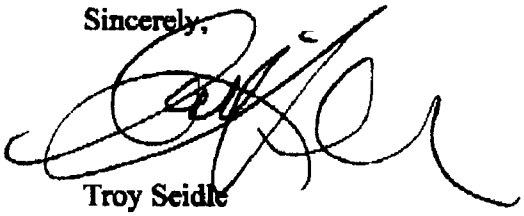
It is unconscionable that while the NTP report recommends "revisiting" the U.S. Environmental Protection Agency's (EPA) existing test guidelines for

AN INTERNATIONAL ORGANIZATION DEDICATED TO PROTECTING THE RIGHTS OF ALL ANIMALS

reproductive and developmental toxicity, these non-validated tests are currently being required by the EPA. We fear that the sloppy approach to test method validation (e.g., lack of standardized protocol, double standard for animal versus non-animal methods, etc.) will continue with the proposed animal screens and tests for endocrine disruption.

Thank you for your attention to these comments.

Sincerely,



Troy Seidle
Research Associate

From: Fred Vom saal
Sent: Wednesday, July 11, 2001 10:26 PM
To: liaison@starbase.niehs.nih.gov
Subject: comments on ED Low Dose Report

vomsaal_low_dose.doc

I am submitting comments concerning the Endocrine Disruptor Low Dose Peer Review report. The attached is a word file.

Regards,
Frederick vom Saal

Frederick S. vom Saal
Division of Biological Sciences
114 Lefevre Hall
University of Missouri
Columbia MO 65211
Tel: 573-882-4367
Fax: 573-884-5020
vomsaalf@missouri.edu

should be submitted to the NTP Office of Liaison and Scientific Review; 919-541-0530 (phone); 919-541-0295 (fax); liaison@starbase.niehs.nih.gov ; 111 T.W. Alexander Drive, P.O. Box 12233, MD A3-01, Research Triangle Park, NC 27709.

Frederick S. vom Saal
Division of Biological Sciences
114 Lefevre Hall
University of Missouri
Columbia MO 65211
Tel: 573-882-4367
Fax: 573-884-5020
vomsaalf@missouri.edu

COMMENTS FOR LOW DOSE PEER REVIEW RELATED TO BISPHENOL A SUBPANEL

p 1-8. Question 3, Item 4. "The control body weights and prostate weights differ between some studies, e.g. some of the Ashby studies and the vom Saal studies. This raises the theoretical possibility that tissues may have already been maximally stimulated by estrogens [in the Ashby studies]..."

We have examined a possible basis for differences in the outcome of fetal exposure to bisphenol A in previously conducted studies (vom Saal, 2001). These studies involved administration of bisphenol A to pregnant mice and examination of male offspring during postnatal life. Of interest is that in two studies that found no effect of either bisphenol A or the positive control estrogenic chemical diethylstilbestrol (DES) on the prostate (Ashby et al., 1999; Cagen et al., 1999), the control animals were markedly heavier than were the animals in the studies in which there were effects of bisphenol A, as well as the positive control chemicals DES and ethinyl estradiol (Nagel et al., 1997; vom Saal et al., 1997; Thayer et al., 2001). In addition, in the Ashby and Cagen studies, the prostate in control animals was enlarged relative to prostate weight reported in the Thayer and Nagel studies, which involved the use of the same strain of mouse at the same ages as Ashby and Cagen. One potential basis for the differences in body weight and prostate weight in the control males in these studies was that in the Ashby and Cagen studies, a different animal feed was used relative to the feed used in the

studies by Nagel and Thayer. In the study described below, we compared a diet formulated based on the NIH-31M diet, which is a casein-based diet, with a diet based on soy protein that contained high levels of phytoestrogens (the Purina 5001 diet).

Adult male and female CD-1 mice were randomly assigned to be fed these two different diets for two weeks prior to being paired, and the two colonies were then maintained on each diet throughout the remainder of the experiment; all of these adult animals had been reared on the Purina 5001 diet. The F1 offspring were examined for differences due to the diet that their parents and that they were maintained on. In comparison to the F1 females on Purina 5001, the NIH-31M females showed early puberty and a greater uterine weight gain in the prepubertal uterotrophic assay (in response to estradiol). Both of these outcomes are consistent with NIH-31M females having been exposed during development to higher levels of estrogen relative to Purina 5001 females.

Both males and females on the NIH-31M diet had dramatically higher (2-fold) amounts of body fat and 2-3 fold higher serum leptin levels relative to Purina 5001 animals. The NIH-31M animals also had an impaired response to a glucose challenge, with glucose levels remaining elevated for a significantly longer period of time. Interestingly, males on the NIH-31M diet were heavier than males on the Purina 5001 diet, but there was no significant weight difference in females; this suggests a difference due to diet in body composition, since the NIH 31M animals had significantly more body fat. In males maintained on the NIH-31M diet, there was a significant decrease in testis weight, seminal vesicle weight and epididymitis weight, but a significant increase in prostate weight, relative to males on the Purina 5001 diet. The NIH-31M animals thus appeared to have been estrogenized, since this is exactly the outcome reported in male CF-1 mice maintained on Purina 5001 diet and exposed to bisphenol A during prenatal life [vom Saal, 1998 #175; Gupta, 2000a #389]. The finding that the phytoestrogen-free NIH-31M diet appeared to estrogenize males was unexpected, since the NIH-31M diet has lower estrogenic activity than does the Purina 5001 diet (due to high levels of the phytoestrogens genistein and daidzein in the Purina 5001 feed). The likely basis for these findings was provided in a study in which we measured endogenous estradiol levels in pregnant females and fetuses maintained on the two different diets. We found that during fetal life, males whose mothers were consuming the NIH-31M diet had significantly higher serum estradiol than did the males whose mothers were consuming the

Purina 5001 diet. This finding thus could explain the similarity of effects of the NIH-31M diet and maternal administration of estrogenic chemicals (such as bisphenol A) on the reproductive system in males from animal colonies being maintained on the Purina 5001 feed. Specifically, the casein-based NIH-31M diet resulted in animals that were already showing substantial estrogen-related effects due to elevated estradiol levels during fetal life.

In our feed study, the NIH-31M animals matched the effect on the prostate in control animals of the different diets used by Ashby and Cagen. Our findings suggest that the animals in the Ashby and Cagen studies may have experienced an elevated background level of serum estradiol relative to the level experienced by the animals in the Thayer and Nagel studies. As a result, bisphenol A did not have a significant effect on the prostate in the Ashby and Cagen studies, since the prostate in the treated males was already near the maximum amount of enlargement that can be achieved due to supplemental estradiol or estrogenic chemical.

Taken together, these findings show that components of feed can alter endogenous hormone levels and change the course of development of the reproductive system, and also influence many other aspects of phenotype, including rate of sexual development, amount of body fat, and body weight. Clearly, the feed is an important variable to consider in endocrine disruptor studies. Both the Ashby and Cagen studies were presented as having replicated exactly procedures used in the Nagel study, yet in the Ashby and Cagen studies different types of feed were used. No mention was made in the publication by either Ashby or Cagen that a different feed was used and that the control animals were markedly different from the control animals in the Nagel study. In addition, in neither the Ashby nor the Cagen publications was it indicated that DES had been included as a positive control chemical in the initial published design of these studies by the Society of the Plastics Industry. The failure to find an effect due to the positive control chemical DES should have alerted these investigators that their studies had produced markedly different outcomes than the studies in which positive effects of this chemical chemicals had been reported. It is important for the reader to know that the Ashby and Cagen publications did not report that they had used different feed, that their positive control had failed, and that their control data were markedly different from data in the experiments that were supposedly being replicated.

In summary, our new findings show that some feed types can lead to elevated endogenous estradiol levels during fetal life and thus produce a physiological state that one would not want to have as a background against which to examine effects of estrogenic endocrine disruptors.

p 1-4. Paragraph 1. Question 1. The findings concerning the effects of bisphenol A on body weight reported in Nagel et al. (1997) and Howdeshell et al. (1999) are presented as contradictory by the panel. However, the procedures used were different. Specifically, Nagel fed pregnant females bisphenol A at 2 μg/kg/day and these females then nursed their pups. In contrast, Howdeshell used the same dose of bisphenol A, the pups were raised by foster mothers that had not been treated with bisphenol A. Dr. Taisen Iguchi and colleagues replicated the study of Howdeshell and reported their findings at the Japanese Endocrine Disruptor Society meeting, Yokahama, Japan, December, 2000. They found that prenatal exposure to bisphenol A accelerated puberty in female CD-1 mice, but the exposed mice were significantly lighter than controls, similar to the finding for body weight by Nagel. Importantly, the effect on puberty was the same as reported by Howdeshell (puberty was accelerated) even though the bisphenol A-treated females had lower body weights. In the Iguchi study, however, the pups remained with their treated mothers rather than being reared by foster mothers. These findings suggest that maternal exposure to bisphenol A could alter maternal behavior and/or lactation after parturition.

p 1-5, paragraph 2. Question 1. Note added in proof. "The statistics subpanel actually found in data provided in advance of a paper to be presented by vom Saal's group in Berlin in November, 2000, that the positional effects on body weight noted in the preceding paragraph were not reproduced in another study from the same laboratory".

<u>This statement is incorrect.</u> No data contradicting findings in Howdeshell et al. (1999) were provided to the panel. The findings presented in Berlin that are referred to here concerned only control males from different intrauterine positions, and differences in prostate weight, but not body weight, due to intrauterine position were reported. In neither the Howdeshell study or any prior study concerning intrauterine position were control animals from different intrauterine positions ever found to differ in body weight. All of these findings are consistent with regard to lack of effects on body weight due to intrauterine position of control animals.

p 1-6. Paragraph 2. Question 1. "The effects [of fetal exposure to bisphenol A on ventral prostate weight in male rats] observed [by Elswick et al., 2001] did not show a clear dose-response relationship".

This statement is inconsistent with the clear statement from the entire Low Dose Panel that non-monotonic dose-response curves can occur due to exposure to hormones or endocrine disrupting chemcials. Prostate weight is an endpoint that shows a non-monotonic dose-response relationship with fetal exposure to estrogenic chemicals. This study by Elswick reported no effect at the lowest and highest doses of bisphenol A, but they found effects at intermediate doses that were within the "low dose" range. The findings by Elswick for bisphenol A are thus similar to findings reported by vom Saal et al. (1997) for estradiol and DES. It is odd that given the subject of this meeting, when non-monotonic functions were found, the findings were referred to as not showing a "clear" dose-response relationship.

p 1-6. Paragraph 4. Question 1. "The evidence [regarding low dose effects of bisphenol A] is limited to one dose level". In addition to the paper by Sakaue et al. (2001) which is discussed below, at a meeting held at the Free University of Berlin in November, 2000, effects in the low part per billion range were reported for bisphenol A in snails, fish and frogs. Since the low dose workshop there have been numerous findings presented at meeting about low dose effects of bisphenol A. It is important to note that this subpanel did not include in their summary the two studies that had been published by Gupta that replicated exactly the findings of Nagel et al. (1997) with regard to effects of fetal exposure to bisphenol A on the prostate and epididymitis.

An article entitled" "Bisphenol-A Affects Spermatogenesis in the Adult Rat Even at A Low Dose" by Sakaue et al. (2001) has provided additional information about low dose effects of bisphenol A. Bisphenol A was fed (by gavage) to male Sprague Dawley rats from week 13-18 of life. A wide range of doses was administered (2 ng - 200 mg/kg/day) spanning 7 orders of magnitude. There was a steady decrease in daily sperm production (DSP) and efficiency (DSP/g testis) from 2 ng - 20 μg/kg/day, with a significant difference from controls occurring at the 20 μg/kg dose. Between the 20 μg/kg and 200 mg/kg doses, there was no further decrease in DSP or efficiency. This latter finding is similar to the effect of fetal exposure to ethinyl estradiol on efficiency of sperm production in male mice reported by Thayer et al. (2001). In addition, vom Saal et al. (1998) reported that

feeding pregnant mice bisphenol A from gestation day 11-17 resulted in a decrease in efficiency of sperm production at a dose of 20 μg/kg/day but not at 2 μg/kg/day. The findings of Sakaue in rats and vom Saal in mice thus show an effect of bisphenol A on efficiency of sperm production at the same dose of 20 μg/kg/day. However, the effect shown by Sakaue was based on exposure in adulthood, while the effect shown by vom Saal was based on exposure only during fetal life.

LITERATURE CITED

Ashby, J., H. Tinwell, and J. Haseman. Lack of effects for low dose levels of bisphenol A (BPA) and diethylstilbestrol (DES) on the prostate gland of CF1 mice exposed in utero. Reg Tox Pharm. 30:156-166, 1999.

Cagen, S.Z., J.M. Waechter, S.S. Dimond, W.J. Breslin, J.H. Butala, F.W. Jekat, R.L. Joiner, R.N. Shiotsuka, G.E. Veenstra, and L.R. Harris. Normal reproductive organ development in CF-1 mice following prenatal exposure to Bisphenol A. Tox. Sci. 11:15-29, 1999.

Elswick, B.A., Welsch, F. and Janszen, D.B. Effect of different sampling designs on outcome of endocrine disruptor studies. Reprod. Toxicol. 14:359-367, 2000.

Gupta, C. Reproductive malformation of the male offspring following maternal exposure to estrogenic chemicals. Proc. Soc. Exp. Biol. Med. 224:61-68, 2000.

Gupta, C. The role of estrogen receptor, androgen receptor and growth factors in diethylstilbestrol-induced programming of prostate differentiation. Urologic Researcch 28:223-229, 2000.

Howdeshell, K.L., Hotchkiss, A.K., Thayer, K.A., Vandenbergh, J.G. and vom Saal, F.S. Exposure to bisphenol A advances puberty. Nature 401:763-764, 1999.

Howdeshell, K.L. and vom Saal, F.S. Developmental exposure to bisphenol A: Interaction with endogenous estradiol during pregnancy in mice. American Zoologist 40: 429-437, 2000.

Nagel, S.C., vom Saal, F.S., Thayer, K.A., Dhar, M.G., Boechler, M. and Welshons, W.V. Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative *in vivo* bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ. Health Perspect. 105:70-76, 1997.

Sakaue M., Ohsako, S., Ishimura, R., Kurosawa, S., Kurohmaru, M., Hayashi, Y. Aoki, Y., Yonemoto, J. and Tohyama, C. Bisphenol-A Affects Spermatogenesis in the Adult Rat Even at A Low Dose. J. Occupational Health, July, 2001).

Thayer, K.A., R. Ruhlen, R., Howdeshell, K.L., Buchanan, D., Cooke, P.S., Welshons, W.V., Haseman, J. and vom Saal, F.S. Altered prostate growth and daily sperm production in male mice exposed prenatally to subclinical doses of 17α-ethinyl estradiol. Human Reproduction 16:988-996, 2001.

vom Saal, F.S. The interaction of background levels of endogenous estradiol with estrogenic and antiestrogenic chemicals. Paper presented at the Canadian Federation of Biological Societies meeting, Ottawa, Canada, June, 2001.

vom Saal, F.S., Timms, B.G., Montano, M.M., Palanza, P. Thayer, K.A., Nagel, S.C., Dhar, M.D., Ganjam, V.K., Parmigiani, S. and Welshons, W.V. Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc. Nat. Acad. Sci., 94:2056-2061, 1997.

vom Saal, F.S., Cooke, P.S., Palanza, P., Thayer, K.A., Nagel, S., Parmigiani, S. and Welshons, W.V. A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production and behavior. Toxicol. Industrial Health, 14:239-260, 1998.

From: Jimmy Spearow
Sent: Saturday, July 14, 2001 1:16 AM
To: liaison@starbase.niehs.nih.gov
Subject: Public Comment

JLScomment.pdf


ATT184996.txt

Dear NTP Office of Liaison and Scientific Review

Please include the attached public comment in the NTP Endocrine Disruptors Low-Dose Peer Review Final Report.

Thank you

Jimmy L. Spearow

Jimmy L. Spearow, Ph.D.
Associate Research Geneticist
Section of Neurobiology, Physiology and Behavior
Rm 196 Briggs Hall
University of California at Davis
Davis, CA 95616

Phone (530) 752-7548
Fax (530) 752-5582
Email jlspearow@ucdavis.edu

National Toxicology Program Office of Liason and Scientific Review
111 T.W. Alexander Drive,
P.O. Box 12233, MD A3-01,
Research Triangle Park, NC 27709.

July 13, 2001

Dear National Toxicology Program Office of Liason and Scientific Review
Please include this public comment to the Endocrine Disruptors Low dose Peer Review.

Studies from our laboratory requested by the NIEHS and included in the Endocrine Disruptor Peer Review demonstrated highly significant differences between strains of mice in sensitivity to endocrine disruption of testes weight and spermatogenesis by pubertal exposure to estradiol-17 beta (E2). Most Significantly, CD1 line mice were far more resistant to disruption by E2 in juvenile males than other strains of mice tested, including the commonly used C57BL/6J strain. The finding of strain differences in Susceptibility to Endocrine Disruption by Estrogenic Agents (SEDE) was heightened since CD1 mice and CD Rats have been previously selected for large litter size and are some of the most commonly used animal models for testing chemicals for endocrine disruptor activity.

This comment concerns the findings of much more recent studies in our laboratory which show that CD1 strain females mice are also much more resistant than B6 strain mice to pregnancy loss, i.e., the induction of abortions, by gestational exposure to E2.

The EPA's endocrine disruptor study program (EDSP) also plans to examine the effects of many potential estrogenic agents on reproductive function and development in tier-two multi-generation studies. Unfortunately, neither the NIEHS or the EPA has examined the effects of genetic differences in SEDE in such bioassays for estrogenic agents. Therefore, we examined the importance of strain differences in SEDE during gestation and neonatal development.

To avoid confounding phytoestrogen exposure, mice were a soy and alfalfa-free diet, Harland Teklad #2019 free choice. This diet is suitable for optimal growth, gestation and lactation. Adult females were paired with males of the same strain and checked for vaginal plugs daily (day of the plug = day 0 of pregnancy). Females were then implanted with silastic implants (10 mm / 20 gm body weight) containing 0, 0.15, 0.6, 2.5 or 40 µg crystaline E2/ 10 mm on day 9 of gestation (corresponding to an estimated exposure of approximately 0µg/Kg /day; 0.18 µg/Kg /day; 0.72 µg/Kg /day; 3 µg/Kg /day; or, 48 µg/Kg /day, respectively). Females were checked for litters daily and litters reduced to a maximum of 10 pups at birth. Maternal E2 exposure was continued through weaning, thereby providing a mid to late gestational exposure as well as an lactational exposure.

While we are still collecting data on this experiment, analysis of preliminary data revealed that the effects of strain and dose of E2 were both highly significant ($P<0.0004$) on total number of pups born and the number of live pups ($P<0.0001$). As shown in the table below, preliminary results to date indicate that B6 strain mice are far more sensitive than CD1 mice to the induction by estrogen of abortions and pregnancy loss.

Table 1.

Strain x µg E2 Implant Treatment Group	# plugged females	Total Number Pups per Plugged Female		SEM	Number Live Pups per Plugged Female		SEM
B6 x 0	8	4.3	±	1.3	4.1	±	1.2
B6 x 0.15	8	4.5	±	1.4	4.1	±	1.3
B6 x 0.6	8	3.6	±	1.2	3.4	±	1.2
B6 x 2.5	12	0.6	±	0.4	0.5	±	0.4
B6 x 40	8	0.0	±	0.0	0.0	±	0.0
CD1 x 0	8	13.4	±	0.9	12.9	±	0.9
CD1 x 0.15	5	11.6	±	2.9	11.6	±	2.9
CD1 x 0.6	6	12.0	±	0.9	10.5	±	0.5
CD1 x 2.5	6	10.2	±	1.2	8.7	±	2.0
CD1 x 40	6	9.5	±	1.0	5.2	±	1.9

Statistical analysis within B6 strain mice showed that the effect of E2 dose was highly significant on the total number of pups per plugged females ($P<0.002$) and the number of live pups per plugged female ($P< 0.002$). Post hoc Analysis using Duncan's multiple range test statistic showed that treatment of B6 females with 2.5 or 40 µ E2 / 20 gm body weight significantly reduced the total number of pups per plugged females and the number of live pups per plugged females ($P<0.05$). Note that treatment of 12 plugged B6 dams with as little as 2.5 µg E2 resulted in the birth of a total of only 6 live pups to date.

Statistical analysis within CD1 strain mice showed that E2 dose did not significant effect the total number of pups per plugged female ($P<0.28$). Nevertheless, E2 dose did effect the number of live pups per plugged CD1 female ($P< 0.05$). Post hoc Analysis using Duncan's multiple range test statistic showed that treatment of B6 females with 40 µ E2 / 20 gm body weight was required to significantly reduced the number of live pups per plugged females ($P<0.05$). While CD1 mice implanted with higher doses of E2 also showed an increased fetal mortality and reduced litter size relative to that of the 0 µg implant controls, many of the CD1 females, nevertheless produced litters in response to 2.5, 10 and even 40 µg E2 implants.

These findings show that CD1 are at least 16-fold more resistant than B6 strain mice to the disruption of gestation and induction of pregnancy loss by estrogen. We are currently repeating this experiment, adding more animals and extending our gestational and lactational exposure treatments to include much lower E2 doses. Nevertheless, these results strongly suggest that CD1 (and likely other large litter size selected animals) are also highly resistant to endocrine disruption of pregnancy by estrogenic agents.

These observations agree with the findings of Dr. Retha Newbold, US EPA reported at the Society for the Study of Reproduction (Newbold, 1998). In developing a model for studying the effects of gestational estrogenic exposure, Dr. Newbold exposed pregnant females from several strains of mice to DES starting at mid gestation. While the CD1 strain mice were able to produce offspring following DES treatment, all of the other strains of mice aborted. Thus, Dr. Newbold used the resulting DES treated CD1 pups to establish an animal model for the effects of gestational exposure to estrogenic agents on reproductive development . Thus, our findings of strain differences in sensitivity to the disruption of pregnancy by estradiol are in agreement with Dr. Newbold's findings of similar strain differences in sensitivity to DES.

These preliminary, but statistically highly significant, findings strongly suggest that the sole use of large litter size selected, strains of laboratory animals in the proposed EDSP multi-generation testing protocol is likely to dramatically underestimate the effects of agents with endocrine disrupting activity on reproductive function, i.e. the ability to maintain pregnancy. Since the results of tier two multi-generation testing protocol in the EDSP will supercede the tier one tests, it is critically important that a sensitive strain of laboratory animals be utilized for such testing. The use of resistant strain animals is very likely to underestimate the effect of endocrine disrupting chemicals on sensitive individuals. These data also show the importance of considering quantitative differences in fecundity, i.e., litter size, as a critical endpoint when testing chemicals for estrogenic or endocrine disrupting potency.

Jimmy L. Spearow, Ph.D.

Associate Research Geneticist

Literature Cited:

Retha Newbold. 1998. Averse effects of estrogenic compounds result from exposure during critical periods of development. Endocrine Disrupters Symposium. 31st Annual Meeting of the Society for the Study of Reproduction. College Station Texas, Aug.11, 1998. Biol. Reprod. 58:Sup1.

From: Thayer, Kristina
Sent: Monday, July 16, 2001 11:55 AM
To: 'liaison@starbase.niehs.nih.gov'
Cc: Liroff, Rich; Colborn, Theo
Subject: WWF Comments on the NTP Endocrine Disruptors Low-Dose Peer Review Final Report

WWF Comments on Low Dose Peer Review.doc

Please find attachted WWF Comments on the NTP Endocrine Disruptors Low-Dose Peer Review Final Report.

Kristina Thayer
Wildlife and Contaminants Program
World Wildlife Fund
1250 24th Street, NW
Washington, DC 20036
Ph: 202-822-3473
fax: 202-530-0743
kristina.thayer@wwfus.org

<<WWF Comments on Low Dose Peer Review.doc>>

NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive
P.O. Box 12233
MD A3-01
Research Triangle Park, NC 27709

RE: NTP final report from the Endocrine Disruptors Low-Dose Peer Review

World Wildlife Fund (WWF) submits the following comments regarding the NTP Endocrine Disruptor Low-Dose Peer Review.

WWF is a non-profit organization with over 1.2 million members in the U.S. WWF is dedicated to using the best available scientific knowledge to preserve the diversity and abundance of life on Earth by conserving endangered spaces, safeguarding endangered species, and addressing global threats to the planet's web of life.

We have four major points regarding the final report of the Low-Dose Peer Review meeting:

- The studies discussed by the panel and presented in subsequent journal publications clearly indicate that certain chemicals can have endocrine disrupting effects at low doses. Many of the studies which found effects at low doses incorporated detailed analysis of endocrine target tissues later in life following a developmental exposure.

- Most of these findings came from assessments of endpoints not historically evaluated in toxicological testing protocols.

- These finding should cause EPA to elevate the importance of endocrine data gaps when determining the FQPA Safety Factor for pesticide tolerances, especially if reproductive studies were conducted prior to 1996 guideline changes which incorporated more estrogenic- and androgenic-related endocrine endpoints.

- EPA should recognize that current guidelines do not adequately address thyroid or adrenal disruption.

World Wildlife Fund
1250 Twenty-Fourth St., NW Washington, DC 20037-1132 USA
Tel: 202.293.4800 Fax: 202.887.5293
www.worldwildlife.org

Affiliated with World Wide Fund for Nature

COMMENTS ON THE BISPHENOL A SUBPANEL

Recent Findings in Laboratory and Wildlife Species

The peer-reviewed literature on bisphenol A has exploded in recent years. Since the October 2000 Low Dose Peer Review Meeting many additional low dose findings in laboratory species have been published. Brief summaries of these papers can be found in Table 1. These citations do not include all recent bisphenol A papers, but only those detecting significant effects at low doses. In addition, recent low dose *in vitro* studies are summarized in Table 2.

Although the effects of bisphenol A on wildlife species were not considered at the low dose meeting, there are numerous reports of low dose effects (≤ 5 ppm) in these species (Table 3). While human health effects and wildlife effects are considered separately in the regulatory process, these wildlife findings provide additional evidence supporting low dose activity of bisphenol A.

COMMENTS ON BIOLOGICAL FACTORS AND STUDY DESIGN SUBPANEL

The Current Multigeneration Test May Be Inadequate To Detect Reproductive Tract Malformation and Estrogenic or Androgenic Disruption

We urge EPA to consider the Biological Factors and Study Design subpanel's comments on the multigeneration test when determining FQPA safety factors. Specifically, the subpanel made several points concerning the multigeneration study design. First, examination of only one pup/sex/litter of F_1 animals in adulthood is a major shortfall resulting in inadequate power to detect significant reproductive tract malformations. Second, certain sensitive endpoints are not evaluated at all (i.e. nipple retention) or are triggered in F_2 animals based on F_1 results (anogenital distance). Finally, the subpanel noted that both linuron and di(isononyl) phthalate (DINP) have been shown to have endocrine activity even though credible multigenerational and prenatal studies were negative using standard designs.

Although multigeneration guideline changes in 1996 (Federal Register: October 31, 1996 Volume 61, Number 212, pages 56274-56322) included more endpoints responsive to estrogen or androgen disruption, they still do not include all endpoints suggested by the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) (Federal Register: December 28, 1998 Volume 63, Number 248, pages 71541-71568). In particular, EDSTAC recommended the inclusion of accessory sex organ function (secretory products), nipple development and retention, androgen and estrogen levels, LH and FSH levels, and testis descent (Table 4).

The Multigeneration Test Scarcely Addresses Thyroid and Adrenal Disruption

Current multigeneration guidelines do not address thyroid disruption. For example, measurement of thyroid weight is not explicitly specified in F0 or F1 animals although such tissues as the uterus, ovaries, testes, epididymides, prostate, seminal vesicle, brain, liver, kidney, adrenals, spleen are listed (Federal Register: October 31, 1996 Volume 61, Number 212, pages 56274-56322). EDSTAC recommended the inclusion of many endpoints responsive to thyroid

disruption in the multigenerational study, primarily as they relate to neurological development and T3, T4 and TSH levels (Table 4). Although dioxin and dioxin like compounds were excluded from the Low Dose Peer Review, it is worth reiterating that TCDD and various PCBs have been shown to interfere with thyroid regulation at low doses.

Certain developmental endpoints in the multigeneration study were determined by EDSTAC to be potential indicators of thyroid disruption (growth, body weight, food consumption, food efficiency, developmental abnormalities, perinatal mortality). These types of effects are commonly observed in developmental or multigeneration studies, although they have never been evaluated as "potential" endocrine disruption. Rather, they are only considered to be significant in the FQPA safety factor determination if they occur at doses below where maternal or parental toxicity was observed, which is proposed to indicate increased sensitivity. This approach actually does nothing to address potential increased sensitivity, which would require comparison of the same endpoints in developmentally exposed and adult animals. It is worth noting that the synthetic estrogen diethylstilbestrol (DES) does not cause these effects on traditional developmental toxicology endpoints at doses known to affect reproductive organs. Rather, the most sensitive indicators of DES action *in vivo* are based on analysis of reproductive tissue function. In addition, DES has been shown recently to cause effects in offspring of exposed animals, though the affected animals were never exposed directly to DES during development (Newbold,1998, Newbold,2000).

EPA Must Consider Endocrine Disruptor Data Gaps Significant When Determining FQPA Safety Factors Because Current Multigeneration Tests Do Not Adequately Assess Endocrine Disruption

In 1999, the Office of Pesticide Programs (OPP) released a document titled "The Office of Pesticide Programs' Policy on Determination of the Appropriate FQPA Safety Factor(s) for Use in the Tolerance-Setting Process" (http://www.epa.gov/scipoly/sap/1999/may/10xpoli.pdf). In this draft the OPP indicated:

> " ... OPP's default position when a newly identified study is lacking will not be that an additional uncertainty factor is necessarily mandated. Rather, OPP's approach will be to evaluate the existing toxicological database on a pesticide to determine if the absence of the new data is ***so key as to warrant an additional uncertainty factor to protect the safety of infants and children***. (p 20-21) Emphasis added.

The FIFRA SAP panel also encouraged OPP to consider not only the number of core database studies that were missing, but in addition the relative importance of their omission (http://www.epa.gov/scipoly/sap/1999/may/final.pdf). Furthermore, in the final report the panel indicated that, if EPA does not have all relevant information regarding developmental neurotoxicity, immunotoxicity and effects on the endocrine system, the Agency "faces a special presumption against relieving the 10X safety factor" (p. 14). The report also indicated that the requirements of the core database be revisited every few years to determine if it is adequate (p. 16).

We believe OPP has demonstrated its concern for ED effects in the tolerance reassessment process as evidenced by a willingness to incorporate endocrine disruption considerations into the safety factor determination *when data exist* indicating potential endocrine disruption (such as for

methyl parathion). Unfortunately, these types of data are not available for the majority of pesticides. Ultimately, EPA should require the registrant to subject pesticides to standardized and validated screens and tests for endocrine disruption, once these have been agreed upon, and registration should be reassessed on the basis of these results. This is especially true if the multigenerational studies conducted by the registrant were conducted prior to 1996.

In the interim, we urge OPP to consider endocrine disruptor data gaps significant when evaluating FQPA safety factor concerns. OPP faces an unenviable struggle with consistent application of the FQPA safety factor, but it must not discount endocrine disruption data gaps simply for the sake of consistency with prior tolerance decisions, particularly if doing so ignores new areas of science especially relevant to embryos, fetuses, infants and children.

In summary, we believe the EPA should take recent low dose findings as an indication that current testing paradigms are not sufficient to detect potential low dose endocrine disruption. We understand that it will take time until low dose responsive endocrine related endpoints are incorporated into testing guidelines. In the meantime, we urge EPA to consider endocrine disruption to be a significant data gap worthy of consideration in the FQPA Safety Factor determination for two reasons. First, embryos, fetuses, infants and children are expected to be more sensitive to endocrine disruption effects than adults. Second, multi-generational studies do not appear to be sufficient to detect endocrine disruption in all cases (i.e. linuron, DINP and bisphenol A). This is especially true of multigeneration studies conducted before 1996 guideline changes which added additional endpoints responsive to estrogenic and/or androgenic endocrine disruption. In addition, current guidelines do not include endpoints designed to specifically address thyroid or adrenal gland disruption.

We appreciate the opportunity to provide these comments in response to the NTP Endocrine Disruptor Low-Dose Peer Review.

Sincerely,

Kristina Thayer, PhD
Program Scientist
Wildlife and Contaminants Program
(202) 822-3473
kristina.thayer@wwfus.org

Theo Colborn, PhD
Senior Program Scientist and Director
Wildlife and Contaminants Program
(202) 778-9643

Documentation:

Arukwe A, Celius T, Walther BT, Goksoyr A. Effects of xenoestrogen treatment on *zona radiata* protein and vitellogenin expression in Atlantic salmon (*Salmo salar*). Aquatic Toxicology 49:159-170(2000).

Christian M, Gillies G. Developing hypothalamic dopaminergic neurones as potential targets for environmental estrogens. J Endocrinol 160:R1-6.(1999).

Gupta C. Reproductive malformation of the male offspring following maternal exposure to estrogenic chemicals. Proceedings of the Society of Experimental Biology & Medicine 224:61-68(2000).

Lindholst C, Pedersen KL, Pedersen SN. Estrogenic response of bisphenol A in rainbow trout (*Oncorhynchus mykiss*). Aquatic Toxicology 48:87-94(2000).

Markey CM, Luque E, Munoz de Toro M, Sonnenschein C, Soto AM. In utero exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. Biology of Reproduction(in press).

Markey CM, Michaelson CL, Veson EC, Sonnenschein C, Soto AM. The mouse uterotrophic assay: a reevaluation of its validity in assessing the estrogenicity of bisphenol A. Environ Health Perspect 109:55-60(2001).

Metcalfe CD, Metcalfe TL, Kiparissis Y, Koenig BG, Khan C, Hughes RJ, Croley TR, March RE, Potter T. Estrogenic potency of chemicals detected in sewage treatment plant effluents as determined by *in vivo* assays with japanese medaka (*Oryzias latipes*). Environmental Toxicology & Chemistry 20:297-308(2001).

Newbold RR, Hanson RB, Jefferson WN, Bullock BC, Haseman J, McLachlan JA. Increased tumors but uncompromised fertility in the female descendants of mice exposed developmentally to diethylstilbestrol. Carcinogenesis 19:1655-63.(1998).

Newbold RR, Hanson RB, Jefferson WN, Bullock BC, Haseman J, McLachlan JA. Proliferative lesions and reproductive tract tumors in male descendants of mice exposed developmentally to diethylstilbestrol. Carcinogenesis 21:1355-63.(2000).

Ramos JG, Varayoud J, Sonnenschein C, Soto AM, Munoz de Toro M, Luque H. Prenatal exposure to low doses of bisphenol A alters the periductal stroma and glandular cell function in rat ventral prostate. Biology of Reproduction(in press).

Rubin BS, Murray MK, Damassa DA, King JC, Soto AM. Perinatal Exposure to Low Doses of Bisphenol A effects Body Weight, Patterns of Estrous Cyclicity, and Plasma LH Levels. Environ Health Perspect 109:675-680(2001).

Sakaue M, Ohsako S, Ishimura R, Kurosawa S, Kurohmaru M, Hayashi Y, Aoki Y, Yonemoto J, Tohyama C. Bisphenol A affects spermatogenesis in the adult rat even at low doses. Journal of Occupational Health(in press).

Takai Y, Tsutsumi O, Ikezuki Y, Hiroi H, Osuga Y, Momoeda M, Yano T, Taketani Y. Estrogen receptor-mediated effects of a xenoestrogen, bisphenol A, on preimplantation mouse embryos. Biochem Biophys Res Commun 270:918-921(2000).

Takai Y, Tsutsumi O, Ikezuki Y, Kamei Y, Osuga Y, Yano T, Taketan Y. Preimplantation exposure to bisphenol A advances postnatal development. Reproductive Toxicology 15(2000).

Yamasaki K, Sawaki M, Takatsuki M. Immature rat uterotrophic assay of bisphenol A. Environ Health Perspect 108:1147-1150(2000).

Yokota H, Tsuruda Y, Maeda M, Oshima Y, Tadokoro H, Nakazono A, Honjo T, Kobayashi K. Effect of bisphenol A on the early life stage in Japanese medaka (*Oryzias latipes*). Environmental Toxicology & Chemistry 19:1925-1930(2000).

TABLE 1: Recent Low Dose Bisphenol A Studies in Laboratory Species

Reference	Species	Dosing Paradigm	Positive Control	Effects
(Markey,in press)	☜ mice (CD-1)	0, 25, 250 µg/kg/d GD9-birth via subcutaneous pump implanted in dam	none	25 and 250µg/kg/d: opposite effects on mammary gland terminal end bud elongation at 1 month of age. Although neither differs significantly from the control group, they differ from each other. 25, 250 µg/kg/d: ↑ relative area of mammary ducts, terminal ducts, terminal end buds (25µg only) and alveolar buds at 6 months of age. 25, 250 µg/kg/d: ↓ incorporation of BrdU into mammary epithelial cells (PND 10, but not at 1 or 6 months). 25, 250 µg/kg/d: dampened age-related changes in bromodeoxyuridine (BrdU) incorporation in mammary epithelial cells. BrdU incorporation significantly increased at 1 month and decreased at 6 months compared to PND10 levels in control animals. This did not happen in BPA-treated mice. 25, 250 µg/kg/d: ↑ incorporation of BrdU into mammary gland stroma (6 months, but not at PND10 or 1 month). 250 µg/kg/d: ↓ incorporation of BrdU into mammary gland stroma (1 month, but not at PND10).
(Ramos,in press)	_ rat (Wistar derived)	0, 25, 250 µg/kg/d GD8-GD23 via subcutaneous pump implanted in dam	none	25, 250 µg/kg/d: ↓ prostatic acid phosphatase (PAP) staining in ventral prostate epithelial cells on PND30. 25, 250 µg/kg/d: ↑ vimentin, a fibroblast cell marker, staining in ventral prostate periductal ring on PND30. 25, 250 µg/kg/d: ↓ α-smooth muscle actin staining in ventral prostate periductal zone on PND30. 25, 250 µg/kg/d: ↓ androgen receptor staining in ventral prostate periductal stroma on PND30.
(Sakaue,in press)	_ rat (Sprague-Dawley)	0, 2ng, 20ng, 200ng, 2µg, 20µg, 200µg, 2mg, 20mg, 200mg/kg/d adult (13 weeks) for 6 days via gavage	none	20µg, 200µg, 2mg, 20mg, 200mg/kg/d: ↓ daily sperm production and daily sperm production/ gram testis. 20 µg/kg (single injection): altered staining of unidentified proteins in three regions in a two dimensional gel analysis of testicular cytosol

TABLE 1 continued: Recent Low Dose Bisphenol A Studies in Laboratory Species

Reference	Species	Dosing Paradigm	Positive Control	Effects
(Rubin,2001)	♀rat (Sprague-Dawley)	0, 1, 10 mg/L (~0, 0.1, 1.2 mg/kg/d) GD6-weaning via drinking water	none	0.1, 1.2 mg/kg/d: ↑ body weight on PND4, 7 and 11. 1.2 mg/kg/d: ↑ incidence of irregular estrous cycling at 4 and 6 months. 1.2 mg/kg/d: ↓ LH levels in ovariectomized females. *BPA did not affect day of vaginal opening or anogenital distance. BPA also was not uterotropic when administered to ovx postpubertal rats (1, 10, 100 mg/L)
(Markey,2001)	♀ mice (CD-1)	0.1, 0..5 1, 5, 50, 75, 100 mg/kg/d PND 23-25	Estradiol 5μg/kg/d	0.1, 100 mg/kg/d: acceleration of vaginal opening. 0.5 mg/kg/d: ↑ in body weight. 5 mg/kg/d: ↓ relative area of uterine lamina propria. 5, 75, 100 mg/kg/d: ↑ luminal epithelial height. 75, 100 mg/kg/dy - ↑ lactoferrin expression. 100 mg/kg/d: ↓ in body weight and ↑ uterine wet weight.
(Yamasaki,2000)	♀ rat (Crj:CD/SD) paper includes plasma BPA concentrations one hour after last administration	0, 8, 40, 160 mg/kg/d PND18 - PND20 (subcutaneous or gavage)	none	8, 40, 160 mg/kg/d (subcutaneous injection):↑ uterine wet weight, blotted weight, relative wet weight and relative blotted weight. 160, 800 mg/kg/d (oral ; gavage):↑ uterine relative wet weight and relative blotted weight. 800 mg/kg/d: ↑ uterine wet weight and blotted weight.
(Gupta,2000)	♂ mice (CD-1)	*In vivo* 50 μg/kg/d GD16-GD18 (oral) *In vitro* 5.0, 50 pg/ml culture of GD17 fetal urogenital sinus for 6 days	*In vivo* DES 0.1, 200 μg/kg/d *In vitro* DES 0.1, 0.5 pg/ml	*In vivo* 50 μg/kg/d: ↑ anogenital distance and ↑ prostate weight in males at PND 3, 21 & 60. 50 μg/kg/d: ↑ epididymal weight at PND 60 (not seen with 0.1 μg/kg DES). 50 μg/kg/d: ↑ androgen receptor (AR) binding activity on PND 21, 60. *In vitro* 50 pg/ml: ↑ prostate growth and branching in absence and presence of testosterone. 50 pg/ml: ↑ AR binding.

*GD = gestation day; PND = post-natal day

TABLE 2: Recent *In Vitro* Low Dose Bisphenol A Studies

Reference	Species	Dosing Paradigm	Positive Control	Effects
(Christian,1999) *reference to preliminary results in discussion (p 160)	rat (GD 18 hypothalami)	1 nM for 2-3 weeks	Estradiol (10^{-14}, 10^{-12}, 10^{-10}, 10^{-8}, 10^{-6} M)	1 nM: ↑ dopamine uptake
(Takai,2000)	mice (B6C3F1 embryos)	1 nM, 100 uM for 2 days In vitro: 2-cell embryos exposed for 48 hrs (after exposure embryos implanted in non-treated mice)	none	1 nM: ↑ developmental rate (↑ number of blastocytes, ↓ number of morula) 100 uM: ↓ developmental rate (↓ number of blastocytes, ↑ number of morula)) 1 nM, 100 uM: ↑ BW at weaning.
(Takai,2000)	mice (B6C3F1 embryos)	0, 0.1, 1, 3, 10, 100, 1000, 10,000, 100,000 nM In vitro: 2-cell embryos exposed for 48 hrs to obtain blastocysts	none tamoxifen added to block ER-mediated effect	3nM: ↑ developmental rate (↑ percentage of 8-cell embryos at 24 hours). 1 and 3nM: ↑ developmental rate (↑ percentage of blastocysts at 48 hours). 100 uM: ↑ developmental rate (↓ percentage of blastocysts at 48 hours). * administration of tamoxifen appeared to antagonize low and high dose BPA effects

TABLE 3: Recent Low Dose Bisphenol A Studies in Wildlife Species

(Metcalfe,2001) *this study measured BPA concentrations in test solutions	Japanese Medaka (*Oryzias latipes*)	0, 10, 50, 100, 200 µg/L 1 day post-hatch to ~ 100 days post hatch	17β estradiol, 17α ethinyl estradiol, estriol, estrone	10 µg/L: ↑ incidence of testis-ova. 50, 100, 200 µg/L: testicular abnormalities. 100, 200 µg/L: ↑ conditioning factor (weight divided by total length).
(Yokota,2000) *this study measured BPA concentrations in test solutions	Japanese Medaka (*Oryzias latipes*)	0, 2.28, 13, 71.2, 355, 1,820 µg/L fertilized egg to 60days post hatch	none	1,820 µg/L: ↓ total length and weight, no males identified, and ↑ incidence of testis-ova.
(Arukwe,2000)	Atlantic Salmon (*Salmo salar*)	0, 1, 5, 25, 125 mg/kg juvenile (~ 1 year) (one ip injection)	17β estradiol	1, 5, 25, 125 mg/kg: ↓ EROD liver activity. 25, 125 mg/kg: ↑ plasma vitellogenin levels. 25, 125 mg/kg: ↑ plasma zona radiata protein levels.
(Lindholst,2000) * includes actual BPA concentrations and BPA muscle and liver levels	Rainbow Trout (*Oncorhynchus mykiss*)	0, 10, 40, 70, 100, 500 µg/L juvenile (90-130 g) for 12 days	17β estradiol	70, 100, 500 µg/L: ↑ percentage of animals with vitellogenin levels more than twice than level of any fish on day 0 or control fish on days 6 and 12 500 µg/L: ↑ plasma vitellogenin levels.

Table 4. Comparison of Current Endocrine Related Endpoints Assessed in Multigenerational Studies and Additional Endpoints Recommended by EDSTAC (Federal Register: December 28, 1998 Volume 63, Number 248, pages 71541-71568).

Test	current endocrine related endpoints	recommended endocrine related endpoints
2-generation mammalian reproductive toxicity	**Estrogen Agonist/Antagonist** gonad development (size, morphology, weight) accessory sex organ (ASO) development (weight ± fluid, histology) acquisition of vaginal patency (VP) preputial separation (PPS) fertility fecundity time to mating mating and sexual behavior ovulation estrous cyclicity gestational length abortion premature delivery dystocia (difficult delivery) spermatogenesis epididymal sperm numbers and morphology testicular spermatid head counts daily sperm production (DSP) efficiency of DSP gross and histopathology of reproductive tissues anomalies of genital tract viability of conceptus in utero (prenatal demise) survival and growth of offspring maternal lactational behaviors (e.g. nursing, pup retrieval, etc.) **Androgen Agonist/Antagonist** altered apparent sex ration (based on anogenital distance) malformations of the urogenital system altered sexual behavior changes in testis and ASO weights effects on sperm numbers, morphology, etc. retained nipples in male offspring altered AGD (now triggered from VP/PPS) male fertility agenesis of prostate changes in androgen dependent tissues in pups and adults (not limited to sex accessory glands) **Thyroid Agonist/Antagonist** growth, body weight food consumption, food efficiency developmental abnormalities perinatal mortality testis size and DSP vaginal patency preputial separation	**Estrogenic/Androgenic** ASO function (secretory products) nipple development and retention androgen and estrogen levels LH and FSH levels testis descent **Thyroid** neurobehavioral deficits TSH, T3, T4 thyroid weight and histology (e.g. goiter) pinna detachment surface righting reflex eye opening acquisition of auditory startle negative geotaxis mid-air righting reflex motor activity on PND 13, 21 postwean includes motor activity on PND21 and postpuberty ages (sex difference) learning and memory PND60 (acitive avoidance/water maze, brain weight (absolute), whole and cerebellum brain histology.

From: Angelina Duggan
Sent: Monday, July 16, 2001 3:31 PM
To: liaison@starbase.niehs.nih.gov
Subject: ACPA Comments: NTP Low Dose Peer Review

ACPA NTP Comments.doc

** High Priority **

Dear NTP Representative:

The American Crop Protection Association is sending comments according to the requested July 16 deadline.

These comments have also been faxed and an original signed copy will be mailed to the NTP Office of Liaison and Scientific Review.

Sincerely yours,

Angelina J. Duggan, Ph.D.
ACPA Director of Science Policy
(202) 872-3885 (tel)
(202) 463-0474 (fax)

American Crop Protection Association

July 16, 2001

NTP Office of Liaison and Scientific Review
111 TW Alexander Drive
PO Box 12233
MD A3-01
Research Triangle Park, NC 27709
FAX: (919) 541-0292
liaison@starbase.niehs.nih.gov.

Dear NTP Representative:

RE: NATIONAL TOXICOLOGY PROGRAM (NTP) ENDOCRINE DISRUPTORS LOW-DOSE PEER REVIEW FINAL REPORT

The American Crop Protection Association (ACPA) appreciates the opportunity to provide written comments on the NTP Endocrine Disruptors Low Dose Peer Review Final Report. These comments are being provided to you as a Word Documents by both FAX and E-mail. A signed original will also be mailed to the NTP Office of Liaison and Scientific Review.

ACPA is a not-for-profit trade organization of U.S manufacturers, formulators and distributors of crop protection products. Our members produce, sell and distribute virtually all of the active ingredients used in such products.

ACPA promotes the environmentally sound use of crop protection products for the economical production of safe, high quality, abundant food, fiber and other crops. We have actively participated in the development of a endocrine disruption screening and testing program that requires validation of all screens and tests, sound science and weight-of-the-evidence in decision-making. Our industry representatives served on the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) and contributed technical expertise and policy perspective to the EDSTAC Working Groups for Priority Setting, Screening and Testing, and Communications and Outreach. We also participated in writing the EDSTAC consensus report, and have continued to work with other stakeholders to implement the EDSTAC program by participating in the EPA Priority Setting Workshops, on EPA's Endocrine Standardization & Validation Taskforce (SVTF) Steering Committee and in the recent organizational meeting for the Endocrine Disruptor Methods Validation Subcommittee (EDMVC).

ACPA continues to support a thorough evaluation of the endocrine disruption low-dose hypothesis that includes a balanced and disciplined evaluation of all available data and conclusions based on the weight-of-the-evidence. On October 10-12, 2000, we also attended and provided oral and written comments for the EPA, NTP and NIEHS Low Dose Peer Review. These comments are reissued and included as an addendum to this document.

Sincerely yours,

Angelina J. Duggan, Ph.D.
ACPA Director of Science Policy
(202) 872-3885 (tel)
(202) 463-0474 (fax)
angelina@acpa.org

cc: EPA - Anthony Maciorowski, Jane Smith, Gary Timm, Vanessa Vu

AMERICAN CROP PROTECTION ASSOCIATION COMMENTS

NATIONAL TOXICOLOGY PROGRAM ENDOCRINE DISRUPTORS LOW-DOSE PEER REVIEW FINAL REPORT

Submitted by
Angelina J. Duggan, Ph.D.
ACPA Director of Science Policy
July 16, 2001

ACPA COMMENTS

Introduction

The member companies of ACPA are dedicated to protecting the environment and serving people through science in agriculture. In the spirit of that commitment, ACPA has supported the implementation of an endocrine disruptor screening and testing program and a balanced evaluation of the low dose hypothesis that is based on validation, sound science and the weight-of-the-evidence in decision-making.[1] Herein, in response to the National Toxicology Program's "Report of the Endocrine Disruptors Low Dose Peer Review" (Report), we are submitting comments that include the following:

a) The low-dose issue relative to the crop protection industry
b) Genesis of the low-dose hypothesis
c) Original charge and the how the NTP peer review process was conducted
d) Major concerns about the peer review and the final Report
e) Summary and recommendations

ACPA would like to thank the NTP and EPA for the opportunity to provide comments on the low-dose peer review Report. However, we would also like to express our grave concern that NTP has published the proceedings from the October 2000 low-dose peer review as a final, rather than a draft, report. In our opinion, it is not sufficient that the NTP will attach comments to the final Report as part of the public record of these proceedings. As a government agency, NTP has an obligation to respond to and consider critical comments from stakeholders and the public and to correct possible inaccuracies before issuing a final Report. Revisions would substantially improve the NTP Report and provide needed balance in evaluating the low-dose hypothesis.

Additionally, issuing a "final report" implies, or may lead some to assert, that the low dose issue is resolved and therefore, should be implemented. Indeed this is not the case. As originally charged, the primary objective of the low dose peer review was <u>not</u> to consider whether there <u>may</u> be low-dose effects, but rather "to evaluate the likelihood and significance of these and/or other potential low-dose effects for humans."

Since the NTP will forward comments to EPA as an attachment to the Report, ACPA directs comments about the low dose hypothesis and the NTP Report to NTP, EPA, interested stakeholders and the public.

The Low Dose Issue Relative to the Crop Protection Industry

ACPA companies develop extensive data to ensure the safety and effectiveness of their products. Both the federal and individual state governments regulate the use and registration of crop protection products, and statutory mandates require at least 120 or more separate tests that include acute, developmental, multi-generation reproduction, sub-chronic and chronic studies to ensure safety and effectiveness.[2] EDSTAC recently recommended new endocrine-related endpoints and developmental landmarks to the rat multi-generation reproduction study.[3] The crop protection industry has supported the EDSTAC process and its final report

recommendations. However, adding dose levels, additional endpoints points or making other changes to an already overburdened study design should be done only for very good reasons.

Some of the more strident proponents of the low-dose hypothesis criticize the use of the NOAEL, because they believe adverse effects may reappear at levels below the NOAEL. Changes in the toxicology testing paradigm for regulatory testing would be appropriate only if low-dose hypothesis was demonstrated in a scientifically disciplined, convincing, clear, reproducible, reliable and meaningful way.

Although limited studies from some laboratories, using only a few animals, have reported evidence for low-dose observations with selected compounds, for example bisphenol A (BPA). However, in our opinion, these studies should be viewed as research observations.[4, 5] Recent publications, two of which are multi-generation toxicology studies, independently conducted by industry and the Japanese government, did not confirm the low-dose research observations for BPA. [6, 7, 8, 9, 10, 11, 12] These recent rat multi-generation tests adhere to OECD and EPA guideline protocols that are relied upon and accepted by regulators worldwide. In both cases, there was no translation of the low-dose research effects to larger-scale more robust whole animal studies that evaluated multiple reproductive hormonal endpoints from many animals with a range of broad range of high to very low doses over two to three generations. Moreover, it has still not been proven that the observed low-dose effects in mice from the research observations cause adverse outcomes or disease in the same test animals or that these same low-dose effects in mice indicate potential disease or adverse human health effects.

ACPA is committed to providing the most useful and reliable toxicology data that is practical and based upon valid scientific principles. We do not believe that the expenditure and waste of additional laboratory resources or test animals for low-dose regulatory testing are justified.

Genesis of the Low-Dose Hypothesis

The low-dose issue can be linked directly to several key publications alleging "low-dose" effects. In 1997, it was reported that prenatal exposure to DES caused a <u>decrease</u> in prostate weight in male mice at high levels. The high-dose effects were considered by the authors to be indicative of an adverse effect. The Nagel et al. studies established a NOAEL for decreased prostate weight. That study then reported modest <u>increases</u> in prostate weight at dose levels orders of magnitude <u>below</u> the NOAEL that the authors also considered adverse.[13] These studies had very few mice (5 to 7 males per group) and the increase was rather small (within the normal range), but some of the changes were statistically significant. The authors have never repeated those studies, but nonetheless, strongly allege that their low-dose observations are indicative of an adverse effect and that prenatal estrogen exposure exhibits a bimodal dose response curve that cannot be detected by conventional toxicology studies.

Another example repeatedly held forth by the same research group as evidence of low-dose effects is a limited study on BPA.[4] This study used only two dose levels and 5 to 7 male mice per group. Both dose levels are below the NOAEL for other effects, and were reported to show an increase in prostate weight. Other researchers, including the original authors, have never been able to repeat these studies. It should also be stressed that the many other attempts have been made with much better designs.

It is not possible to show a bimodal response with only two dose levels. Despite these and other significant issues, the low-dose theory or possible presence of a bimodal dose response curve received vocal support by some members of the EPA EDSTAC committee. This support led to a compromise deferring any recommendation on how EPA should deal with the lowdose issue to further consideration by EPA. EDSTAC recommended that EPA should continue to study the issue. This led EPA to ask the NTP to convene a low-dose peer review.

It is important to note that we are talking about changes in a single end point - e.g., prostate weight. While adverse effects are seen at high levels and a NOAEL is established, it is alleged that new and adverse effects are observed at lower levels on that same target organ. This is not a question of whether or not endocrine active agents can cause biological effects at low levels. Normal hormone activity is typically seen at very low levels - this not new. It is also not a question of whether higher levels of hormonally active agents can cause permanent adverse effects in adults who were exposed during development. This is well known and addressed in other toxicology study protocol issues.

The low-dose issue raises the question whether or not hormonally active agents have a unique capacity to cause adverse effects on a target organ at levels far below (orders of magnitude) the demonstrated NOAEL for that target organ. The low-dose issue hypothesizes a bimodal dose response curve for a specific end point with an adverse effect in a target organ that disappears at intermediate doses, but is then again affected at much lower levels.

Charge to the NTP Panel

The EPA provided guidance to the NTP and the public on the scope, intent and procedural rules for the low-dose peer review. The peer review was supposed to focus on the interpretation of major data sets showing or refuting effects at low doses. EPA defined low doses for this purpose as "doses below the currently accepted No Observed Adverse Effect Level [NOAEL] for that substance." EPA also spelled out the intent to evaluate "the presence or absence of low dose effects in specific studies, then evaluate the likelihood and significance of these and/or other potential low does effects to humans."

The main topic of the review included considering "evidence for defining the shape of the dose response curve's for endocrine active substances in the low dose region." The review was expected to "examine all evidence, including such things as relevant pharmacokinetic and mechanistic information, which may have a bearing on the low dose issue." EPA intended closure on this "central issue of whether there are sufficient grounds to change the traditional dose setting paradigm for endocrine active substances."

Other issues that were not considered central by EPA, but open to discussion, included existence of the inverted U-shaped dose response curves as a general phenomenon in toxicology, completeness of the list of endpoints into generation toxicity tests, and the definition of adversity. These issues and other major comments about the review are raised in more detail in these comments below.

NTP Low-Dose Peer Review Process
The NTP designed the peer review process to select and consider the critical low-dose studies, and required the authors to provide raw data from those studies. The contributing authors made a presentation at the public meeting and were asked to respond to questions from the sub-panels. This process allowed the NTP statistics sub-panel to reevaluate the data independently from the original authors. The NTP developed an excellent independent statistical review of the data, but left the scientific and regulatory interpretations to the various sub-panels. The sub-panels wrote reports that were used to prepare the first and only public draft of the Report.

Comments About the Process
The NTP Report reflects inconsistent sub-panel processes and considerations. Moreover, in some cases the sub-panels draw conflicting conclusions, and in certain cases the published NTP procedures were not followed and the issues were not addressed. In other cases, the investigators' presentations failed to address the critical issue of the dose- response curve.

A major problem with the peer review process became apparent very early in the public meeting, when the EPA definition of low-dose was changed by the NTP. The definition articulated by the NTP in their executive summary does not address the EPA needs. The NTP defined low dose effects as "biological changes that occur in the range of human exposures or at doses that are lower than those typically used in the EPA's standard testing paradigm for evaluating reproductive and developmental toxicity." This definition is flawed on various levels.

First, the NTP definition bears no relationship to the scientific issue that has to be addressed. The NTP considered "biological changes" rather than adverse effects despite the clear instructions from EPA. EPA is legally mandated to require testing and regulation to protect the public from adverse effects. Realistically, there is virtually no adverse effect for which one could not find some biological change with a more sensitive test at a lower level. If this were in the definition of the low dose effects, no panel would be needed because it would be generally accepted as true without argument. Hormones are active and naturally occurring substances that cause physiological and biological changes at low levels. It was not the charge to this peer review to confirm that hormones have such activity. Even if biological changes are observed, EPA will need to regulate on the basis of adverse, rather than just any, effects.

Second, the NTP definition also relates low-dose effects observed in the range human exposure. This part of the definition has absolutely nothing to do with the low dose hypothesis. The low dose hypothesis is that levels of hormonally active agents may cause adverse effects at levels below the NOAEL for that specific target. If the low-dose hypothesis were shown to be generally correct, routine-testing systems might lead to setting a NOAEL at a level above those that actually cause other adverse effects. There are many substances that cause biological changes in the range of human exposure and would meet the NTP definition of low dose. For example any drug, vitamin or hormone supplement would cause low-dose effects because they are causing biological changes at relevant levels of human exposure.

The NTP report justifies their approach with this statement that "in many cases, the long-term health consequences of altered endocrine function during development have not been fully

characterized." Even if this conjecture proves to be true, that is an entirely separate issue from the low-dose hypothesis. EPA has already drawn conclusions about long-term end points following developmental exposure. While that issue can always be revisited, it is separate from the low-dose issue. The low-dose issue is a dose response issue, not a developmental toxicity issue. For the determination of low dose, EPA is concerned with the shape of the dose response curve and the ability of existing toxicity studies to detect changes below the NOAEL for that particular target. Reviewing the individual sub-panel reports confirms that the definition was not applied consistently through the NTP peer-review process.

Another significant issue with the conduct of the NTP peer review was that selected studies were considered, and subsequently given considerable weight, though the data were not submitted for review by this statistical panel. Submission of raw experimental data was an abiding rule for consideration by the low-dose panel. This was clearly inconsistent with the established procedures and in direct conflict with the process established by the NTP/EPA prior to convening the panel.

Major Concerns About the NTP Low Dose Peer Review and the Report

ACPA requests that EPA carefully considers all the data on the hypothesized low-dose effects and ensures discipline in the process. We have serious concerns and comments about the NTP Peer Review process and Report. We believe the NTP peer review represents only a first step in examining this important issue as it possibly relates to regulatory toxicology guidelines and risk assessment. We do not find the Report makes a valid case for changes in either testing or risk assessment based upon the low-dose issue or the hypothetical bi-modal response. The major points are:

- The definition of low dose used by NTP was not consistent with the issue as discussed by the EDSTAC panel, the EPA charge, or the scientific issue as it has been espoused for the past five or more years. This has already been discussed above;
- The use of the definition of low dose was not consistent among the sub-panels;
- The panel did not discuss the biological relevance of the changes that were described by various investigators;
- The panel failed to distinguish between biological and adverse effects;
- The panel failed to consider the importance of the route of exposure (such as subcutaneous injection and subcutaneous Silastic® tubing implants) in comparing the potency of the various dose levels;
- The panels did not discuss the shape of the dose response curve;
- The panel relied on one and two dose studies, even though such a design can provide only extremely limited information about the shape of a dose response curve;
- Some sub-panels made recommendations, that could require extensive changes in testing protocols, without recognizing that significant protocol changes have only recently been implemented that add developmental landmarks and endocrine sensitive endpoints. Calls to change testing paradigms are premature, considering the limited demonstration that "endocrine disruptors" cause adverse effects at relevant exposure levels;
- The research recommendations often described research needs in areas that the panel had not reviewed any data. In certain cases the sub-panels incorrectly assumed that no data existed, when it simply was not asked for, or was outside the scope of this review;

- The sub-panels relied on data that did not meet the original peer review criteria for acceptance. The NTP failed to follow its own criteria, which would have limited the review to the studies that submitted data to the statistics sub-panel or considered studies that otherwise failed to meet the NTP Panel's acceptance criteria;
- The Panel mixed endpoints in comparing the dose response curve at low doses to regulatory NOAELs; and
- Some sub-panels, such as the BPA Sub-panel, failed to clearly describe a weight of the evidence evaluation. Although BPA is not a crop protection chemical, the manner in which the sub-panel described its conclusions could provide an unfortunate precedent for other reproductive and developmental toxicity data. Based on these evaluations, essentially, no amount of solid negative data could ever demonstrate that the initial, limited observation was a false positive. While that may be statistically correct, it does not make sense in evaluating toxicity studies, or studies on any biological system, where conflicting observations will be made and need to be reconciled based on established biological/toxicology principles.

Specific Comments About the Critical Issues

- ***Risk Assessment Considerations***

  The panel failed to consider the potential adversity or biological relevance of the low dose findings. Risk assessment is based upon hazard, exposure and risk, not simple biological response. There needs to be a link between the exposure and an adverse effect. Endpoints in a risk assessment must be based upon reproducible, robust and reliable studies. The route of exposure also can create responses that are not relevant to risk.

  The route of exposure has to be a relevant route. The peer review panel considered single dose, non-replicated and irrelevant routes of exposure such as Silastic® tubing implants and subcutaneous injections. Many substances are altered by the gut flora or are absorbed at vastly different rates by different routes of exposure. Bypassing the skin is not at all relevant for predicting oral exposure to a food additive or similar substance. For example, BPA is barely absorbed from the gastrointestinal tract.

  The BPA sub-panel relied heavily on some of the studies by the subcutaneous route of exposure, which do not predict risk, and cannot be used to compare dose. Nevertheless, they cited them as evidence of low-dose effects. The BPA sub-panel also relied on research studies from Ben-Jonathan's laboratory, and referred to them as credible, even though these studies were not provided to the statistics sub-panel for their independent reanalysis.[14, 15, 16] While the studies from Ben-Jonathan's laboratory may well be credible, the BPA sub-panel misused the data on several levels, and the findings were cited as providing evidence of low-dose effects, even though this work was not part of the low dose peer review.

  A major substantive issue is that the endpoints from these studies included uterine growth and changes in prolactin levels. These are not discussed as adverse effects, and no adverse effects were reported. Just as critical, the dose levels used by such irrelevant

routes of exposure cannot be compared to robust regulatory guideline GLP studies that showed a NOAEL for adverse effects when administered by relevant routes of exposure.

The "Other Environmental Estrogens and Estradiol Sub-panel" reported that these substances (i.e., estradiol and DES) produced dose-response curves that were either low-dose linear, threshold-appearing or non-monotonic. For example, low dose effects for estradiol included changes in serum prolactin, LH and FSH in ovariectomized rats. This is not a model for measuring adverse effects, and the study does not use a hormonally-active animal. Like the Ben-Jonathan studies on BPA, these studies have research objectives that are not consistent with risk assessment or toxicology. The only evidence that seems to fit the low-dose definition is the unreplicated findings on DES in mice.

- ***Weighing the Evidence and Research Recommendations***

  The NTP panel focused too much on single studies, and not, as a whole, on the weight-of- the-evidence. No single study can show that the low-dose hypothesis is a reproducible phenomenon. At this point, there are not two studies for any one endpoint that show it to be reproducible. However, this is not for lack of effort.

  By reviewing data collectively, one can sort out the sources of variability and describe a more complete picture of what is actually happening. The NTP failed to do this. EPA and other regulatory agencies routinely conduct a weight-of-the-evidence assessment because no data set is completely consistent. For example, the BPA sub-panel could have critically considered the key factors in the original Nagel et al. and vom Saal et al.studies and compared them to the six studies from five laboratories and determined the weight for each. The BPA Sub-panel called the initial low-dose research studies credible, because they focused on these as single isolated studies. Each data point by itself appeared real, and was collected and reported in an appropriate manner. That does not mean that each study should get equal weight. There were large differences in design, number of dose levels, number of animals, and data collection methods. These should be considered when the evidence is considered as a whole.

  The sub-panel concluded that they were "not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding," and that the "data are insufficient to establish the shape of the dose response curve for BPA in the low dose region, and the mechanism and biological relevance of reported low-dose effects are unclear."

  However, the panel also recommended research to determine the mechanism of action. If the panel was not convinced that the BPA low-dose effects are reproducible, why do they suggest additional research into the mechanism of action? Truly heroic studies have been conducted that have failed to replicate the initial low-dose BPA findings (See references 6-12). If the BPA low dose effects cannot be consistently evaluated or reliably replicated within and across laboratories, then how can the mechanism be conclusively determined? Or additionally, how can these low-dose studies be validated?

The "Biological Factors and Study Design" groups also focused on areas that were not directly related to the low-dose issue. They made certain recommendations about changes in testing design because of a concern to add certain endpoints to the existing multi-generation rat reproduction study design. Those issues are outside the scope of low dose. They do not consider the dose response curve; instead they dealt with endpoint selection. EPA should consider those issues in an entirely separate context, which is what EDTAC has already accomplished.[3]

Summary Conclusions and Recommendations

- The National Toxicology Program's "Report of the Endocrine Disruptors Low Dose Peer Review" (Report) describes the conclusions of various panels that were supposed to address critical and central issues about EPA's current testing guidelines related to hormonally active agents.
- Although the NTP is to be commended for its rigorous and comprehensive (prior) review of the statistical evaluation of the data that were provided to the low-dose panels, it should also be noted that the some of the panels went outside the intended scope of the low-dose review by considering, in a number of cases, studies that were not intended to be available to the panel. In many cases the statistical review confirmed the conclusions and approaches of the original investigators, in other cases the statistical review panel identified serious flaws or discrepancies.
- The NTP review modified, or nearly completely eliminated, considering the critical issue that EPA needed to have addressed. The revision of the definition of low dose led directly to a number of conclusions and recommendations that were not relevant to EPA's testing and research needs.
- Even with the availability of several large, well-conducted GLP studies using many animals and an exceptionally wide range of dose groups, the panel still concluded that the data on low dose, from very small two-dose research studies were somehow comparable and credible. In some cases, a single dose was used, which certainly cannot yield useful information about a dose-response curve. By themselves, the studies are credible, but in the context of the weight-of-evidence, they should be clearly classified as research observations with insufficient evidence for the existence or confirmation of a low-dose effect.
- The NTP low-dose peer review was not intended to address issues outside of the low-dose question. Therefore, EPA should not accept conclusions or recommendations from the NTP peer review that exceed the original charge and intended scope. The weight-of-evidence from the October 2000 peer review <u>does not</u> support the following recommendation in the Report Executive Summary.

  "The findings of the Panel indicate that the current testing paradigm used for assessments of reproductive toxicity should be revisited to see if changes are needed regarding dose selection, animal model selection, age when animals are evaluated and the endpoints being measured following exposure to the endocrine active agents."

- The Panel Report strayed from the original definition and the original scope of the review. Despite that, the Report supports the conclusion that the potential adverse effects resulting from a bi-modal response that began this review has not been shown to be a reproducible phenomenon.

To summarize, ACPA believes that no further research recommendations or regulatory testing or risk assessment changes should follow based upon the findings of any of the sub-panel reports, as they relate to the low-dose issue. Moreover, observed low-dose effects in laboratory animals do not prove potential disease or adverse health effects for humans. EPA should implement a validated Endocrine Disruption Screening Program, EDSP. The consideration of any recommendation to change classical guideline toxicology or dose-setting paradigms, either from expanding the scope of endocrine endpoints (beyond those already recommended by EDSTAC) or by increasing the number of doses, should be peer-reviewed through the EPA SAB/SAP process. ACPA also supports that EPA should refer the NTP Low Dose Peer Review and the final Report to the Endocrine Disruptor Methods Validation Subcommittee (EDMVC).

References

1. ACPA submitted the following comments to EPA: DC #OPPTS – 42208A dated February 22, 2000; DC# OPPTS – 42212A dated August 25, 2000 and DC # OPPTS – 42212C dated May 31, 2001. See also www.acpa.org

2. Refer to Code of Federal Regulations 40 CFR Part 158, Chapter I, EPA, subchapter E, pesticide programs, part 158, "Data Requirements for Registration."

3. Endocrine Disruption Screening and Testing Advisory Committee (EDSTAC) Final Report: Volume 1, Chapter 5 pgs. 1-80 and Chapter 7 pgs. 14-18. See also www.epa.gov/scipoly/oscpendo/index.htm

4. Nagel, S. C.; vom Saal, F. S.; Thayer, K. A.; Dhar, M. G.; Boechler, M., and Welshons, W. V.. (1997) Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ Health Perspect. 105(1):70-6; ISSN: 0091-6765 and Environ Health Perspect (1997). Jun 105 (6): 570-2.

5. vom Saal, F. S.; Cooke, P. S.; Buchanan, D. L.; Palanza, P.; Thayer, K. A.; Nagel, S. C.; Parmigiani, S., and Welshons, W. V. (1998). A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production, and behavior. Toxicol Ind Health 239-60; ISSN: 0748-2337.

6. Cagen, S. Z.; Waechter, J. M. Jr; Dimond, S. S.; Breslin, W. J.; Butala, J. H.; Jekat, F. W.; Joiner, R. L.; Shiotsuka, R. N.; Veenstra, G. E., and Harris, L. R. (1999). Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A. Toxicol Sci. 50(1): 36-44; ISSN: 1096-6080.

7. Ashby, J.; Tinwell, H., and Haseman, J. Lack of effects for low dose levels of bisphenol (1999). A and diethylstilbestrol on the prostate gland of CF1 mice exposed in utero. Regul Toxicol Pharmacol:156-66; ISSN: 0273-2300.

8. Cagen, S. Z.; Waechter, J. M. Jr; Dimond, S. S.; Breslin, W. J.; Butala, J. H.; Jekat, F. W.; Joiner, R. L.; Shiotsuka, R. N.; Veenstra, G. E., and Harris, L. R. (1999). Normal reproductive organ development in Wistar rats exposed to bisphenol A in the drinking water. Regul Toxicol Pharmacol. (2 Pt 1): 130-9; ISSN: 0273-2300.

9. Tyl, R.W., C.B., Myers, Marr, M.C., Change, T.Y., Seely, J.C., Brine, D.R., Veslica, M.M., Fail, P.A., Joiner, R.L., Butala, J.H., Dimond, S.S., Shiotsuka, R.N., Stropp, G., Veenstra, G.E., Waechter, J.M., and Harris, L.R. (2000). Two-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats. *RTI Study No 65C-07036-000 (Draft Final Report)*. Presented by R.W. Tyle at the International Symposium on Environmental Endocrine Disruptors 2000. Yokahama, Japan, December 18, 2000.

10. Ema, M. and Kanno J. (2000). Two-generation reproduction study of bisphenol A in rats. *(Unpublished Study Report)*. Presented by M. Ema at the International Symposium on Environmental Endocrine Disruptors 2000. Yokahama, Japan, December 18, 2000.

11. Welsch. F., B.A., Elswick, et.al., (2000). Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats. Toxicological Sciences 54 (supplement: 256A.

12. Welsch. F., B.A., Elswick, et.al., (2000). Lack of effects of perinatal exposure to low doses of bisphenol A on male rat offspring ventral prostate glands.

13. vom Saal, F.S.; Timms, B. G.; Montano, M. M.; Palanza, P.; Thayer, K. A.; Nagel, S. C.; Dhar, M. D.; Ganjam, V. K.; Parmigiani, S., and Welshons, W. V. (1997). Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc Natl Acad Sci U S A. 94(5):2056-61; ISSN: 0027-8424.

14. Steinmetz, R.N., Brown, N.G. and Ben-Jonathan, N. et.al., (1997). The Environmental estrogen bisphenol A. stimulates prolactin release in vitr and in vivo. Endocrinology, 138 (5): 1780-1786.

15. Steinmetz, R.N., Mitchner, N. and Ben-Jonathan, N. et.al., (1998). The xenoestrogen bisphenol A induces growth, differentiation and c-fos gene expression in the female reproductive tract. Endocrinology 139 (6): 2741-2747.

16. Long, X., Steinmetz, R.N. and Ben-Jonathan, N. et.al., (2000). Strain differences in vaginal responses to the xenoestrogen bisphenol A. Endocrinology 108 (3): 243-247.

ADDENDUM

AMERICAN CROP PROTECTION ASSOCIATION (ACPA) COMMENTS

NATIONAL TOXICOLOGY PROGRAM'S PEER REVIEW OF LOW-DOSE ISSUES FOR ENDOCRINE DISRUPTORS

Reissued on July 16, 2001

Presented by:
Angelina Duggan, Ph.D.
Director of Science Policy
American Crop Protection Association
October 10, 2000

Background and Introduction

The American Crop Protection Association (ACPA) commends EPA, NIEHS and NTP for organizing the Peer Review of Low Dose Issues for Endocrine Disruptors and the opportunity to provide public comments to the Low-Dose Panel.

ACPA is a not for-profit trade organization of U.S. manufacturers, formulators and distributors of crop protection and pest control products. ACPA members produce, sell and distribute virtually all of the active ingredients used in such products. We promote the environmentally sound use of crop protection products for the economical production of safe, high quality, abundant food, fiber and other crops.

Our member companies are dedicated to serving the public through science in agriculture while protecting the environment and human health. In the spirit of this commitment, ACPA has endorsed the implementation of a validated endocrine disruptor screening and testing program and a priority setting process based on sound science and weight-of- evidence in decision making.

Existing statutory mandates, the Federal Insecticide Fungicide and Rodenticide Act (FIFRA), require agrochemicals to undergo at least 120 or more separate evaluations before entering the marketplace. These tests include acute, developmental, multi-generation reproduction, sub-chronic and chronic studies to ensure both safety and effectiveness. These toxicology tests have a long-history of use and are similar to those required by the FDA for new pharmaceuticals. Although the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) has recommended additional endpoints, EPA and toxicology experts have deemed the current safety evaluations appropriate to address at least 25 estrogen/anti-estrogen, 13 androgen/anti-androgen and 6 thyroid relevant endpoints some of which include endocrine mediated cancers.

ACPA has actively participated in the evaluation of endocrine issues. Our industry representatives served on the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) and contributed technical expertise and policy perspective to the EDSTAC Working Groups for Priority Setting, Screening and Testing, and Communications and Outreach. We also contributed to writing the EDSTAC Report and have continued to work with other stakeholders and EPA to implement the recommendations from EDSTAC by participating in the EPA Priority Setting Workshops and on EPA's Endocrine Standardization & Validation Task Force (SVTF) Steering Committee.

EDSTAC acknowledged that the low-dose hypothesis is a complex issue that requires critical analysis. In reaching consensus, EDSTAC recommended that more research was necessary in order to properly address the low-dose issue. Overall, this view was corroborated by the recommendations of the National Academy of Science (NAS) Report, "Hormonally Active Agents in the Environment" released August 1999. The Report called for additional research, including investigation of the low-dose issue, in order to reduce scientific uncertainty about effects of hormonally active agents in the environment.

The science concerning the low-dose hypothesis has continued to advance since EDSTAC and the NAS report. However, there remains and most likely will continue to be divergent scientific opinions about the relevance and interpretation of low-dose effects.

Comments

ACPA welcomes a rigorous evaluation of all the data and a balanced representation of all perspectives in carrying out the objectives of this peer review. Setting aside classical toxicology principles and embracing new paradigms for endocrine mediated thresholds and dose responses are serious ventures. We strongly advocate that the low-dose hypothesis be implemented only if the weight-of-the-evidence supports the conclusion that the occurrence and frequency of low-dose effects can be well characterized. Moreover, low-dose observations should also correlate with definable adverse effects in the animal models, including the thorough understanding of species and strain variability. In our opinion, implementation of a new low-dose paradigm is not appropriate unless:

(1) the body of scientific knowledge supports that adverse effects are clearly demonstrated at low-doses and
(2) the relevance and implications of these effects in humans are clear.

Assessing Adverse Effects and Relevant Doses for Low-Dose Effects

In assessing adverse effects, measurements should focus on toxicology endpoints with clear hormonal significance . Evidence of an unusual dose response with endpoints that have no clear toxicological significance for endocrine mediated events , or endpoints that cannot be included in routine toxicology testing protocols, should not influence the dose selection process for routine toxicological studies.

The NTP has charged the Panel to consider "low-dose effects" as "biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in EPA's standard toxicity testing paradigm." In our opinion, this charge is an enigma since a general criteria for "environmentally relevant doses" has not been defined. Overall, it is more likely that an environmentally relevant dose will fluctuate on a case-by-case basis due to temporal considerations and site (geographical) and/or compound specificities, Thus, arbitrarily testing at doses below a previously observed NOAEL may or may not correlate to a relevant environmental exposure or dose.

Moreover, the correlation between adverse endocrine effects and the low dose hypothesis is weak at best. For example, the link for chemical exposures and adverse endocrine effects in wildlife has had validity only at very high environmental doses - accidental spills or severe environmental pollution. Additionally, the environmental cause and effect for humans have not been conclusively established. EPA, NAS and foreign governments have conducted independent assessments of the endocrine disruption issue, and the follow-on reports have called for

additional research. To date there has been no evidence to substantiate that the human endocrine system is being disrupted, adverse or otherwise, by low-level environmental exposures to hormonally "active" substances.

Weight -of-the- Evidence

In considering weight-of-the-evidence, the Low-Dose Panel should evaluate the source of the evidence, the weight and quality of the evidence and the consistency of the evidence. Weight-of-the-evidence needs to consider whether the experimental design (protocol) is capable of measuring adverse effects over the range of doses and whether the study has sufficient statistical power to measure these effects. Moreover, the uncertainty factors and confidence limits of low dose effects should be clearly defined and available for examination along with the raw data.

It is also important to note whether the study data has been audited and conducted according to Good Laboratory Practices (GLP). The Panel should eliminate from contention those studies for which the raw data are not available. Moreover, low-dose results should be consistent and easily reproducible within laboratories and by other researchers – this will be absolutely necessary in order to validate the experimental design of a particular low-dose study. Finally,confounding factors, that may lead to inconsistencies or erroneous results between laboratories, should also be reported.

These factors could include:

- Sample size
- Differences in experimental controls
- Dietary phytoestrogens
- Caging of multiple animals to induce potential hormonal stress and dominant males
- Position of the selected fetus in the uterus
- Laboratory animal strain and species differences
- Temporal measurements
- Sample purity

Shape of the Dose-Response Curve for Endocrine Active Substances In The Low Dose Region

NTP has also requested the Panel to address the definition and criteria of the shape of the dose response curves for endocrine active substances in the low-dose region. The hypothesis at issue is whether a non-monotonic or "U-shaped" curve may result from a very low exposure to a hormonally active chemical substance. In doing so, the Panel should not consider any study that provides only two data points (for the low dose region). It is not possible to demonstrate with accuracy a "U-shaped" curve with so few data points. The Panel should also consider the comparative mode and/or mechanism of action for low versus high doses. This is necessary in order to establish whether the mode/mechanism of action are consistent, one and the same, or different over the range of doses.

Additionally, during the last several years, receptor-mediated endocrine research has indicated that binding processes and resultant biological effects may be more complex than originally thought. For example, organ-specific responses have been confirmed for two different estrogen receptors, alpha and/or beta-estrogen receptors, both of which may or may not be in present in specific organs and in varying concentrations. Moreover, the strength of binding to either the alpha-or beta-receptor may vary on a compound-by compound-basis. For example, many of the naturally occurring phytoestrogens, while binding very strongly to beta-estrogen receptors, display only weak agonist activity at the alpha-estrogen receptor.

Conclusions

As charged, the primary objective of this peer review is not to consider whether there may be low-dose effects, but rather "to evaluate the likelihood and significance of these and/or other potential low-dose effects for humans." With this consideration firmly in mind, we urge the Panel to determine the implications, if any, "for the development, validation and interpretation of test protocols for reproductive and developmental toxicity."

In our opinion, at this time the complete body of scientific evidence does not support a new paradigm for endocrine mediated thresholds or regulation based on potential low dose effects. Moreover, there is no clear evidence to link adverse low dose effects to humans.

Additional doses or arbitrarily repeating reproductive and developmental toxicology testing on the basis of a low-dose hypothesis are not warranted and would result in a needless expenditure of laboratory resources, including animals. While some small-scale studies have provided support for a low-dose phenomenon, more recent independent studies have not seen the translation of these unusual low-dose effects to large-scale whole animal studies.

ACPA expresses their appreciation to EPA, NIEHS and the NTP for conducting a peer review of low-dose issues for endocrine disruptors. We also thank the Panel and wish them well in evaluating the large body of low-dose information.

American Plastics Council

VIA FAX URGENT

DISTRIBUTE TO:

NTP Office of Liaison and Scientific Review Fax # 919-541-0295

From: Steve Hentges Phone #: 703-253-0636
Number of Pages 24 including cover)

Comments:

Original will be sent overnight for delivery tomorrow.

Please call the number above if this transmission is incomplete.

Nancy Kirks
American Plastics Council
1300 Wilson Boulevard
Arlington, VA 22209
703-253-0635
Fax – 703-253-0701
Nancy_Kirks@plastics.org

American
Plastics
Council®

1300 Wilson Boulevard
Suite 800
Arlington, VA 22209
Tel 703.253.0700 Fax 703.253.0701
www.plastics.org

July 16, 2001

NTP Office of Liaison and Scientific Review
111 T. W. Alexander Drive
P. O. Box 12233
MD A3-01
Research Triangle Park, NC 27709

Re: Comments on the National Toxicology Program Endocrine Disruptors Low-Dose Peer Review Final Report

Dear Sir or Madam:

Attached are comments from the Polycarbonate Business Unit of the American Plastics Council on the final report of the NTP Endocrine Disruptors Low-Dose Peer Review.

If you have any questions or need additional information, please feel free to contact me by phone on (703) 253-0636 or by e-mail at steve_hentges@plastics.org.

Sincerely,

Steven G. Hentges, Ph.D.
Executive Director
Polycarbonate Business Unit

COMMENTS ON THE NATIONAL TOXICOLOGY PROGRAM ENDOCRINE DISRUPTORS LOW-DOSE PEER REVIEW FINAL REPORT

Submitted by
The Polycarbonate Business Unit of
The American Plastics Council

July 16, 2001

Comments by the American Plastics Council on the National Toxicology Program Endocrine Disruptors Low-Dose Peer Review Final Report

1. Summary of Comments

For many years, the manufacturers of Bisphenol A (BPA) have supported and conducted an extensive amount of research on the safety of BPA. These studies have consistently demonstrated that BPA and products derived from BPA are safe for their intended uses and pose no health issues for consumers. Included in this large body of research are the most comprehensive studies conducted to date to assess the potential for BPA to cause low-dose health effects. Accordingly, these comments are focused primarily on the conclusions of the Bisphenol A Subpanel as supported by the Statistics Subpanel.[1]

A summary of key comments is provided below in this section, with supporting details provided in the following sections.

- **The Bisphenol A Subpanel did not complete a weight-of-evidence assessment, which would have concluded that low-dose effects from BPA have not been demonstrated.**

  Instructions from the Organizing Committee of the Low-Dose Peer Review indicate that the Subpanels were to use a weight-of-evidence approach in their assessment of information relevant to the topic assigned to each Subpanel.

[1] Chapter 1: Report of the Bisphenol A Subpanel and Chapter 5: Report of the Statistics Subpanel; National Toxicology Program's Report of the Endocrine Disruptors Low Dose Peer Review.

Regarding reported low-dose effects of BPA, the BPA Subpanel concluded that these effects have:

- not been conclusively established as a general or reproducible finding;
- an uncertain mechanism(s) (e.g. are not known to be caused by an endocrine mediated mechanism); and
- unclear biological relevance (e.g. there is no evidence that the reported effects are adverse or have relevance to human health).

Nevertheless, in spite of the overwhelming weight of evidence demonstrating the lack of low-dose effects, the BPA Subpanel did not complete a weight-of-evidence assessment. The weight of evidence is fully adequate to conclude that low-dose effects from BPA have not been demonstrated and that reported low-dose effects of BPA are not credible. The BPA Subpanel should have reached this conclusion.

- **The BPA Subpanel improperly used or ignored the ground rules established by the Organizing Committee for admission of scientific evidence to the peer review process.**

  Studies for which data was requested but not submitted by principal investigators were to be used only as background information by the Subpanels. However, a number of such studies were fully considered by the BPA Subpanel and apparently contributed to their overall conclusion and recommendations. Without the requested data and information, it is not possible to meaningfully assess and

weigh studies that have not been fully reported and, in fact, may lead to erroneous interpretations and conclusions. The BPA Subpanel's reliance on such non-validated studies was contrary to the ground rules established by the Organizing Committee and a contributing factor in their failure to complete a weight-of-evidence assessment.

- **The BPA Subpanel did not fully and completely examine all of the submitted background published references.**

  Significant research on the pharmacokinetics and metabolism of BPA was not fully examined by the BPA Subpanel. This research, along with additional research completed subsequent to the BPA Subpanel's deliberations, supports the overwhelming weight of evidence for the lack of low-dose effects of BPA.

- **The BPA Subpanel's call for mechanistic and additional low-dose studies is inappropriate since the weight of evidence does not support a low-dose effect for BPA.**

  Given the overwhelming weight of evidence that BPA does not cause low-dose effects, it is unclear how the proposed mechanistic or molecular research on BPA is warranted or could contribute to further resolution of the reported low-dose effect that evidently does not exist.

2. Introduction to Detailed Comments

The American Plastics Council (APC) is a major trade association for the plastics industry in the United States. The membership of APC includes 23 of the leading plastics resin manufacturers plus one affiliated trade association, which collectively represent more than 80 percent of the monomer and plastic resin production capacity in the United States. Included in APC's membership are manufacturers of bisphenol A (BPA), epoxy resins and polycarbonate plastic resin.[2] These comments are submitted on behalf of the members of the Polycarbonate Business Unit (PCBU) of APC.

For many years, the manufacturers of BPA, polycarbonate plastic and epoxy resins have supported and conducted extensive research to evaluate the safety of BPA and products made from BPA. These studies have consistently demonstrated that BPA and derived products are safe for their intended uses and pose no health issues for consumers. Included in this large body of research are the most comprehensive studies conducted to date to assess the potential for BPA to cause health effects at low doses. Accordingly, these comments from PCBU are focused primarily on the conclusions of the Bisphenol A Subpanel as supported by the Statistics Subpanel.[3]

---

[2] Bisphenol A is used primarily as an ingredient to make polycarbonate plastic and epoxy resins. These products have been widely and safely used for many years and are highly valued by consumers for their beneficial uses and properties. Polycarbonate is an ideal material for use in a wide range of applications such as eyeglass lenses, medical equipment, five-gallon water bottles, baby bottles, kitchen ware (e.g. microwaveable food storage containers, drinking glasses, dinner plates), digital media (e.g. CDs and DVDs), and automotive uses that take advantage of its light weight, shatter-resistance, clarity and durability. Epoxy resins are used a variety of protective coating including on metal cans to maintain the quality and integrity of foods and beverages. These coatings provide essential public health benefits by helping to protect food from spoilage (e.g. botulism) and contamination from the metal containers. Epoxy resins are also used in a wide variety of engineering applications, such as the manufacture of electrical laminates for printed circuit boards.

[3] Chapter 1: Report of the Bisphenol A Subpanel and Chapter 5: Report of the Statistics Subpanel; National Toxicology Program's Report of the Endocrine Disruptors Low Dose Peer Review.

3. **Detailed Comments on the Report of the Bisphenol A Subpanel**

   a. **The Bisphenol A Subpanel did not complete a weight-of-evidence assessment, which would have concluded that low-dose effects from BPA have not been demonstrated.**

Each Subpanel of the Low-Dose Peer Review was charged to answer seven questions after examination of the report of the Statistics Subpanel and reading the background references and "selected studies" pertinent to the topic assigned to the Subpanel. The first two questions to be addressed by the BPA Subpanel were to define the "extent of empirical evidence demonstrating low-dose effects" and "extent of empirical evidence demonstrating the lack of low-dose effects" of BPA. The term "extent of empirical evidence" clearly implies that it was the intention of the Organizing Committee of the Low-Dose Peer Review to use a weight-of-evidence approach. Both of the first two questions also indicate that the evidence should be examined "within and across studies", which further indicates that the BPA Subpanel was expected to consider the evidence comprehensively rather than as a set of isolated studies.

To complete their assigned task to weigh the evidence, the BPA Subpanel had to overcome a number of difficult challenges, including:

- no clear definition of what dose should be considered as a "low dose";
- a large number of "selected studies" to review in a limited period of time;
- raw data that was not provided for a number of the "selected studies"; and

- relevant studies (e.g. metabolism and pharmacokinetics) were not considered because of time constraints.

Despite these challenges, the BPA Subpanel made substantial progress in weighing the evidence supporting and refuting low-dose effects of BPA to reach an overall conclusion. Most notably, regarding reported low-dose effects of BPA, the BPA Subpanel concluded[4] that these effects have:

- not been conclusively established as a general or reproducible finding;
- an uncertain mechanism(s) (e.g. are not known to be caused by an endocrine-mediated mechanism); and
- unclear biological relevance (e.g. there is no evidence that the reported effects are adverse or have relevance to human health).

Nevertheless, despite the overwhelming weight of evidence demonstrating the lack of low-dose effects, the BPA Subpanel did not complete a weight-of-evidence assessment, which resulted in the inconclusive statement that there is "credible evidence" both for and against low-dose effects. By not fully weighing the evidence, this statement gives the impression that the single, small preliminary study with mice showing an effect on reproductive organ weights, for which data was not provided to the Statistical Subpanel, has equal standing with the many larger, carefully designed and conducted studies that

[4] The overall conclusion of the BPA Subpanel is presented on page 1-11 of the final report:
"There is credible evidence that low doses of BPA can cause effects on specific endpoints. However, due to the inability of other credible studies in several different laboratories to observe low dose effects of BPA, and the consistency of these negative studies, the Subpanel is not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding. In addition, for those studies in which low dose effects have been observed, the mechanism(s) is uncertain (i.e. hormone related or otherwise) and the biological relevance is unclear."

examined multiple doses, some of which examined many animals over multiple generations and for all sensitive life stages, and all of which found no evidence for low-dose effects. The data from the remaining low-dose studies were either not provided to the Statistics Subpanel for analysis or the reported effects were not confirmed upon re-evaluation by the Statistics Subpanel and therefore should not have been considered further.

Specifically,

1. The BPA Subpanel chose to state that the evidence of increased prostate weight from seven mice per group in a single laboratory was found to be "credible", even though two major studies with nearly identical treatment conditions and procedures, which had substantially greater numbers of animals per group (up to 28), more dose groups and evaluated more end-points, were unable to replicate these findings (Cagen *et al.* 1999; Ashby *et al.* 1999). Interestingly, the Statistics Subpanel could not validate any of the other low-dose effects reported in these same mice in a separate paper where the data were submitted for evaluation (vom Saal *et al.* 1998).
2. The BPA Subpanel also chose to include as evidence for low-dose effects a preliminary study in CD1 mice, which was presented only as an oral presentation at the meeting, although the offspring in this study were taken by Caesarian section prior to delivery and examined for a completely different endpoint (3-D reconstruction by serial sectioning, a technique that has not been subjected to verification of accuracy or relevance). No data from this study were submitted

(or have been submitted subsequently) to the Statistics Subpanel for independent evaluation and, therefore, under the procedures established for the review, the study should not have been considered further.

3. The BPA Subpanel considered a series of studies by Ben-Jonathan *et al.* (pp. 1-5 to 1-6) to be "very credible and consistent", although the data were not provided to the Statistics Subpanel for analysis and independent confirmation, nor were the doses used within the definition of low dose established by the BPA Subpanel. The BPA Subpanel stressed that "these studies did NOT find any low dose effects in SD animals, and the "low dose" effects of BPA seen in F344 animals were observed at what would be considered the "high edge" of the low dose range."[5] Furthermore, BPA administration in all of these studies was via sc [subcutaneous] injection or release from silastic implants, and both these routes of administration would be expected to have far higher levels of bioavailability than oral administration of BPA."

   The reported effects of Ben-Jonathan *et al.* do not represent true low-dose effects for BPA, nor did they receive appropriate peer review of the data, yet the BPA Subpanel considered them as "credible" evidence of low-dose effects. It is even

[5] The lowest doses reported to cause effects, 0.3 - 0.5 mg/kg/day and higher, are within about a factor of ten of the estimated NOAEL for oral exposures to BPA, 5 mg/kg/day (p. 1-3). The NTP Organizing Committee defined low dose as doses "lower than those typically used in EPA's standard toxicity testing paradigm," and thus interpreted doses at or below the NOAEL as low doses. However, the bioavailability of BPA from subcutaneous injections (implants would be expected to be similar to injections since both are subcutaneous exposures) is ten-fold or more higher than from oral exposures (L.H. Pottenger *et al.*, 2000, "The relative bioavailability and metabolism of bisphenol A in rats is dependent upon the route of administration," *Toxicological Sciences* **54**: 3 – 18). Consequently, the doses reported to cause effects by Ben-Jonathan *et al.* may be at or above the NOAEL expected from subcutaneous exposures. Doses at the NOAEL would just meet NTP's definition of low dose, while doses above the NOAEL would not. These doses are in fact well above EPA's safe or reference dose (RfD) for long-term oral exposure to BPA of 0.05 mg/kg/day, and thus, were not originally considered low dose by the BPA Subpanel (p. 1-3).

more surprising that these papers were considered significant at all since the BPA Subpanel reported the findings to be "counterintuitive" to known endocrine signaling mechanisms.

4. Included in the weight of evidence against these findings, in addition to the studies of Cagen and Ashby mentioned above, were two major multi-generation reproduction studies in rats (Ema 2000; Tyl *et al.* 2000), both demonstrating a lack of low-dose effects of BPA on prostate weights (as well as many other parameters) across doses and generations. Indeed, the study of Tyl *et al.* is one of the most comprehensive animal studies of reproduction ever conducted and was described by the Statistics Subpanel as "arguably the most comprehensive of the studies we evaluated.... The statistical methods were well thought out and appropriate for the data".

Both the Ema and Tyl *et al.* studies were deliberately designed to detect reproductive or offspring effects, including endocrine-mediated effects, from low doses of BPA. The Tyl *et al.* study explored key indicators of possible male and female toxicity at all stages of development through three generations in 30 females and males per group. The study design included seven dose levels ranging from 1 μg/kg/day to 500 μg/kg/day and analyzed more than a million individual data points. Conducted under Good Laboratory Practices, the study met or exceeded current international standards for reproductive health testing. The results of the Ema and Tyl *et al.* studies were fully consistent and neither study found evidence for low-dose effects from BPA.

5. Additional evidence against a low-dose effect for BPA is provided by the extensive studies in rats conducted at CIIT (Elswick *et al.* 2000. "Effect of different sampling designs on outcome of endocrine disruptor studies." *Reproductive Toxicology* **14**: 359-367; Elswick *et al.* 2000. "Effects of perinatal exposure to low doses of bisphenol A in male offspring of Sprague-Dawley rats." *Toxicological Sciences* **54(Supplement)**: 256A), which demonstrated no effects on prostate weight, as well as no effects on many other parameters examined, following oral administration of low doses of BPA in drinking water. These conclusions are further supported by new information from another study from the CIIT group by Kwon *et al.* 2000 ("Pubertal development and reproductive functions of Crl:CD BR Sprague-Dawley rats exposed to bisphenol A during prenatal and postnatal development", *Toxicological Sciences* **55**: 399-406), which showed that a gavage study of BPA given from GD 11-PND 20 at doses of 3.2, 32, and 320 mg/kg/day to rats did not produce any changes in testis, epididymis, seminal vesicle, and ventral and dorsolateral prostates of male offspring.
6. The BPA Subpanel, in reviewing the evidence against low-dose effects, did state that there were many large and well-conducted studies that found no low-dose effects. "Collectively, these studies found no evidence for a low dose effect of BPA, despite the considerable strength and statistical power they represent, which the Subpanel considered especially noteworthy."

The overwhelming "credible evidence" against low-dose effects from BPA is fully adequate to support a weight-of-evidence conclusion that low-dose effects from BPA have not been demonstrated. The weight of evidence is also adequate to conclude that the reported results of one study with only seven mice per group are not credible, particularly as the results of the study could not be replicated and the Statistics Subpanel could not validate any of the other low-dose effects reported in the same mice in a separate paper. The BPA Subpanel should have reached this same conclusion.

At odds with the overwhelming weight of evidence against the validity of low-dose effects, the BPA Subpanel discussed a number of factors that they believed may have differed between the studies reporting low-dose effects and those reporting no such effects (pp. 1-7 to 1-9). Interestingly, because low-dose effects have not been demonstrated to be a general or reproducible finding, no conclusion can logically be drawn regarding the factors that were different between the studies. Consequently, the discussion is highly speculative and does not in any way offer possible explanations for the different results observed.

Further, much of this discussion is unsupported. For example, the report mentions a difference in phytoestrogen levels between Purina 5001 and 5002 diets (p. 1-8) as a possible cause for the disparate results between research groups. The manufacturer of Purina Rodent Chows 5001 and 5002 reports similar amounts of total phytoestrogens, mainly genistein and diadzein (*Personal Communication*, Dr. Torrance Haught, Purina Mills, Inc., St. Louis, MO, January 8, 2001), ranging from 1.1 to 1.4 times as much in

5001 as 5002 (total combined content, measured as aglycone units, averaged 356 for 5001 and 259 for 5002). These small differences and variation between feed lots likely represent batch-to-batch variation. More recent research by Ashby and colleagues reported that the similar marked differences in body weight, sexual development, and reproductive tissue weights seen for the phytoestrogen-free diet AIN-76A and the phytoestrogen-rich diet Purina 5001 indicated that these effects were probably caused by nutritional differences between the diets having centrally-mediated effects on rodent sexual development rather than individual dietary components affecting peripheral estrogen receptors (Odum *et al.*, 2001, "Effect of Rodent Diets on the Sexual Development of the Rat", *Toxicological Sciences* **61**: 115-127). Consequently, these data show that the differences in phytoestrogen content of the diet alone do not explain the differences in observed results among laboratories with BPA.

Another example is the discussion of whether higher prostate weights in control animals in some studies explain the observation of no low-dose effects (p. 1-8). A further review of the data reveals that prostate weights higher than those in the BPA-treated animals were observed in an earlier study from the same research group claiming low-dose effects of other compounds, natural estrogens.[6] Consequently, prostate weight appears to vary considerably in control animals and, hence, this normal variation alone may explain the differences in observed results. In fact, this reported variability raises further questions as to whether the "low-dose effects of BPA" originally reported could have been observed in a repeat study by vom Saal and colleagues (Nagel *et al.* 1997), given the small numbers of animals used in the study and the inherent variability shown in their

[6] Reviewed in J. Ashby, "Towards resolution of the divergent effects of estrogens on the prostate gland of CF-1 mice," *Environmental Health Perspectives* **109**: A109 (2001).

own colony of CF-1 mice (see footnote 6). Unfortunately, these investigators have reportedly destroyed the colony of laboratory animals and chose to not repeat their own study.

One factor previously known in the published literature to have impact on male reproductive organ weights (including the prostate gland) is the condition of single *vs.* group housing, a factor that was controlled through singly housing mice in the study of Cagen *et al.* (1999). Electing to group house mice in the study of Nagel *et al.* automatically confounded the results of this study, since not controlling this factor introduced a variable that was known to have impact on prostate organ weights. Although Nagel *et al.* claim that singly housing mice for a short time before sacrifice "reverses" any impact on prostate weights, no data have been published that supports this hypothesis, nor were data supporting this point presented to the Statistics Subpanel for independent evaluation and corroboration.

b. The BPA Subpanel improperly used or ignored the ground rules established by the Organizing Committee for admission of scientific evidence to the peer review process

According to the Executive Summary of the NTP report (p. i), "studies for which requested data were not submitted by principal investigators for independent review by the Statistics Subpanel were used as background information by the Panel." However, in contrast to this statement, a number of studies without the requested data were fully considered by the BPA Subpanel and apparently weighed heavily in their conclusions and

recommendations. Notable examples include the Ben-Jonathan *et al.* studies and two studies by Howdeshell *et al.*[7], discussed on pages 1-4 to 1-5 of the NTP report.

This is not simply a matter of following an established protocol for the peer review. The quality of experimental data, including study design, statistical analysis of raw data and interpretation of results is of critical importance in determining the credibility and relative weight of competing studies in a weight-of-evidence assessment. As noted in the Executive Summary of the NTP report (page ii), the analyses conducted by the Statistics Subpanel "provide greater insight on the experimental data than is typically apparent in most peer-reviewed research articles, consequently, the statisticians' report was critical for each of the subpanel reviews." Lacking the requested data and information, it is not possible to meaningfully assess and weigh studies that have not been fully reported, and in fact, may lead to erroneous extrapolations and conclusions. The BPA Subpanel's reliance on such non-validated studies was a contributing factor in their failure to complete a weight-of-evidence assessment.

c. The BPA Subpanel did not fully and completely examine all of the submitted background published references

The BPA Subpanel notes in several places that parenteral administration of BPA would be "expected" to result in higher bioavailability of BPA, and hence, studies using subcutaneous or other parenteral routes of administration must be interpreted with caution. Nevertheless, the BPA Subpanel chose to disregard "the route or duration of

[7] Howdeshell *et al.* 1999 and Howdeshell *et al.* 2000 represent two of the five studies from vom Saal for which data were requested but not provided.

administration or the age of the animal used for a particular protocol." In fact, the route-dependency in the pharmacokinetics and metabolism of BPA has been fully demonstrated and published (Pottenger *et al.* 2000). Also, the BPA Subpanel appears to have ignored a paper providing evidence that the fetal mouse ovary does not produce estrogen (Greco and Payne 1994), which challenges the hypothesis that additional xenoestrogen exposure added to background estrogen levels *in utero* is a key factor in producing low-dose effects. If other fetal tissues are producing the estrogen, that fact has not been shown experimentally or, at least, data were not presented to the Statistics Subpanel for independent evaluation and corroboration.

The BPA Subpanel noted (p. 1-11) that the available time precluded their undertaking of a rigorous consideration of pharmacokinetic issues. Since it appears that the Subpanel did not review the data submitted, it therefore called for more research on BPA pharmacokinetics, including comparison of mice and rats. In fact, an extensive study of the pharmacokinetics of BPA has been published[8] and was provided to the NTP peer review. This study demonstrated that BPA was rapidly eliminated from the animal following oral exposure. These findings support the overwhelming weight of evidence for the lack of low-dose effects of BPA.

A study comparing liver cells from mice, rats, and humans was in progress at the time of the NTP review and has since been completed. The results show similar patterns of metabolism of BPA, which support the safety of products made from BPA by

[8] L.H. Pottenger *et al.*, 2000, "The relative bioavailability and metabolism of bisphenol A in rats is dependent upon the route of administration," *Toxicological Sciences* 54: 3 -18.

demonstrating that the capacity of the human liver to metabolize BPA greatly exceeds any likely human exposure.[9] Another study recently completed and being prepared for journal submission shows that the fetus has the enzymes and microsomal mechanism for metabolizing BPA to its glucuronide conjugate. This ability increases after parturition and becomes maximal in young adulthood.[10]

d. The BPA Subpanel's call for mechanistic and additional low-dose studies is inappropriate since the weight of evidence does not support a low-dose effect for BPA

The BPA Subpanel report indicates that low-dose effects must be established as a "general, reproducible phenomenon" before follow-up research on modeling and mechanisms can be done, or biological significance assigned. Yet the Subpanel went on to propose much mechanistic-type research. Given the overwhelming weight of evidence that BPA does not cause low-dose effects, it is unclear how the additional mechanistic or molecular research on BPA proposed by the BPA Subpanel are warranted or could contribute to further resolution of a reported low-dose effect that evidently does not exist. In other words, most of the research proposed appears to be oriented towards searching for the mechanism of an effect that cannot be clearly or reproducibly demonstrated.

---

[9] The rate of metabolism by 3 hours was similar in all species tested and all BPA was converted to conjugates by 3 hours. Initial rates of metabolism in hepatocytes followed the order of mice > rats > humans. However, when extrapolated to the whole liver, the differences in liver mass reverse this relationship to humans > rats > mice. These data show that the *in vitro* metabolic profile of BPA is similar in hepatocytes from rats, mice, and humans. Pritchett *et al.*, 2001, "Differential Metabolism of Bisphenol A in Primary Cultured Hepatocytes from Mice, Rats and Humans," *The Toxicologist* **60(1)**: 96.

[10] Although the microsomal glucuronidation rate is about 50% lower in fetal tissues than in 77-day old adults (38 vs. 72 nmol/min/mg), the increased liver weight as percentage of body weight, 9% (fetal) vs. 4% (adult), may allow for more extensive metabolism. These data show that the glucuronsyltransferase responsible for BPA metabolism develops in the fetus and begins increasing soon after birth. Kuester *et al.*, 2001, "Glucuronidation of Bisphenol A in Hepatic Microsomes: Age-Dependent Differences", *The Toxicologist* **60(1)**: 95.

Thus, the studies proposed to determine "estrogen receptor occupancy during critical periods of development, effects of specific receptor antagonists, and responses in estrogen-receptor knock-out mice"; "genetic and epigenetic factors that affect responses to bisphenol A", or "regulation of transcriptional activity, from gestation through adulthood" are unlikely to provide any valuable evidence for or against the low-dose hypothesis. None of these studies involves rigorously tested and validated procedures because many of those procedures are still in the early research stages in academic institutions.

The Executive Summary appears to misinterpret the research proposed by the BPA Subpanel and "assumes" that the low-dose effects are real (page iv, third bullet). From this assumption, proposed additional research includes: "additional low-dose studies, including the development and use of sensitive and easily measured molecular endpoints, following *in utero* or early neonatal exposure to **conclusively establish low-dose effects of bisphenol A as a general, reproducible phenomenon"** [emphasis added]. The BPA Subpanel actually stated "...If experimental paradigms can be developed to conclusively establish low dose effects of the chemical as a general, reproducible phenomenon, these should be used to obtain sufficient data points to perform credible physiologically based pharmacokinetic modeling." The BPA Subpanel did not "assume" the effects were real and just needed "establishing", but recommended research to more clearly determine the validity of the low-dose theory for BPA. Even this seems to be an unwarranted recommendation and wasteful of both animals and research effort since the

overwhelming weight of evidence clearly shows no effects from exposure to BPA below the currently accepted NOAEL of 5 mg/kg/day.

The BPA Subpanel also proposed additional pharmacokinetic studies. In fact, the BPA Global Industry Group has already sponsored the generation of *in vitro* metabolism data in the species and strains suggested, although these data are not yet published nor were reported and available for the low-dose peer review (see footnotes 9 and 10). In addition, many other data on pharmacokinetics and metabolism have recently been generated in rodents and primates, including both monkeys and humans (both *in vitro* and *in vivo*), from several different investigators.

If any additional research were to be conducted, three of the suggestions proposed by the BPA Subpanel may provide additional insight and information for assessing the potential hazard and risk of exposure to BPA. In each case though, the proposed research is highly speculative since the overwhelming weight of evidence does not support the conclusion of a potential hazard or risk from exposure to low doses of BPA to any life stage.

First is the suggestion to conduct additional, carefully controlled, double-blind (with some laboratories involved receiving only placebos for use as the treatment), and "refereed" low-dose studies with standardized oral administration. However, the weight of evidence against low-dose effects by this route is already substantial.

Second, the suggestion to pursue studies on intrauterine position may also assist in determining if the theoretical basis for xenoestrogen exposure adding to the background "estrogen load" is defensible. However, it should be noted that the BPA Subpanel found the positional effects reported by Howdeshell *et al.* to be "counterintuitive" and not consistently observed (page 1-5).

Third, additional carefully controlled and statistically sound repeat-dose pharmacokinetic studies could resolve the potential for BPA to "bioaccumulate". However, pharmacokinetic data currently available does not support the hypothesis of bioaccumulation of BPA in tissues.

In summary, the BPA Subpanel report concludes with recommendations for further research, beginning with research to establish, if possible, low-dose effects of BPA (pp. 1-12 to 1-13). The report suggests the need for additional markers and endpoints to investigate low-dose effects, although a very wide range of endpoints was examined in the six comprehensive studies that found no low-dose effects. In fact, the report provides little guidance as to which additional markers and endpoints to investigate for possible low-dose effects of BPA and instead, suggests "basic" research not designed to elucidate potential hazard or risk.

4. Errors and corrections

a. ***Page 1-3, point 7, paragraph 1***

"Nevertheless, since the incidence of testis tumors was significantly elevated in the low dose (1000 ppm) male rat group, the EPA apparently used these data to support their conclusion that this dose, which corresponds to an oral dose of 50 mg/kg/day based on typical food consumption rates, represents a LOAEL for BPA." This statement incorrectly attributes the LOAEL to findings of carcinogenesis in the NTP chronic bioassay. The EPA IRIS document used compound-related effects on body weight as the critical effect to establish the LOAEL. Furthermore, carcinogenesis is not an endpoint that is considered in determination of an RfC. In addition, the incidences of testes tumors are well within the range of historical control data of the F344 strain of rats.

b. ***Page 1-7***

The Subpanel discusses the issue of the studies that tried to replicate the findings of vom Saal and colleagues not producing positive results for DES concomitantly tested. Although the "maximum effect" oral dose of DES was used in this study as previously reported and recommended by vom Saal (vom Saal 1996, vom Saal *et al.* 1997), no effects on prostate weight from DES treatment were observed in these studies. This was not an unanticipated finding since the dose of DES recommended by vom Saal was nearly three orders of magnitude below that reported by other researchers to produce effects on the male offspring of dams dosed by sc injection (100 μg/kg/day; McLachlan 1981; Newbold 1995). The NOEL for male offspring in the studies of McLachlan and

Newbold was 10 μg/kg/day, which is nearly two orders of magnitude higher than the dose recommended and used by vom Saal. Pylkkanen and coworkers (1996) reported that sc injection of 2 μg of DES per pup into 1- to 3-day old male mice neonates resulted in an increase in prostate weight; however, this direct sc dose to the males was about 727 μg/kg/day (about 3600 times higher than the oral dose to the dams used in these studies). Further, the NOEL expected for oral exposure to DES would be even higher than the 10 μg/kg/day NOEL determined by Newbold and McLachlan because of the extensive conjugative metabolism of DES in both the intestinal wall and liver (Metzler, 1981). Therefore, although the absence of any effect of DES was inconsistent with the findings reported by vom Saal *et al.* (1997), these findings are consistent with the previously reported NOELs for DES effects to male offspring.

c. ***Page 1-7, Question 3, point 1***

"Animals in these studies may have adapted so that they did not show a response to BPA at some "critical" time". This is speculative and does not consider the number of other studies with BPA that used only single, "critical" windows, and did not show the effects reported in the single study from vom Saal and colleagues. The basis for such a statement was not given at the Peer Review and no supporting data were presented to the Statistics Subpanel for evaluation.

5. Conclusions

The NTP Endocrine Disruptors Low-Dose Peer Review was intended to resolve a controversial and important environmental health issue. The organizers of the Peer Review should be commended for acquiring and distributing a large body of work from scientists throughout the world. The Subpanels had the proper charges from the Organizing Committee, namely, to evaluate the submitted data and reach conclusions about the shape of the dose-response in the low-dose region if such a response existed. The Statistics Subpanel made a Herculean effort and did a superb job of evaluating extensive data sets within a short period of time.

Unfortunately, despite the overwhelming weight of evidence against low-dose effects from BPA, the BPA Subpanel did not reach a definitive conclusion. Our comments are intended to complete the weight of evidence assessment and reach the conclusion that should have been reached by the BPA Subpanel. The overwhelming weight of evidence against low-dose effects from BPA is fully adequate to support a weight-of-evidence conclusion that low-dose effects from BPA have not been demonstrated and that reported low-dose effects of BPA are not credible.

We encourage EPA to carefully consider our comments in regard to the agency's review of reproductive and developmental toxicity testing guidelines and for any specific actions that may be directly related to BPA.

From: Karen_Neale@americanchemistry.com
Sent: Monday, July 16, 2001 4:32 PM
To: liaison@starbase.niehs.nih.gov
Subject: Low-Dose Comments

ACC Low-Dose Comments final 07-16-01.doc

(See attached file: ACC Low-Dose Comments final 07-16-01.doc)

To: National Toxicology Program, NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive, P.O. Box 12233, MD A3-01, Research Triangle Park, NC 27709.
919-541-0530 (phone); 919-541-0295 (fax);
liaison@starbase.niehs.nih.gov

cc: Vanessa Vu, OPPTS-OSCP, EPA
Gary Timm, OPPTS-OSCP, EPA
Anthony Maciorowski, OPPTS-OSCP, EPA
James Kariya, OPPTS-OSCP, EPA

Subject: Comments on the National Toxicology Program's Report of the Low Dose Peer Review (released on 5/14/01)

Dear Sir or Madam,

Attached are the American Chemistry Council's (the Council or ACC) comments on the National Toxicology Program's (NTP) Report of the Low Dose Peer Review. The Council is committed to improving the basic understanding of the endocrine issue through scientific research, and to developing a scientifically sound basis for public policy and product stewardship decisions relating to this issue. Member companies of the Council will be responsible for endocrine screening and testing under EPA's Endocrine Disruptor Screening Program (EDSP) when the program is implemented. The Council and its members played a constructive role in the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), contributing scientific and technical expertise over the two years of the Committee's work, which resulted in an August 1998 consensus report to EPA. The Council remains committed to working with EPA to implement a sound and dependable screening and testing program by working together with other

stakeholders to ensure that validation and standardization of proposed test methods receives top EPA priority.

The Council acknowledges the Environmental Protection Agency (EPA) for its efforts to address the low-dose hypothesis by requesting the National Toxicology Program (NTP) to conduct a peer review of the issue. Despite a number of serious shortcomings of the peer review process, as detailed in the attached comments, ACC believes that the NTP peer review sufficiently addressed the low-dose issue to allow EPA to move forward with implementation of the EDSP. Looking at the report and its conclusions as a whole, ACC believes that overall weight of the scientific evidence does not warrant changes to the traditional dose-setting approach because:
1) non-monotonic dose-responses in the low-dose region were not demonstrated across a variety of hormonally active agents;
2) non -monotonic dose-responses were not consistently observed across studies or laboratories; and,
3) the biological significance of the reported low-dose effects is scientifically uncertain, in particular with respect to relevance of such effects, if any, to adverse effects upon health of the organism and to establishment of No Observed Adverse Effect Levels for purposes of human risk assessments.

A hard copy of these comments will also be sent to your office by mail.

Sincerely,

Richard A. Becker, Ph.D., DABT
Senior Director, Public Health Team
American Chemistry Council
1300 Wilson Blvd.
Arlington, VA 22192
phone (703) 741-5000
e-mail Rick_Becker@AmericanChemistry.com

American Chemistry Council *Good Chemistry Makes It Possible*

COMMENTS ON THE NATIONAL TOXICOLOGY PROGRAM ENDOCRINE DISRUPTORS LOW-DOSE PEER REVIEW FINAL REPORT

Submitted by
The American Chemistry Council

July 16, 2001

Responsible Care®

1300 Wilson Boulevard, Arlington, VA 22209 ▪ Tel 703-741-5000 ▪ Fax 703-741-6000 ▪ http://www.americanchemistry.com

C-76

The American Chemistry Council represents the leading companies engaged in the business of chemistry. Council members apply the science of chemistry to make innovative products and services that make people's lives better, healthier and safer. The Council is committed to improved environmental, health and safety performance through Responsible Care®, common sense advocacy designed to address major public policy issues, and health and environmental research and product testing. The business of chemistry is a $460 billion enterprise and a key element of the nation's economy. It is the nation's largest exporter, accounting for ten cents out of every dollar in U.S. exports. Chemistry companies invest more in research and development than any other business sector.

EXECUTIVE SUMMARY

The American Chemistry Council (ACC) acknowledges the Environmental Protection Agency (EPA) for addressing the low-dose hypothesis by asking the National Toxicology Progam (NTP) to conduct a peer review. ACC and a number of its member companies were very active in and supportive of the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) process and consensus recommendations to EPA.

One of EDSTAC's recommendations was for EPA to undertake a project to address the controversy over whether endocrine-active substances demonstrate non-monotonic dose-response curves at environmentally relevant dosage levels. Specifically, EDSTAC recommended that the purpose of the project should be "to address the nature of the dose-response curves for exogenous estrogenic substances in order to allow more informed judgments about appropriate toxicology study designs for substances that have hormonal activity." EDSTAC Final Report at 5-50.

EPA asked NTP to undertake a peer review of the relevant studies to address this issue, which is pertinent to selecting doses for EPA's Endocrine Disruptor Screening Program (EDSP). NTP originally proposed a low-dose review that was consistent with EDSTAC's recommendation. However, the NTP Peer Review Organizing Committee failed to provide formal weight-of-the-evidence guidance to ensure a systematic comparison of studies to account for adequacy, strength and quality of the experimental design as well as consistency and coherence of results. Further, the Committee expanded the scope of the peer review when it expanded the definition of "low-dose" to include "biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in U.S. EPA's standard toxicity testing paradigm." This led the NTP Peer Review subpanels to deliberations that diverted focus away from the central topic: the shape of the dose-response curves. As a result, there were many shortcomings in the peer review and the report (see detailed comments).

ACC believes, however, that the NTP peer review sufficiently addressed the low-dose issue (i.e., non-monotonic dose-response) to allow EPA to move forward with implementation of the EDSP. Looking at the report and its conclusions as a whole, ACC believes that the overall weight of the scientific evidence does not warrant changes to the traditional dose-setting approach because:

1) The low-dose findings have not been demonstrated consistently across different studies of the same substance in independent laboratories;
2) The findings are not consistent for all substances with similar mechanisms of action; and,
3) The biological significance of the reported low-dose effects is scientifically uncertain, in particular with respect to relevance of such effects, if any, to adverse effects upon health of the organism.

In sum, the <u>overall</u> weight of the scientific evidence considered by the peer review subpanels does <u>not</u> demonstrate a non-monotonic dose-response in the low-dose region that is common

across a variety of hormonally active agents, consistent across studies or laboratories or relevant to establishment of No Observed Adverse Effect Levels for purposes of human risk assessments. Therefore, ACC has several recommendations to EPA.

- First, we urge the Agency to move forward with implementation of the EDSP using the existing, widely accepted dose-setting paradigm.

- Second, we believe that EPA should require within the EDSP the use of a formal weight-of-the-evidence approach for assessing additional or conflicting data, recognizing that additional research – by itself – will not resolve the issue unless a well-prescribed process for weighing the evidence is utilized. To do so, EPA needs to apply clear and consistent guidelines for establishing causality for adverse effects and weighing experimental results within and across studies.

- Third, EPA should consult with its Endocrine Disruptor Methods Validation Subcommittee (EDMVS) and appropriate peer review bodies such as EPA's Science Advisory Board and Science Advisory Panel before rejecting or modifying well-established assays and protocols (and requiring the use of many additional animals). EPA should ensure that the data demonstrate the inadequacy of such assays and protocols to detect adverse, endocrine-related effects, and that any suggested changes have undergone adequate validation for relevance and reliability. EPA should anticipate that there will be other instances when isolated, non-replicated findings will be used to challenge well-established toxicological tenets.

- Finally, we recommend that if EPA concludes that it should change any well-established assays, it should work with OECD to ensure international harmonization. Mutual acceptance of data could be adversely affected by modifications of internationally harmonized test guidelines made by EPA unilaterally.

TABLE OF CONTENTS

Page

COMMENTS OF THE AMERICAN CHEMISTRY COUNCIL ON THE NTP LOW-DOSE PEER REVIEW FINAL REPORT

I. Introduction.

In 1998, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) recommended that EPA undertake a project to critically review and weigh the evidence to address the controversy over whether endocrine active substances demonstrate bimodal dose-response curves. The question EDSTAC wanted EPA to address was whether the current dose-setting paradigm was sufficient for selecting doses for the Endocrine Disruptor Screening and Testing Program (EDSP). There was great disagreement among some EDSTAC members concerning this issue because, although a few small studies suggested the possibility of bimodal dose responses for some endocrine active agents (bisphenol A (BPA) and diethylstilbestrol (DES), other larger, more sensitive studies were unable to replicate the findings. The EDSTAC report was published in August 1988 and subsequently, EPA asked the National Toxicology Program to undertake a peer review of the relevant studies to address this issue.

The American Chemistry Council (ACC) and a number of its member companies were very active in and supportive of the EDSTAC process, and continue to be involved in related matters such as the validation of the EDSTAC recommended screens and tests. ACC also supported the proposed NTP low-dose peer review.[1] NTP proposed a low-dose review that would primarily address the evidence for "defining the shape of the dose/response curves for endocrine-active substances in the low-dose region" and "evaluate the likelihood and significance of . . . low-dose effects for humans." 65 Fed. Reg. 58097 (Sept. 27, 2000) In its February 2000 comments, ACC agreed that the low-dose peer review should focus on whether available data clearly demonstrate a non-monotonic dose-response in the low-dose region that is common, predictable, and relevant to humans.

Initially, NTP's objective was clearly stated, and was in alignment with the primary question posed by EDSTAC and EPA. However, the NTP Peer Review Organizing Committee failed to provide appropriate scope and direction for the peer review. As a result, the NTP peer review addressed issues that were well beyond the scope of the low-dose issue and did not fully address the primary questions posed by EPA. Such additional issues included whether endocrine active substances can exert any effects or changes at "low doses," whether non-adverse effects can be observed below a previously established NOAEL, whether specific assays effectively test for endocrine-related effects, and whether EPA's proposed EDSP screening and testing framework is adequate to identify endocrine disruptors. These issues exceeded NTP's originally proposed scope for the low-dose peer review meeting.

---

[1] See ACC's February 22, 2000 comments on the NTP proposed low-dose peer review, Docket OPPTS – 42208A. In February 2000 ACC was known as the Chemical Manufacturers Association and submitted comments under that name.

C-82

ACC is disappointed with the Organizing Committee's actions because they resulted in an unsystematic and rather unfocused, undertaking among the subpanels to address issues that were outside the originally proposed scope of the low-dose peer review. The low-dose peer review was not designed to address such issues. For example, a significant body of the scientific literature relevant to the adequacy of specific protocols or testing programs was not obtained in advance of the meeting, thus preventing a comprehensive scientific review of such issues by the expert subpanelists. Also, experts that could have provided important information concerning these issues were not consulted or asked to be members of the subpanels. Indeed, only very limited opportunities were provided for public participation in the low-dose peer review meeting.

The NTP charge to the peer reviewers resulted in a final report that is inconsistent and confusing. ACC is concerned that isolated comments in the NTP peer review report might be taken out of context to support unfounded claims concerning the possibility of low-dose effects or unwarranted challenges to the EDSP and current testing protocols. ACC is especially concerned that the peer review, in general, failed to fully and systematically weigh the scientific evidence concerning low-dose effects – resulting in an incorrect implication (especially in the NTP Executive Summary) that the low-dose hypothesis is widely and generally supported in the scientific community.

ACC also believes that a number of the subpanel conclusions are factually incorrect. Some of these opinions simply attempt to supplant the EDSTAC recommendations (which were developed by a large number of scientists and other stakeholders over a number of years) and EPA-endorsed testing guidelines. Unfortunately, outside of the panel participants, neither the public nor other knowledgeable and experienced scientists were provided an opportunity to review a draft NTP report and provide written comments that could correct mistaken conclusions.

Because the NTP low-dose peer review was not structured to address issues outside of the low-dose question, ACC believes that EPA should not accept conclusions of the NTP peer review that exceed the proper scope of the review. Instead, to the extent the low-dose peer review raises scientifically testable and verifiable issues that EPA wishes to pursue, the Agency should utilize appropriate advisory and peer review processes, including its Endocrine Disruptor Methods Validation Subcommittee (EDMVS), Scientific Advisory Panel (SAP) and Science Advisory Board (SAB). The Agency must also allow for adequate stakeholder input.

Despite shortcomings in the NTP low-dose peer review and report, ACC believes that the NTP peer review sufficiently addressed the primary low-dose question to allow EPA to move forward with implementation of the EDSP without changing the existing, widely accepted dose-setting paradigm. The <u>overall</u> weight of the scientific evidence considered by the peer review subpanels does <u>not</u> demonstrate a non-monotonic dose-response in the low-dose region that is common across a variety of hormonally active agents, consistent across studies or laboratories, common across strains and species, or relevant to establishment of NOAELs for purposes of human health risk assessments.

In sum, ACC believes that the NTP low-dose peer review raises a number of issues EPA must address when examining such controversial areas in the future. The agency must find a more rational and efficient mechanism to examine isolated claims, such as that of low-dose effects. We believe that EPA should develop (and require the use of) a weight-of-evidence approach for assessing conflicting data. To do so, EPA would need to provide clear and consistent guidelines for establishing causality for adverse effects and weighing experimental results within and across studies. In that regard, EPA should develop a policy for weighing very limited findings that cannot be replicated. In addition, although it supports the use of peer review to resolve scientific controversies, ACC believes that EPA should not undertake such peer reviews until it has established a clear scope of review and provided strict guidelines for such reviews. Furthermore, once established, such guidelines should be rigorously followed.

II. Comments Concerning the Scope of the NTP Low-Dose Peer Review.

A. The NTP review should have focused on the shape of the dose-response curve for endocrine-active substances.

In 1998, EDSTAC recommended that EPA undertake a project "to address the nature of the dose-response curves for exogenous estrogenic substances in order to allow more informed judgments about appropriate toxicology study designs for substances that have hormonal activity." EDSTAC Final Report at 5-50. Specifically, EDSTAC wanted EPA to critically review and weigh the evidence concerning whether endocrine active substances exert a bimodal dose-response. There was great disagreement among some EDSTAC members concerning this issue because, although there was some very limited data suggesting the possibility of bimodal dose responses for some endocrine active agents (BPA and DES), other larger, more sensitive studies were unable to replicate the findings. A number of EDSTAC members questioned the validity of studies purporting to demonstrate a bimodal dose-response given the limited nature of those studies, significant contrary findings by other laboratories, and the fact that the purported findings contradicted the well-established toxicological understanding for dose-response relationships. In response, EPA asked the National Toxicology Program to undertake a peer review of the relevant studies to resolve this issue.

In January 2000, NTP proposed conducting the low-dose peer review that was consistent with the EDSTAC recommendation. Defining "low doses" as "doses below the currently accepted No Observed Adverse Effect Level for [a] substance," NTP stated that the intent of the review would be to "evaluate the presence or absence of low-dose effects in specific studies, then evaluate the likelihood and significance of these and/or other potential low-dose effects to humans." NTP established that "the main topic to be addressed is evidence for defining the shape of the dose/response curves for endocrine-active substances in the low-dose region." See 65 Fed. Reg. 784 (Jan 6, 2000).[2] In September 2000, NTP maintained that the purpose of the peer review was to "evaluate dose-response relationships for endocrine disrupting chemicals" and that a "main topic to be addressed is defining the shape of the dose-response curves for endocrine-active substances in the low-dose region." 65 Fed. Reg. 58097 (Sept. 27, 2000).

Therefore, the issue raised by the low-dose hypothesis concerns the shape of the dose-response curve – *i.e.*, whether sufficient evidence supports the claim that the dose-response curve for endocrine active agents is non-monotonic. Although there appears to have been some

---

[2] In its criteria for selection of studies for review, NTP suggested that it intended to compare adverse low-dose effects to established NOAELs. NTP stated that "studies . . . will be considered critical if there is reason to believe that normal procedures for establishing a NOAEL would set NOAELs at a higher level than those indicated by the study in question..." 65 Fed. Reg. 784. NTP also stated that "studies which show effects at low doses but whose central issue in setting a NOAEL is either the definition of adversity or the completeness of the list of endpoints for which observations are made will not be considered relevant to the dose/response issues that this peer review will address." Id. Therefore, NTP rejected the notion that a low-dose effect could be a non-adverse effect observed below a NOAEL. It also established that it was not the purpose of the peer review to determine whether current assays address all relevant end points.

confusion during the NTP low-dose peer review, the issue has never been whether endocrine-active substances can exert an effect (adverse or non-adverse) at low doses or whether non-adverse effects can be observed below a NOAEL. The issue also is not whether specific assays adequately test for endocrine-related effects or whether EDSTAC's recommended screens and tests for EPA's Endocrine Disruptor Screening and Testing Program (EDSP) are adequate to identify endocrine disruptors. Indeed, EDSTAC addressed these issues for over two years and such issues will be further addressed as necessary by EPA, and by the agency's EDMVS and appropriate peer review bodies (such as EPA's SAP and SAB).

B. The NTP review exceeded the appropriate scope of a low-dose peer review.

Initially, NTP set the appropriate scope for a low-dose review -- to critically review evidence concerning whether endocrine active substances demonstrate bimodal dose-response curves -- that was consistent with EDSTAC's and EPA's desire to determine whether the current dose-setting paradigm was sufficient for selecting doses for the Endocrine Disruptor Screening and Testing Program. The NTP Peer Review Organizing Committee deviated from and expanded that scope when it changed the definition of "low-dose" to include "biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in the U.S. EPA's standard toxicity testing paradigm."[3] The Organizing Committee did not provide clear direction and criteria for conducting the review. The actions of the Organizing Committee resulted in the Peer Review Subpanels' adoption of arbitrary, inconsistent, and confusing operational definitions of "low-dose." The confusing focus of the review also apparently prompted some subpanels to consider peripheral matters not directly related to a review of the central issue of the shape of the dose-response curves, without the benefit of proper documentation, statistical re-analyses, etc., that would be required to thoroughly address those issues.[4]

The Organizing Committee's new definition required the peer reviewers to determine whether sensitive endpoints (here, biological responses) might be observed below doses that cause clearly adverse effects. This is not a particularly useful inquiry since biological effects (such as receptor binding, biochemical changes, or adaptive responses) are expected to occur below a NOAEL (i.e., the terms "NOAEL" and "NOEL" are not synonymous). Further, it has been suggested that some of these biological effects may actually be beneficial (Calabrese and Baldwin, 1998). Also, because the Organizing Committee's new scope for the low-dose review results in an exercise to search for more sensitive endpoints, it is contrary to EDSTAC's recommendation and NTP's earlier pronouncements. The inquiry envisioned by the Organizing

---

[3] See the "Overview" in the agenda for the October 10-12, 2000 low dose peer review. Whether biological changes may occur at environmentally relevant exposure levels is, to a large extent, irrelevant for determining the shape of a dose-response curve. The relevant question for the NTP peer review would have been whether endocrine active substances exhibit a bimodal dose-response curve for adverse effects. Therefore, the Organizing Committee should have defined "low-dose effects" as adverse effects that occur below established NOAELs.

[4] It is important to note that ACC does not oppose, per se, the expansion of the scope of a peer review to examine scientific issues of interest. ACC is, however, disappointed with the expansion of the scope of the low-dose peer review because the expansion confused the critical issue concerning the existence of non-monotonic dose-response curves and led to examination of other important issues that the panelists were inadequately prepared to address.

Committee does little to define the shape of dose-response curves. Rather, it results in a comparison of different endpoints that differ in, among other things, their level of adversity. NTP had earlier rejected such an inquiry (see note 2, above).

Because the Organizing Committee did not provide clear scope and direction for the review, each subpanel individually re-defined "low-dose," resulting in different operational definitions and, consequently, different scopes of review. As a result, the subpanels failed to clearly focus on the critical question for the review: whether there is evidence for bimodal dose-response curves. Instead, the subpanels wandered into issues beyond the proper scope of the review including: (1) whether established assays can detect all relevant endpoints (an issue NTP initially stated was not part of the review); and (2) whether non-adverse effects (e.g., biological changes) occur at doses below a NOAEL. The subpanels' use of different definitions of "low-dose" also makes it very difficult to draw a single conclusion from the subpanel reports concerning the possible existence of low-dose effects across chemicals and test systems.

The BPA Subpanel defined low-dose for BPA as a LOAEL for testis tumors[5] divided by a safety factor of 10 (apparently an effort to calculate a NOAEL). The Subpanel then determined whether doses of BPA below this calculated level (5 mg/kg/day) resulted in biological effects (*e.g.*, increased prostate weight). The Subpanel attempted to compare doses that might increase prostate weight to the calculated NOAEL for testis tumors. Such an analysis, however, cannot possibly determine whether BPA exhibits a bimodal dose response curve – the Subpanel compared two different effects.[6] Moreover, testis tumors are not an appropriate endpoint for BPA since NTP's own peer-reviewed conclusion was that there "was no convincing evidence that bisphenol A was carcinogenic for F344 rats or B6C3F$_1$ mice of either sex."

The Other Environmental Estrogens and Estradiol Subpanel correctly defined "low-dose effects" as effects "occurring when a nonmonotonic dose-response resulted in significant effects below the presumed NOEL (no-observed-effect level) expected by the traditional testing paradigm." The Subpanel, however, did not attempt to apply its own definition. Rather, it considered any difference between treated animals and control animals occurring at a dose lower than current NOAELs a "low-dose effect." It did not consider whether changes were an effect (adverse or not), or whether changes represented a nonmonotonic dose response or simply differences in unusual endpoints not previously measured. The Subpanel made no attempt to distinguish between "normal physiologic reaction" and "effect" ["effect" being used to signify "adverse", although not explicity stated].

---

[5] The use of testis tumors as an endpoint for BPA, however, was inappropriate because BPA treatment could not be judged to have increased testis tumor incidence since the rat strain used (F344) has an extremely high incidence of testis tumors in controls. The 50 mg/kg/day LOAEL in rats was based on reduced body weights and food consumption in the 103 week study.

[6] As discussed below, the weight of evidence demonstrates that increased prostate weight does not occur at doses below 5 mg/kg/day. In any event, the significance of increased prostate weight it is unknown. To the extent increased prostate weight is not an adverse effect, it was improper to compare a biological effect to an adverse effect for purposes of determining whether current assays are sufficient or whether BPA exhibits a bimodal dose-response curve.

Finally, the Androgens and Antiandrogens Subpanel chose to adopt the Organizing Committee's broad definition of low-dose. The Subpanel admitted that it had "arbitrarily defined the 'low-dose range' as one below the currently recognized NOAEL/LOAEL." Low-Dose Report at 3-2. However, the Subpanel concluded that no studies have been conducted below the NOAEL/LOAEL and, therefore, it could not determine whether any effects occur in the low-dose range. Although this should have been the end of the Subpanel's inquiry, it moved well beyond the appropriate scope of the review and criticizing the sensitivity of the Tier 2 multigeneration test despite the fact that a significant body of the scientific literature relevant to a thorough evaluation of specific protocols or testing programs were not obtained in advance of the meeting or made available to the peer reviewers.

Even after adopting definitions of "low-dose effects", the panels considered effects above the low-dose range. In some cases, the effects being considered were clearly not in the accepted "low-dose" range, being well above the established NOAEL values (see the BPA Subpanel's discussion of the Ben-Jonathan et. al. studies, Low-Dose Report at 1-5). In another case, the Other Estrogens Subpanel noted effects at doses or concentrations above its defined "low-dose" level. Even though no effects were seen in the "low-dose" area, the Subpanel chose to report the observed effects as "indicative of low-dose effects" (see the Subpanel's discussion concerning Methoxychlor, Low-Dose Report at 2-5).

III. Comments on the NTP Low-Dose Peer Review Conclusions Concerning the Low-Dose Hypothesis

A. The NTP review established that the evidence is not adequate to support the low-dose hypothesis.

The overall conclusions of the peer review panel demonstrate that the available data do not support a non-monotonic dose-response in the low-dose region that is common across a variety of hormonally active agents, consistent across studies or laboratories, common across strains and species, or relevant to establishment of NOAELs for purposes of human health risk assessments. Claims concerning the low-dose hypothesis have relied primarily on limited evidence for low-dose effects of BPA or "low-dose" effects from the natural estrogen, estradiol (E2), and DES – although by definition, the levels at which the natural estrogen exerts an effect are not "low-dose" effects" merely because they are small.[7] The BPA Subpanel concluded, after reviewing the relevant information, that it was "not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding."

---

[7] For example, the Other Estrogens Subpanel considered any change in hormone concentration biologically relevant, without reference to lack of consistency with other hormone levels and/or responses. Even if the stated responses for estradiol were clearly acceptable as "effects", however, there is no clear indication that these are "low-dose effects", since estradiol is a potent estrogen. In fact, the doses at which there seemed to be effects were within the physiologically active range for estradiol.

Proponents of the low-dose hypothesis cited one study by Nagel and colleagues (Nagel et al., 1997) that purported to observe increased prostate weight associated with low-dose exposures to BPA. This study used only two doses of BPA with only seven animals in each dose group. However, proponents of the low-dose hypothesis claimed that the study demonstrated a low-dose effect because it showed an effect at a low dose (2 μg/kg/day) when other studies showed no effect at higher doses. Proponents of the low-dose hypothesis also cited another study by vom Saal and Timms (vom Saal et al., 1997) that purported to demonstrate prostate enlargement at very low doses of DES (0.02 to 2 μg/kg/day) but not at a high dose of 200 μg/kg/day (the authors of this study did not provide data for independent verification and statistical review despite having been requested to do so).

Numerous other studies, however, were unable to repeat the Nagel et al. (1997) and vom Saal et al. (1997) findings. Ashby et al. (1999) observed no BPA effects on prostate weight at 2 and 20 μg/kg/day using 37 and 29 mice, respectively. They also observed no DES effects on prostate weight at 0.2 μg/kg/day. Ashby was unable to repeat the Nagel and vom Saal findings despite attempts to closely replicate the vom Saal protocol. A very large study conducted by Welsch (Elswick and Welsch, 2000) using multiple pups per litter also found no BPA effects on prostate weight at 0.005, 0.05, 0.5, or 5 mg/L drinking water (0.001 to 10 mg/kg/day).

Three other very large studies, conducted as GLP studies, also failed to show low-dose effects of BPA. Ema (2000) did not observe increased prostate weight in rats exposed to 0.2, 2, 20, or 200 μg/kg/day (25 male rats/dose) in a two-generation reproductive study. Likewise, Tyl et al. (2000) in a three-generation reproductive toxicity evaluation of BPA observed no increased prostate weight at 0.015, 0.3, 4.5, 75, 750, or 7500 ppm (1, 20, 300, 5000, 50,000, or 500,000 μg/kg/day) in any of the three generations examined (30 animals per sex per dose). Cagen et al. (1999a) also observed no BPA effects on prostate weight in mice at 0.2, 2, 20, or 200 μg/kg/day or in rats (Cagen et al., 1999b) at 0.01, 0.1, 1, or 10 ppm drinking water (ranging from 1-4, 12-38, 100-391, or 888-4022 μg/kg/day, depending on the stage of pregnancy/lactation and hence the amount of water consumption). Cagen also observed no DES effects on prostate weight in mice at 0.2 μg/kg/day or in rats at 0.1 ppm drinking water (4-36 μg/kg/day). Like the Ashby study, the Cagen mouse study attempted to replicate the vom Saal protocol.

Concerning the large number of studies that were unable to replicate the findings reported by vom Saal and co-workers or that were unable to observe elevated prostate weights, the panel found that: (1) the studies were very consistent; (2) their conclusions were supported by appropriate statistical analysis; and (3) the Statistics Subpanel confirmed the lack of BPA effects. The Subpanel concluded: "Collectively, these studies found no evidence for a low dose effect of BPA, despite the considerable strength and statistical power they represent, which the Subpanel considered especially noteworthy." The BPA Subpanel's overall conclusion was:

> "There is credible evidence that low doses of BPA can cause effects on specific endpoints. However, due to the inability of other credible studies in several different laboratories to observe low dose effects of BPA, and the consistency of these negative studies, the Subpanel is not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding. In addition, for those

studies in which low dose effects have been observed, the mechanism(s) is uncertain (i.e., hormone related or otherwise) and the biological relevance is unclear."

Peer Review Report at 1-11.

Therefore, the Subpanel discounted the purported low-dose observations as not established and generally not reproducible. Importantly, the Subpanel noted that the reported low-dose effects (e.g., increased prostate weight) may not be endocrine mediated and, in any event, the relevance of such effects is unclear (i.e., there is no evidence that such effects are adverse and such effects have no known relevance to human health).

Concerning DES, studies from the vom Saal and Newbold laboratories, which purport to show low-dose effects, were not subjected to independent review by the Statistics Subpanel (the data were not provided). By comparison, according to the BPA Subpanel, "other credible studies [*e.g.*, Ashby and Cagen] have not observed such low dose effects of DES or other estrogens." Significantly, data for these studies were submitted and fully assessed by the Statistics Subpanel. The lack of reproducibility brings the original findings into question.

B. The NTP review, in general, failed to properly weigh the scientific evidence.

The purpose of the low-dose review was to weigh the evidence concerning whether endocrine active chemicals may exhibit a bimodal dose-response curve.[8] Specifically, the Subpanels should have critically analyzed and weighed studies suggesting "low-dose effects" and those that were unable to replicate such findings to arrive at a definitive conclusion concerning the overall support for the low-dose hypothesis. As discussed above, the BPA Subpanel considered the quality and quantity of the evidence to the extent that it determined that a low-dose effect of BPA had not been conclusively established as a general or reproducible finding. The BPA Subpanel also considered the relevance of the purported low-dose findings stating that the mechanism for such effects is uncertain and their biological relevance is unclear. Nonetheless, ACC believes that the BPA Subpanel could have done more to weigh the evidence concerning the low-dose hypothesis. The other subpanels and the overall NTP Low-dose Panel did not employ a systematic method for establishing causality for adverse effects and weighing experimental results within and across studies.

The NTP Statistics Subpanel analyzed the raw data from the 1997 Nagel BPA study. It was unable to review the 1997 vom Saal DES study, however, because vom Saal did not provide the data sets for that study despite having been requested to do so. The Subpanel also analyzed the raw data from the Ashby, Welsch, Ema, Tyl, and Cagen studies discussed above concerning BPA and DES (the authors provided all data sets). Concerning findings for prostate effects of DES and BPA, the Subpanel found that Nagel et al. (1997) did find statistically significant differences in prostate weights in both BPA groups. The Subpanel also found that other

---

[8] In the Charges to the subpanels, the Organizing Committee asked the subpanels to define the "extent of empirical evidence demonstrating" or "demonstrating the lack of low-dose effects." The term "extent of empirical evidence" clearly implies that the Organizing Committee intended a weight-of-evidence approach. Nevertheless, none of the Subpanels used a systematic weight-of-evidence approach.

endpoints for which Nagel et al. (1997) had reported statistically significant changes could not be confirmed following statistical re-analysis. The Subpanel also confirmed the statistical analyses and accuracy of the findings of all the other investigators who had not observed increased prostate weights for BPA and DES, and had not observed the other changes in other endpoints vom Saal reported. The Statistics Subpanel, however, did not compare and contrast results across different studies, as had been requested.

Although the BPA Subpanel appeared to undertake some weighing of the low-dose studies, it is unclear whether the BPA Subpanel fully considered the weight of the Nagel et al. (1997) BPA prostate data when it found the Nagle et al. (1997) study to be "credible." It appears the BPA Subpanel mistook the Statistics Subpanel's statement that the statistics were "credible" to endorse the findings of the study as "credible."[9] Because of the severe limitations of the Nagel et al. (1997) it needs to be interpreted with considerable caution. Such limitations include:(1) many findings in that study originally reported as statistically significant could not be confirmed upon reanalysis (Low Dose Report at A-50); (2) the study was a very small (only seven mice per group), non-GLP study that purported to demonstrate subtle effects of uncertain relevance (Not following GLPs does not, in itself, make the study invalid, but it means that the data is not readily available for review and auditing by others who attempted to repeat these findings or independently analyze the results); (3) the findings contradict the well-established toxicological understanding for dose-response relationships; (4) the colony of CF-1 mice no longer exists so no further attempt to replicate is possible ; and (5) a number of other large, well conducted studies were unable to replicate the findings. Given these factors, one must wonder whether the Nagel findings were simply an artifact, whether the reported effects were related to something other than the BPA treatment, or whether something else unrelated to a true low-dose effect could explain the findings. The subpanel did not look for such explanations. The Subpanel, in reviewing the evidence that contradicts the Nagel findings, stated: "Collectively, these studies found no evidence for a low dose effect of BPA, despite the considerable strength and statistical power they represent, which the Subpanel considered especially noteworthy." In light of these findings and considerations, it is unclear how the BPA Subpanel found the Nagel study "credible."

Concerning DES, the BPA Subpanel stated that there are "credible reports that DES may produce low-dose effects, including studies from both the vom Saal and Newbold laboratories." Low-Dose Report at 1-10. This conclusion was not reached appropriately given that vom Saal did not provide the requested raw data on his DES study to the Statistics Subpanel despite being requested to do so and that the Statistics Subpanel did not analyze the Newbold data. The BPA Subpanel also noted that "other credible studies have not observed such low dose effects of DES or other estrogens." Unlike the vom Saal and Newbold DES studies, the studies that showed no low-dose effects for DES were submitted for review and were fully assessed by the Statistics Subpanel. The BPA Subpanel did not, however, compare and weigh studies concerning DES

---

[9] The BPA Subpanel's finding that the Nagel study was credible can be reconciled with the facts concerning the low-dose data only if the Subpanel meant by "credible" that the data were statistically significant and that the investigators observed the effects they purported to see. "Credible" can not mean that the study provides any degree of certainty to the low-dose hypothesis. This interpretation of the Subpanel's findings is consistent with its overall conclusion that a low-dose effect of BPA had not been conclusively established as a general or reproducible finding.

(studies not reviewed by the Statistics Subpanel should not have been considered or given equal weight to those studies that were reviewed by the Statistics Subpanel), leaving the incorrect impression that there is equal weight on both sides of the DES low-dose question.

The Other Environmental Estrogens and Estradiol Subpanel, without providing any analysis, simply concluded there is "very clear empirical evidence of a low dose effect" of DES on prostate size at 0.02 μg/kg in CF-1 mice and "supportive evidence" in CD-1 mice. Low Dose Report at 2-4. Presumably, the Subpanel was referring to the Nagel et al. (1997) study and the vom Saal et al. (1997) study, for which data were not provided for independent verification and statistical review. The Subpanel did not examine the studies and did not consider contrary data. In fact, the findings on CD-1 mice are unpublished and were presented for the first time during a presentation at the low-dose meeting. Clearly, the Subpanel did not attempt to fully weigh the low-dose data.

The Executive Summary also failed to report a balanced conclusion concerning the weight-of-the-evidence supporting the low-dose hypothesis. The Executive Summary reported that "several studies provide credible evidence for low-dose effects of bisphenol A" and that "several large studies in rats and mice, including multigenerational studies in Sprague-Dawley rats found no evidence for bisphenol A despite the considerable strength and statistical power those studies represent." Low Dose Report at iii. The Executive Summary does not discuss the limitations or strengths of the BPA studies. Rather it seems to imply that there is significant support on both sides of the low-dose controversy, when in fact there very limited findings suggestive of any low-dose non-monotonic effects versus six well-conducted studies in rats and mice by multiple routes of exposure from industry, governmental, and academic laboratories using statistically powerful numbers of animals that found no effects in any organ (including the prostate). Concerning DES, the Executive Summary simply concludes that "there is clear evidence of a low-dose effect" on prostate size by DES (at 0.02 μg/kg) in mice. This is not a conclusion that represents the weight-of-the-evidence, since it fails to recognize or take account of the studies that did not show DES effects on the prostate.

C. The NTP review departed from established procedural guidelines.

1. The subpanels improperly considered studies for which raw data were not provided for statistical analysis.

One of the primary issues surrounding the low-dose controversy concerned data quality. Given the importance of the low-dose issue and the greatly divergent findings from studies of significantly different statistical power, many (including EDSTAC members) felt that it was imperative that independent scientists examine the raw data underlying low-dose studies. This is because the quality of experimental data (including issues related to study design, statistical analysis and data interpretation) is not readily apparent in most peer-reviewed research articles.[10]

[10] The Low Dose Report states that the Statistics Subpanel analyses "provide greater insight on the experimental data than is typically apparent in most peer-reviewed research articles, consequently, the statisticians' report was critical for each of the subpanel reviews." Low Dose Report at ii.

This was to be one of the primary purposes of the low-dose peer review. Indeed, much effort was put into identifying, collecting, and analyzing key data.

Prior to the NTP low-dose peer review, fifteen principal investigators of primary research groups active in low-dose research were asked by organizers to provide raw data for independent statistical re-analyses. According to the ground rules specified by the organizers before the meeting, studies for which raw data were not submitted were not to be included in the low-dose review. Although, almost all of the principle investigators provided the requested data, a few data sets were not provided.[11] Despite the ground rules set for the low-dose review before the workshop, some of the data that was not provided and, therefore, not reanalyzed was considered and utilized by the low-dose subpanels. In fact, as discussed above, some of that data inappropriately formed the primary basis for subpanel conclusions.[12]

The subpanels should not have accepted claims of low-dose effects when data supporting such claims were not submitted in advance for independent review. The acceptance and utilization of such information contrary to ground rules prevented a meaningful assessment and weighing of the data, and set a poor precedent for future peer reviews. In the future, investigators may not go to the trouble and expense of organizing and collating data for agency peer reviews (and undergo a detailed scrutiny of their work) when they see that the conclusions of other investigators who don't submit data are readily accepted and utilized. Certainly, investigators who sought a fair and balanced evaluation of their data and contradictory data from other laboratories must be disappointed by the Organizing Committee's allowance of such a deviation from the fundamental principles of this independent peer review process.

2. Some of the subpanels considered effects from non-standard, unvalidated tests with unknown toxicological significance.

The Organizing Committee instructed that "low-dose effects" were biological changes that occur in the range of human exposures "or at doses that are lower than those typically used in the EPA's standard testing paradigm for evaluating reproductive and developmental toxicity." Low-Dose Report at i (emphasis added).[13] The subpanels considered, however, effects from non-standard assays that reported effects or changes in endpoints of unknown biological

---

[11] It is important to note that one of the primary proponents of the low-dose hypothesis failed to provide five of the eleven data sets requested. Low Dose Report at A-3.

[12] In addition to the studies discussed above, the BPA Subpanel extensively discussed the Howdeshell *et al.* and Ben-Jonathan et al. papers. Low-Dose Report at 1-4 to 1-5. The Subpanel found these studies "very credible and consistent" despite the fact that data for those studies were not provided to the Statistics Subpanel for analysis and independent confirmation of statistical significance.

[13] As discussed above, ACC believes that this is an impermissibly broad definition of "low-dose" for purposes of the low-dose review. Nonetheless, ACC believes the Organizing Committee was correct in directing the Subpanels to consider only evidence from standard toxicity testing protocols.

relevance based on methods of unknown reliability with few quality controls.[14] For example, the BPA Subpanel relied on the work reported by Ben-Jonathan's group, which used silastic mini-pumps implanted subcutaneously to deliver BPA. This is clearly not a "standard testing paradigm" recommended by EPA test guidelines. The Subpanel also relied on the vom Saal studies that used very small numbers of animals and limited doses under "experimental" conditions. These studies also do not conform to EPA-recommended standard testing guidelines. Further, without a clear understanding of the reliability of test methods and relevance of endpoints, it is inappropriate to compare doses from such studies that may elicit effects of unknown toxicological significance with results derived from standardized and validated methods for purposes of evaluating EPA's standard testing paradigm.

The Other Environmental Estrogens Subpanel relied on the NCTR studies that used non-standard and unvalidated immunotoxicologic tests to evaluate nonylphenol, methoxychlor, and genistein. The NCTR researchers measured the SDN-POA and proliferation of splenic T-lymphocytes stimulated with anti-CD3 as indicators of biologic activities, which the Subpanel concluded were "low-dose effects in F1 rats." Such assays have not been validated: the relevance of these endpoints has not been established nor has the reliability of the test methods.

Finally, the Other Environmental Estrogens Subpanel also concluded that "low-dose effects in F1 rats following dietary exposure to 25 ppm [of nonylphenol] include a decrease in SDN-POA in males, an increase in relative thymus weight, an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3, and a prolonged estrus in females." Low-Dose Report at v. However, the biological/toxicological significance of the differences between treated and control animals is unknown. These effects appear to occur at all doses and for several chemicals that may have different mechanisms of action, and they exhibited no clear dose response (monotonic or non-monotonic). In fact, the Subpanel concluded, "as stated above the SDN-POA changes are difficult to interpret at the moment, as are the anti-CD3 findings." Because these endpoints have not been well characterized or shown to be reproducible with an accepted degree of consistency in validation studies, and are not routine measurements in toxicity tests, it is problematic and difficult if not impossible to use such observations even as purported indicators of "low-dose effects."

D. The Executive Summary of the NTP Low-Dose Report overstates the support for the low-dose hypothesis.

The Executive Summary states as an overall conclusion: "Low-dose effects as defined for this review, were demonstrated in laboratory animals exposed to certain endocrine-active agents." This statement reflects bias and fails to recognize that evidence for low-dose effects is dependent on the definition of "low-dose," which was improperly broad for purposes of the low-

---

[14] ACC is not suggesting that such data are not useful. Indeed, basic research often employs non-validated methods or non-guideline protocols. However, it is problematic to place such investigations on an equal plain with standardized and validated guideline studies. Further, without a clear understanding of the reliability of test methods and relevance of endpoints, it is inappropriate to compare doses from such studies that may elicit effects of unknown toxicological significance with results derived from standardized and validated methods for purposes of evaluating EPA's standard testing paradigm.

dose review. The Executive Summary simply misses the point of the low-dose review: to determine whether endocrine active substances demonstrate a bimodal dose-response curve. In any event, the statement is misleading. Taken as a whole, the overall report indicates that the available data do not demonstrate a non-monotonic dose-response in the low-dose region that is common across a variety of hormonally active agents, consistent across studies or laboratories, common across strains and species, or relevant to establishment of NOAELs for purposes of human risk assessments.

In a number of places, the Executive Summary overstates the support for the low-dose hypothesis. For example, it states that the "shape of the dose-response curves for these [low-dose] effects varies with the endpoint and dosing regimen, and may be low-dose linear, threshold-appearing, or non-monotonic." This incorrectly implies that the shapes of the dose-response curves for endocrine-active agents have been conclusively determined in the low-dose region and that non-monotonic curves have been observed for a number of substances. A major flaw of the low-dose review, however, is that it did not systematically evaluate the shape of the dose-response curves. Therefore, it is not clear upon what NTP based its conclusion: it is surmised that these conclusions were drawn from the hypothetical models provided by the Dose-Response Modeling Subpanel. The only subpanel statement that provides support for the Executive Summary conclusion is the following statement concerning estradiol: "Overall, the TIER 1 dietary study demonstrated the three types of dose-response relationships based on particular endpoints, tissue responses and time points." Low Dose Report at 2-5. However, the meaning and basis for this statement is unclear. The Subpanel compared different endpoints, from different types of assays. It admitted that estradiol is "difficult to evaluate in an in vivo system because of the homeostatic mechanisms of the test systems." Finally, the subpanel concluded that the "remaining data gap for estradiol is to determine the shape of the dose-response curve . . . at the low effect and high no effect levels." Therefore, the Executive Summary's conclusion is an overstatement.

The Executive Summary also did not convey the tone set by the BPA Subpanel concerning the overall evidence for low-dose effects. The Executive Summary reported that (1) "several studies provide credible evidence for low-dose effects of bisphenol A"; and (2) "several large studies in rats and mice, including multigenerational studies in Sprague-Dawley rats found no evidence for bisphenol A despite the considerable strength and statistical power those studies represent." Low Dose Report at iii. The Executive Summary omitted the BPA Subpanel's additional comment that it found the "considerable strength and statistical power" of the studies that did not show a BPA effect on prostate weight "especially noteworthy."[15] The Executive Summary also does not provide any discussion concerning limitations or strengths of the BPA studies, which demonstrate that the Nagel, et al. (1977) study was a single, very small study, whereas numerous large, powerful studies did not show a BPA effect at low doses.

Finally, the Executive Summary stated that there is clear evidence of a low-dose effect on prostate size by DES. It failed to point out, however, that the evidence is supported by only two

[15] It is also interesting to note that although the BPA Subpanel insisted that its overall conclusion concerning low-dose effects was to be presented in its entirety, the Executive Summary excluded an important part of that conclusion: "for those studies in which low dose effects have been observed, the mechanism(s) is uncertain (i.e., hormone related or otherwise) and the biological relevance is unclear."

studies for which the data were not submitted for independent re-evaluation prior to the meeting, and that other larger studies (for which raw data were provided) found no low-dose DES effect.

ACC believes that the Executive Summary is misleading concerning the overall weight of the evidence supporting the low-dose hypothesis and is concerned that the casual reader, who is likely to read only the Executive Summary, may conclude that there is a greater degree of empirical evidence in support of the low-dose hypothesis than is actually the case. ACC believes that EPA will have to ensure an accurate portrayal of the entirety of the scientific information when addressing concerns triggered by unempirical perceptions of the low-dose hypothesis.

E. EPA should move forward to implement its EDSP without changing the existing dose-setting paradigm.

When viewed in its entirety, the data on potential low-dose effects of endocrine active agents do not support a change to the current dose-setting paradigm. Despite the shortcomings of the NTP low-dose peer review and report, ACC believes that the overwhelming weight-of-the-evidence from the NTP peer review does not support the need to change the dose-setting paradigm for the EDSP. Therefore, ACC believes that EPA should move forward with implementation of its EDSP without changing the existing dose-setting paradigm.

As discussed above, the BPA Subpanel stated that in light of the inability of "credible studies in several different laboratories to observe low dose effects of BPA, and the consistency of these negative studies," it was "not persuaded that a low dose effect of BPA has been conclusively established as a general or reproducible finding." The Subpanel further concluded that "for those studies in which low dose effects have been observed, the mechanism(s) is uncertain (i.e., hormone related or otherwise) and the biological relevance is unclear." This last statement is particularly relevant to the EDSP because Tier 1 Screening was designed to use validated assays to detect substances capable of interacting with one or more components of the endocrine system, and Tier 2 Testing was designed to use validated tests to quantify adverse effects and delineate the dose-response relationship for use in human health risk assessment.

ACC believes that the overall findings of the BPA subpanel are adequate to resolve the low-dose issue for purposes of EPA's implementation of the EDSP.[16] In fact, the overall conclusions of the peer review subpanels indicate that the available data do not demonstrate a non-monotonic dose-response in the low-dose region for endocrine-active substances that is common across a variety of hormonally active agents, consistent across studies or laboratories,

---

[16] To the extent EPA concludes, however, that further research is necessary to test the low-dose hypothesis (the Peer Review Report suggested conducting further low-dose studies), it should not recommend such research until it first establishes a weight-of-evidence process for evaluating relevant studies. Indeed, the failure to universally utilize a weight-of-evidence approach was one of the most serious shortcomings of the NTP review. The Report itself demonstrates the challenges of examining a scientifically controversial issue, such as the low-dose hypothesis, when the overall weight of studies is not assessed. EPA must ensure that this problem is not repeated.

common across strains or species, or relevant to establishment of NOAELs for purposes of human risk assessments.

EPA should determine the relevance of biological effects and the reliability of assay methods before it even considers making changes to the current risk assessment paradigm or the current methodology for selecting target doses.[17] As discussed above, only unequivocal and reproducible evidence of adverse effects observed below an established NOAEL would support consideration of changes to the dose-setting paradigm for the EDSP.

IV. Comments Concerning the NTP's Conclusions Related to the Sufficiency of Specific Assays in the EDSP to Detect Endocrine-Related Effects.

A. The NTP Low-Dose Peer Review was not the appropriate forum for assessing the adequacy of the multigeneration assay and the EDSP for detecting endocrine-related effects.

As discussed above, the low-dose controversy is focused on whether endocrine active agents may exhibit non-monotonic dose-responses. This issue framed the appropriate scope of the low-dose review. There was no indication in the *Federal Register* announcements prior to the meeting that adequacy of protocol designs other than dose selection procedures would be considered. Indeed, in order to properly and thoroughly assess the adequacy of complex protocols like the multigeneration reproduction study would require consideration of an enormous data base that was not made available to the meeting participants, and would have required direct participation of many additional reproductive and developmental toxicologists. The low-dose issue is not whether specific assays are adequate to screen and test for endocrine-related effects or whether EDSTAC's recommended screens and tests for EPA's EDSP are adequate for risk assessment purposes. Indeed, EDSTAC addressed these issues for over two years and such issues will be further addressed as necessary by EPA, and by the agency's EDMVS and appropriate peer review bodies (such as EPA's SAP and SAB).

Some of the subpanels discussed the sufficiency of the current multigeneration reproduction study protocol to characterize the potential hazards (such as developmental effects) of endocrine-active chemicals. In addition to criticizing the EDSTAC's Tier 2 multigeneration assay, one panel criticized the adequacy of the EDSP to detect environmental androgen and antiandrogens. ACC believes that a number of the subpanel opinions, such as these, do not reflect wider scientific views. Some of these opinions are contrary to the EDSTAC recommendations (which were developed by a large number of scientists and other stakeholders over a number of years) and EPA-endorsed testing and risk assessment guidelines.

---

[17] The BPA panel questioned the biological significance of purported low-lose effects and the Executive Summary acknowledge that the "toxicological significance of many of these effects has not been determined." Low Dose Report at vii.

The low-dose peer review was not structured to address such issues. A large body of relevant studies and data that address issues related to the adequacy of specific protocols or testing programs were not obtained or circulated to peer reviewers in advance of the meeting, thus preventing a comprehensive scientific review of such issues. Also, experts that could have provided important information concerning these issues were not consulted. Indeed, only very limited opportunities were provided for public participation in the low-dose peer review. These failures led to erroneous conclusions (see the discussion below) concerning the adequacy of the multigeneration assay and the EDSP to detect and characterize endocrine-related adverse effects. Unfortunately, neither the public nor knowledgeable and experienced scientists were provided an opportunity to review a draft NTP report and provide written comments that could correct mistaken conclusions.

Because the NTP low-dose peer review was not structured to address issues outside of the low-dose question, particularly complex matters like the adequacy of the multigeneration study, ACC believes that EPA should not accept conclusions of the NTP peer review that exceed the proper scope of the review. Instead, to the extent the low-dose peer review raised questions EPA might wish to pursue, the agency should utilize appropriate advisory and peer review processes, including its EDMVS, SAP and SAB. The Agency also should allow for adequate stakeholder input.

B. The evidence does not support a conclusion that the mulitgeneration assay is insufficient to detect certain endocrine-related effects.

During the low-dose review a few panelists claimed that the multigeneration assay may lack the sensitivity and power to detect and characterize endocrine-active chemicals. Although it was not asked to evaluate the adequacy of existing test guidelines in areas other than dose-setting procedures, and not provided adequate information to do so, the Biological Factors Subpanel concluded that "there are several severe limitations to the 'Tier 2' [multigeneration] test, particularly with regard to the detection of low incidence phenomena." Low Dose Report at 4-5. The Androgen Subpanel concluded that the "existing multigenerational tests . . . have missed malformations and low-dose effects of EACs." Low Dose Report at 3-5. As discussed more fully below, the Subpanels were incorrect in concluding that the multigeneration studies have failed to detect anti-androgenic effects. The Subpanels' erroneous conclusions stemmed from the Subpanels' reliance only on selective studies, and not on the entire body of relevant studies from the available scientific literature.

Citing the examples of linuron and diisononyl phthalate (DINP), the Biological Factors Subpanel concluded that "a number of agents have now been shown to have endocrine activity even though well conducted multigeneration and prenatal studies by competent laboratories were negative in the standard study design." Low Dose Report at 4-5. In its evaluation of the capacity of the current multigeneration study to detect endocrine active compounds, the Biological Factors Subpanel did not appear to review studies conducted according to the newest guidelines, which were issued in 1998. These guidelines include numerous end points that are endocrine-sensitive but were not routinely included previously in the multigeneration study design (e.g., estrous cycling, sperm analysis, puberty onset, enumeration of implantation sites

and primordial follicles, reproductive and accessory sex gland organ weights and histology, and anogenital distance as a triggered endpoint). Additional endocrine-sensitive end points also are under consideration for addition to the EDSTAC Tier II multigeneration study design. Indeed, neither the linuron nor the DINP reports cited by the subpanel included all the endpoints listed in the 1998 multigeneration study guideline.

Specifically, the Subpanel appeared to rely only on McIntyre et al., 2000, as a purported indication of the shortcomings of the mammalian multigeneration protocol. In this study, the authors state that their findings of reproductive tract malformations "is in contrast to previous developmental toxicity studies that demonstrated linuron was neither teratogenic nor embryotoxic when administered to pregnant rats from GD 6 to GD 15 or to rabbits from GD 8 to GD 16 (Khera et al, 1978)." However, the claims by McIntyre et al. are not substantiated based on a wider review of readily available public information.

With respect to developmental toxicity, the EPA Registration Eligibility Decision (EPA RED) document for Linuron (EPA 728-R-95-003, March 1995) states that the LOEL for developmental toxicity in rats (49.8 mg/kg/day; NOEL 12.1 mg/kg/day) was based on "on increases in postimplantation loss and increases in the litter and fetal incidences of resorptions." (MRID 00018167). Clearly, the study cited in the EPA RED showed linuron exhibited embryotoxic effects. In rabbits, the EPA RED reported that the "developmental toxicity NOEL was determined to be 25 mg/kg/day, based upon an increased number of abortions, decreased mean number of fetuses per litter, decreased fetal body weight, and increased incidence of fetuses with skeletal variations of the skull at the 100 mg/kg/day level (the developmental toxicity LOEL) (MRID 00153867)". Again, in contrast to the statement by McIntyre et al. (2000), this earlier study clearly demonstrated the embryotoxic potential of linuron.

With respect to reproductive toxicity, McIntyre et al state that the "original multigenerational reproductive study, which utilized 125 ppm (approximately 6.25 mg/k/day) dietary levels of linuron (Hodge et al., 1968) did not indicate reproductive toxicity." McIntyre et al. also state that in their study "linuron induced testicular lesions in two offspring in two separate litters at 12.5 mg/kg/day. This is currently the lowest dose of linuron that has been shown to induce reproductive toxicity in laboratory animals." In actuality, as indicated in the 1995 EPA RED, the reproductive toxicity LOEL from the 3-generation reproduction study in rats was 6.25 mg/kg/day.

The EPA RED, which is publicly available, and was published in 1995, cites two multigenerational rat reproduction studies. The first study, a three-generation study, is apparently the one cited by McIntyre et al. First, it is important to note that this study used three dose levels of linuron, 25, 125 and 625 ppm in diet. McIntyre et al. imply that the 125 ppm level was the highest dose tested, and this is not the case. More importantly, however, is that reproductive toxicity effects were clearly reported to have been observed in this 3-generation study. The EPA RED states "The reproductive toxicity NOEL was 25 ppm (1.25 mg/kg/day) and the reproductive toxicity LOEL was determined to be 125 ppm (6.25 mg/kg/day) based on the following findings. Fertility was reduced in generations at 625 ppm F2a through F3a. Pup survival was consistently decreased for pups at 625 ppm, with most deaths occurring in the first 24 hours postpartum, and a trend for decreased viability from days 1-4. Weanling body weights

were decreased for F1b and F2b male and female pups at 125 ppm and 625 ppm. Absolute liver and kidney weights of weanlings (both sexes) were decreased, and histopathology of the 625 ppm F2b weanlings identified a frequent incidence of liver atrophy (decreased cytoplasmic clear spaces of hepatocytes)."

Furthermore, McIntyre et al. fail to cite the 2-generation rat reproduction study of linuron, which was reported in the EPA RED. The EPA RED states that in this study, at the 625 ppm level "testicular and epididymal abnormalities (testicular atrophy and intratubular fibrosis; epididymal deformities, inflammatory response and/or oligospermia) and ocular abnormalities (mineralization of the cornea; lens degeneration) were observed at histopathological evaluation of the F1 adults. Further evaluation of reproductive organ weight and hormone data from the F1 adults of this 2-generation study combined with an in vitro analysis of the ability of linuron and its metabolites to compete for binding to the androgen receptor resulted in the conclusion that linuron is a weak androgen receptor antagonist. These results support the hypothesis that rats exposed to linuron could develop interstitial cell hyperplasia and subsequent adenomas (Leydig cell tumors) of the testicular tissue via a mechanism of sustained hypersecretion of luteinizing hormone induced by the antiandrogenic potential of linuron". (MRID 41463401, 41864701, 41630101)

This case illustrates the problems arising out of the unfocused charge developed by the Organizing Committee. In actuality, and in contrast to the apparent conclusions reached, the scientific evidence does not support a sweeping conclusion that the mammalian multigeneration assay is insufficient to characterize certain endocrine-related effects. Had the body of the scientific literature relevant to the adequacy of specific protocols or testing programs been examined comprehensively, it is unlikely that such a panel would have reached such an erroneous conclusion. Furthermore, the likelihood of detecting linuron or other endocrine-active compounds is greater with the endocrine-sensitive end points added to the reproduction toxicity study in 1998.

The Subpanel also cited the example of DINP as a compound not detected using standard multigeneration and/or prenatal study designs.[18] A multigeneration study with DINP found no reproductive effects at dose levels up to approximately 500 mg/kg/day in the diet (Waterman, et al., 2000). Gray, et al. (2000) reported reproductive system effects (e.g., nipple retention in males, small/atrophic testes, flaccid fluid-filled testes) with DINP treatment, which is the basis for the subpanel's concern about the adequacy of the Waterman study. However, the dose in the Gray study was 750 mg/kg/day administered by gavage from gestation day 14 to lactation day 3. There were 4 of 52 pups with effects – two had testicular effects and two had retained nipples. By combining these effects, Gray reported a statistically significant increase ($p<0.04$). However, there is question whether this was an appropriate grouping or whether an appropriate statistical test was applied. Even granting the increase was real, however, it occurred only at a relatively high exposure, by the route of administration (gavage) that results in the highest blood levels of the compound. In Waterman et al. (2000), the principal effect was on body weight gain, which

[18] There was no notice to stakeholders that DINP data would be considered at the workshop and thus no opportunity to provide comment to the subpanel on the DINP data. Therefore, we believe it was inappropriate for the subpanel to consider DINP data, and certainly is inappropriate to include DINP as an example in the workshop report.

was reduced even in the low dose group. Thus, both the Waterman et al and Gray et al data clearly demonstrate that the effects of DINP on male reproductive development were at levels that were quite high relative to other effects. The effects identified by Gray were not low dose effects and did not occur at levels below the NOAEL, as defined by Waterman. Thus, the DINP data do not support a finding that the multigeneration protocol is inadequate for detecting low-dose endocrine effects.

Further, the multigeneration study is not meant to be a screening study for endocrine effects, but rather is a Tier 2 study to help define risk assessment parameters if effects are seen in Tier 1. The point of the EDSTAC program should not be to determine whether a chemical can ultimately cause endocrine effects at some dose, however high. Rather, it should be to determine whether endocrine effects may occur at doses that are relevant for human risk assessment. Where effects occur only at doses well above other sensitive endpoints and well above human exposures,[19] they may be academic interest, but are not of practical interest for risk assessment purposes.

In addition, in its assessment of guideline study requirements, the Biological Factors Subpanel compared the prenatal toxicity and multigeneration study designs. In that comparison, the Subpanel made several misleading statements that fail to adequately describe the current requirements for guideline testing. For example, the Subpanel mentioned that every pup is examined in at least 20 litters in the prenatal study. While this statement is true, the examination that is given to all pups is an external examination for gross anomalies. A similar examination to detect gross external alterations is also conducted on all available pups in a multigeneration study. In order to detect the alterations alluded to by the Subpanel (e.g., altered male reproductive tract development in the 2 referenced studies), an internal examination of visceral organs would be required. This type of examination is not conducted on all pups in a developmental toxicity study, but rather one-half of the pups are examined. Although decreased epididymal size would not necessarily be detected during these examinations, missing epididymides could be detected. Furthermore, the Subpanel recognized that only 10 animals per group are scheduled for pathological investigation in a multigeneration study; however, this statement applies to the EPA's multigeneration study design. Although most aspects of multigeneration study protocols are harmonized, the OECD and Japan's MAFF guidelines require histological examination of reproductive organs and associated tissues from all high dose and control P and F1 parental animals (one/sex/dose). Given the cost of mulitgeneration studies, most studies would be conducted to include histology on these additional 10-20 animals/group to meet the more stringent OECD guidelines so that the study would be accepted internationally.

Without providing any analysis, the Androgen Subpanel stated that: "Existing multigenerational tests sample only a small number of pups for necropsy and have missed malformations and low-dose effects of EACs. The problem is more serious when the endpoints

---

[19] The biomonitoring data released last March by the Centers for Disease Control and Prevention shows that median human exposure to DINP is below detection, and the 90th percentile exposure is 0.9 µg/kg/day – over 800,000 less than the dose in the Gray study. *See* CDC National Report on Human Exposure to Environmental Chemicals (www.cdc.gov/nceh/dls/report) *and* Phthalate Esters Panel of the American Chemistry Council's Statement on What does the CDC National Exposure Report Say About Phthalate Exposures?, June 7, 2001 (http://www.phthalates.org/penewsite/html/cdcref061501.htm)

are low frequency events." Low Dose Report at 3-5. The Subpanel provided the single example of DDT studies that, it claimed, failed to detect the androgenic effects of p,p'-DDE. The example, however, does not demonstrate a deficiency in the methodology of the multigenerational test and is not related to sample size. The issue appears to concern the strain of rat used for testing (see O'Connor et al., 1999).[20] The Sprague Dawley rat is commonly used in the multigeneration reproductive study. Evidence indicates that low doses of DDE do not exert an antiandrogenic effect in Sprague Dawley rats. Other assays, namely the Hershberger assay, using Long-Evans rats have detected antiandrogenic effects for DDE. These findings do not provide a basis for rejecting the use of the multigeneration assay or Sprague Dawley rats. In toxicological testing, species differences are not unexpected and the choice of species must take into consideration a number of issues, such as applicability of results to humans and the reasonable use of experimental animals. In addition, testing batteries may be designed to identify weak acting compounds such as DDE without having to alter the multigeneration assay. O'Connor et al., (1999) concluded that although strain selection may affect the ability to detect certain weak EACs, "a Tier I screening battery consisting of both *in vivo* and *in vitro* endpoints would reduce the chance that weak-acting compounds such as p,p'-DDE would not be identified as potential EACs." In any event, species-related issues can not be adequately addressed in a closed review meeting that is not designed to address such issues. To the extent EPA decides to address such issues, they should be addressed by the EDMVS.

Finally, the NTP panel drew sweeping conclusions in the Executive Summary stating, for example, that "critical endpoints such as cancer have not been evaluated in multigenerational studies." Report at vii. This is not a fair criticism of the multigeneration assay. The multigeneration assay is not designed to evaluate endpoints such as cancer and, therefore, the assay is not deficient in that respect. It is simply improper to expect every assay to pick up every endpoint. The mulitgeneration assay is designed to examine numerous endpoints, such as reproductive effects, birth defects, developmental effects and endocrine effects. Indeed, the multigeneration assay is highly relevant to an endocrine screening and testing program. However, the multigeneration assay, as currently designed, does not retain animals for purposes of detecting tumors. Cancer bioassays are designed for that purpose.[21] To the extent specific cancers may be related to endocrine effects (such as chronic stimulation or alteration in feedback loops), however, the multigeneration assay and other EDSP assays are likely to detect mechanisms or reproductive effects that frequently (but not always) result in tumors.

Finally, the panelists provided no evidence to support the sweeping claim in the Executive Summary that "certain endocrine active chemicals" were negative in standard multigeneration studies. The NTP panel did not identify the "certain" chemicals, but presumably were referring to linuron and DINP. As discussed above, it is not true that linuron was negative in the multigeneration assay and claims concerning the inability of the multigeneration assay to detect certain reproductive effects of DINP are based on improper comparisons between studies.

---

[20] O'Connor, J.C., Frame, S.R., Davis, L.G. and Cook, J.C. (1999). Detection of the Environmental Antiandrogen p,p'-DEE in CD and Long-Evens Rats Using a Tier I Screening Battery and a Hershberger Assay. *Toxicol. Sci.* 51, 44-53.

[21] It has been shown that prenatal exposures in cancer bioassays do not improve the ability of these assays to detect tumorgenicity. Therefore, it is not necessary to rely on a multigeneration assay (and its dosing regime) to detect tumors. See, e.g., Chhabra, et al., 1993, 1993a and 1994.

C. EPA should consult with its EDMVS, and appropriate peer review and international bodies if it has any concerns regarding the adequacy of the multigeneration assay or the EDSP for detecting endocrine-active substances.

The evidence does not support a conclusion that the EDSP is insufficient to detect endocrine disruptors. The current EDSP screening and testing scheme was arrived at by a large group of scientists with experience in screening and testing. These scientists, along with other stakeholders (including agency representatives), developed the screening and testing scheme after over two years of consultation. The scheme also underwent extensive peer review. EPA should be careful not to attach inappropriate significance to the recommendations arising from this selective and limited review of screening and test methods. Any questions concerning the adequacy of the EDSTAC screening and testing scheme should be referred to EPA's EDMVS and the Agency's SAP/SAB.

Before rejecting well-established assays and protocols (and requiring the use of many more additional animals), EPA should ensure that the data demonstrate the inadequacy of such assays and protocols to detect adverse, endocrine-related effects. The low-dose review subpanels failed to provide such data although it freely criticized the multigeneration assay and the EDSP. To the extent the subpanels cited scientific literature, such literature was not appropriately evaluated, as demonstrated by the incorrect conclusions drawn concerning linuron and DINP. In any event, these issues were outside the proper scope of a low-dose review and, therefore, were not adequately considered. For these reasons, EPA should not rely on the NTP low-dose peer review conclusions concerning the adequacy of the multigeneration assay and the EDSP to detect endocrine disruptors.

To the extent the NTP low-dose peer review raises issues concerning specific assays or the EDSP that EPA wishes to examine, the agency should work with its EDMVS, appropriate peer review bodies (such as its SAB and SAP) and scientific experts (especially those with experience conducting such assays). EPA must also utilize appropriate public comment and peer review processes when proposing any modification to established assays or the endocrine screening and testing program or the dose setting paradigm. Finally, if EPA concludes that it should alter well-established assays, such as the multigeneration assay, it should work with OECD to ensure international harmonization.

V. EPA Should Establish a More Rational and Efficient Mechanism to Examine Isolated Claims Such as That of Low-Dose Effects.

The NTP low-dose peer review, while representing a substantial effort by many competent and well-respected scientists, lacked adequate focus and did not squarely address the low-dose issue. This was due, in large part, to the overall failure of the Organizing Committee to provide clear direction and focus, as evidenced by the variety of definitions for low-dose adopted by different subpanels, the varying scopes of review undertaken by the subpanels, and the undertaking by some of the subpanels to address issues that were well outside the proper scope

of the low-dose review. The peer review also did not follow the Organizing Committee's original guidelines by considering studies for which raw data were not provided.

The findings and recommendations of the peer reviewers were diminished in light of the shortcomings by the Organizing Committee – making the NTP low-dose peer review less useful to EPA than it might otherwise have been. Importantly, the shortcomings of the NTP peer review may inhibit future attempts to resolve scientific controversies. For example, it may be much more difficult in the future to get investigators to collate and submit raw data when they believe other investigators will be evaluated merely on the basis of their published papers or oral presentations at a peer review meeting. EPA must find a more rational, efficient and fair mechanism to resolve scientific controversies, such as the low-dose controversy. In that regard, EPA must ensure that peer reviews are focused and the peer review issues are well defined. EPA must also ensure that the ground rules for the peer review are set out in advance, and ensure adherence to those rules during the course of the meetings.

EPA should anticipate that there will be other instances when isolated, non-replicated findings will be used to challenge well-established toxicological tenets. The agency will want to encourage other laboratories to attempt to replicate such findings. The NTP low-dose peer review demonstrates a significant shortcoming that EPA would need to address, namely the weight and significance of small, limited non-GLP study studies. Therefore, to encourage research and to properly assess scientific evidence, it is imperative that EPA ensures that scientific evidence is appropriately weighed.

ACC recommends that before new controversies arise (and before EPA further examines the low-dose controversy), EPA develop a weight-of-the-evidence approach for assessing conflicting evidence and insist on its use. In that regard, EPA should develop clear and consistent guidelines for establishing causality for adverse effects and for weighing results within and across studies. EPA should also develop an overall policy for dealing with very limited findings that cannot be replicated.

VI. References

1. Ashby, J., H. Tinwell, et al. (1999). Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF1 mice exposed in utero. *Regulatory Toxicol. and Pharmacol.* 30:156-166.

2. Cagen, S.Z., J.M., Waechter, et al. (1999a). Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A. *Toxicol. Sci.* 50:36-44.

3. Cagen, S.Z., J.M., Waechter, et al. (1999b). Normal reproductive organ development in wistar rats exposed to bisphenol A in the drinking water. *Regulatory Toxicol. and Pharmacol.* 30:130-139.

4. Calabrese, E.J., Baldwin, L.A. (1998). A general classification of U-shaped dose-response relationship toxicology and their mechanistic foundations. *Hum. Exp. Toxicol.* 17(7):353-364.

5. Chhabra, R.S., Bucher, J.R., Haseman, J.K., Elwell, M.R., Kurtz, P.J., Carlton, B.D. (1993). Comparative carcinogenicity of polybrominated biphenyls with or without perinatal exposure in rats and mice. *Fundam. Appl. Toxicol.* 21(4):451-460.

6. Chhabra, R.S., Eustis, S., Haseman, J.K., Kurtz, P.J., Carlton, B.D. (1992). Comparative carcinogenicity of ethylene thiourea with or without perinatal exposure in rats and mice. *Fundam. Appl. Toxicol.* 18(3):405-417.

7. Chhabra, R.S., Bucher, J.R., Haseman, J.K., Elwell, M.R., Kurtz, P.J., Carlton, B.D. (1993). Comparative carcinogenicity of diphenylhydantoin with or without perinatal exposure in rats and mice. *Fundam. Appl. Toxicol.* 18(21):174-186.

8. Elswick, B.A., F., Welsch, et al. (2000). Effect of different sampling designs on outcome of endocrine disruptor studies. *Reprod. Toxicol. (in press).*

9. Ema, M. (2000). Two-generation reproduction study of bisphenol A in rats. *(Unpublished Study Report).*

10. Gray, L.E., Ostby, J., Furr, J., Price, M., Veeramachanemi, D.N.R., and Parks, L. (2000). Perinatal exposure to the phthalates DEHP, BBP, and DINP, but Not DEP, DMP, or DOTP, Alters Sexual Differentiation of the Male Rat. *Toxicol. Sci.* 58:350-365.

11. McIntyre, B.S., Barlow, N.J., Wallace, D.G., Maness, S.C., Gaido, K.W. and Foster, P.M.D. (2000). Effects of *in utero* exposure to linuron on androgen-dependent reproductive development in the male Crl:CD(SD)BR rat. *Toxicol. Appl. Pharmacol.* 167:87-99.

12. Nagel, S.C., F.S., vom Saal, et al. (1997). Relative binding affinity-serum modified access assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. *Environ. Health Persp.* 105(1):70-76.

13. O'Connor, J.C., Frame, S.R., Davis, L.G. and Cook, J.C. (1999). Detection of the Environmental Antiandrogen p,p.'-DEE in CD and Long-Evens Rats Using a Tier I Screening Battery and a Hershberger Assay. *Toxicol. Sci.* 51:44-53.

14. Tyl, R.W., C.B., Myers, et al. (2000). Two-generation reproductive toxicity evaluation of bisphenol A administered in the feed to CD (Sprague-Dawley) rats. *RTI Study No 65C-07036-000 (Draft Final Report).*

15. vom Saal, F.S., B.G., Timms, et al. (1997). Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. *Proceedings of the National Academy of Sciences* 94:2056-2061.

16. Waterman, S.J., Keller, L.H., Trimmer, G.W., Freeman, J.J., Nikiforov, A.I., Harris, S.B., Nicolich, M.J., and McKee, R.H. (2000). Two-generation reproduction study in rats given di-isononyl phthalate in the diet. *Reprod. Toxicol.* 14:21-36.

WSH\51828.1

From: Maria Bianchi
Sent: Monday, July 16, 2001 4:47 PM
To: liaison@starbase.niehs.nih.gov
Cc: bobf@regnet.com
Subject: APERC NTP Comments

Final NTP Low-dose Comments.doc

Please find attached comments being filed by Bob Fensterheim, Executive Director of the APE Research Council, on the NTP Endocrine Disruptors Low-Dose Peer Review Final Report. If you are unable to access the attached file, please contact me at the phone number or e-mail below. A hard copy has also been forwarded via mail.

Maria Bianchi
APE Research Council
Phone: (202) 637-9040
E-mail: regnet@ricochet.net

APE RESEARCH COUNCIL

1250 CONNECTICUT AVENUE, NW, SUITE 700
WASHINGTON, DC 20036

406-350 SPARKS STREET
OTTAWA, ONTARIO K1R 7S8

TOLL FREE: 866-APERC-NA WWW.APERC.ORG INFO@APERC.ORG

July 16, 2001

NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive
P.O. Box 12233
MD A3-01
Research Triangle Park, NC 27709
liaison@starbase.niehs.nih.gov

Re: Comments on the NTP Endocrine Disruptors Low-Dose Peer Review Final Report

Dear NTP Liaison:

The Alkylphenols & Ethoxylates Research Council (APERC) appreciates the opportunity to submit the following comments on the National Toxicology Program's (NTP) Endocrine Disruptors Low-Dose Peer Review Final Report. APERC members include manufacturers, processors, distributors and users of various alkylphenols, including nonylphenol and octylphenol, two of the specific compounds reviewed as part of the low-dose peer review effort.

APERC fully supported the proposed NTP low-dose peer review. It was our understanding that the primary purpose of the review was to consider the available evidence on the "shape of the dose-response curve for endocrine active substances in the low-dose region." Regrettably, the report that NTP prepared falls far short of this objective. Perhaps more disturbingly, the report fails to live up to the high standards of quality expected from NTP. As discussed in more detail below, APERC believes that an objective review of the available science does not support many of the conclusions and inferences contained in the report.

Much of the difficulty with the report's conclusions are related to the varying definitions of "low-dose" used throughout the report including, at times, the lack of adherence to any specific definition. For example, the Executive Summary states that for this peer review, low-dose effects refer to:

> *biological changes that occur in the range of human exposures or at doses that are lower than those typically used in the EPA's standard testing paradigm for evaluating reproductive and developmental toxicity. (Page i)*

The Executive Summary further notes that the *Other Environmental Estrogens and Estradiol* Subpanel "developed an operational definition for 'low-dose effects' that

was based on the dose-response data for the selected endpoints for each agent under evaluation. Low-dose effects were considered to be occurring when a non-monotonic dose-response resulted in significant effects below the presumed NOEL expected by the traditional testing paradigm." (page iv)

Regrettably, the Subpanel did not adhere to even its own modified definition and essentially considered any difference between treated and control animals occurring at a dose lower than the current NOAEL to be a "low-dose effect." No consideration was given to whether the change was an effect (adverse or not), nor was it considered whether the change represented a non-monotonic dose-response or simply a difference in an unusual endpoint not previously measured. For example, in the case of estradiol, the report suggests that low-dose phenomena occurred based on observations from studies of ovariectomized animals, which found changes in serum prolactin, LH, and FSH levels. However, there is no rationale for associating changes in circulating hormones as an important biological response under such an altered physiological state. Stated differently, the occurrence and significance of changes seen in prolactin, LH, and FSH have not been established in normal animal models. Further, any change in hormone concentration was considered biologically relevant without consideration for consistency with other hormone levels and/or responses or without consideration of whether the change was an increase or decrease.

Specific Comments Regarding Nonylphenol

Executive Summary is Misleading

The Executive Summary clearly states that nonylphenol has "low-dose effects" even though this was not the conclusion presented at the summary of the conference, and this conclusion is not supported by the positions espoused in the body of the report.

In summarizing the overall evaluation with regards to the "low-dose" issue, the Chair of Subpanel 2 indicated that, with regards to nonylphenol, there were "No effects seen." This view was not carried over into the Subpanel's written report, which describes presumed effects at so called low levels. At the same time, the report acknowledges that the significance of these effects are unknown and makes no mention of a non-monotonic dose-response, two criteria specifically stated in the definition of "low-dose effects." The report on page 51 states:

The "low-dose" of nonylphenol is questionable.

Unfortunately, this qualified viewpoint is completely missing from the Executive Summary of the report, which appears to have been written to imply that there are low-dose effects:

> ***Low-dose effects*** *[emphasis added] in F1 rats following dietary exposure to 25 ppm include a decrease in SDN-POA in males, an increase in relative thymus weight, an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3, and a prolonged estrus in females.*

APERC believes that at a minimum, the Executive Summary of the report should have reflected the viewpoints expressed in the body of the report.

The Findings Reported by NTP Do Not Justify a Conclusion of "Low-Dose Effects" For Nonylphenol

The conclusions regarding nonylphenol are stated to be based on observations principally concerning sexually dimorphic nuclei of the preoptic area (SDN-POA) in males, an increase in relative thymus weight, an increase in proliferation of splenic T-lymphocytes stimulated with anti-CD3, and a prolonged estrus in females. For the following reasons, APERC does not believe the reported observations provide sufficient evidence for such a conclusion:

1. **The changes reported are not scientifically defensible as adverse effects.** The principle measurements reported as "effects," SDN-POA and anti-CD3 effects, are not routinely evaluated in general or reproductive toxicity tests. Therefore, the biological/toxicological significance of these differences between treated and control animals is unknown. In fact, the Subpanel concluded, "As stated above, the SDN-POA changes are difficult to interpret at the moment, as are the anti-CD3 findings." It is also significant that changes in these endpoints resulting from exposure to endocrine active substances, including endogenous hormones, have not been validated as reliable indicators of endocrine activity, nor have these changes been reproduced in the same or other laboratories. The changes mentioned for thymus weight and prolonged estrus were not, to the best of our knowledge, reviewed by the Statistics Subpanel, thus failing to meet an NTP criterion for consideration in the Low-Dose Peer Review. Further, these changes are inconsistent with more robust studies evaluated by the Subpanel and the Statistics Subpanel.

2. **There is no evidence of a non-monotonic response.** There is no evidence that nonylphenol exhibits a non-monotonic dose-response in any of the studies reviewed for any of the endpoints noted.

3. **The purported changes should not be considered to be occurring at "low-doses."** Even if, for argument sake, the reported changes were considered to be biologically significant, the fact that they are reported to occur at 25 ppm means they should not be considered "low-dose effects."

4. **There is no specificity to the changes stated to be "low-dose effects" for nonylphenol.** Based on all of the studies reviewed, there is no specificity to changes in the endpoints. In addition to the stated results for nonylphenol, the changes in SDN-POA and anti-CD3 were observed for several other chemicals at all doses and for several chemicals that have different mechanisms of action. In all cases, there was no clear dose response (monotonic or non-monotonic).

The Additional Recommended Research Will Not Provide Meaningful\Information

The Subpanel concludes with the following recommendation, "More data are needed at the level of ng/kg/day, the human exposure level." Regardless of the significance or relevance of the changes discussed by the Subpanel, there are no data that would suggest such resource-intensive scientific investigations, including the use of experimental animals, are warranted. The lack of evidence for a non-monotonic dose-response for nonylphenol clearly indicates that dosing of nonylphenol in the cited dose range, equivalent to parts per trillion of nonylphenol in animal diets, would not provide meaningful information.

In general, APERC believes that the NTP report would have been significantly more useful and robust if there had been a constant focus on whether available data clearly demonstrated a non-monotonic dose-response in the low-dose region that is common, predictable, and relevant to humans. Instead, the report:

- Did not limit itself to evaluating effects at environmentally relevant exposures but considered effects at any dose, including doses higher than those typically used in standard assays.

- Focused on several observations that may have no relevance to human health, including observations that have no specificity for an endocrine mechanism.

- Drew conclusions from studies that were not designed to assess biological relevance to humans.

- Failed to reach the clear conclusion that nonylphenol does not cause low-dose effects.

In summary, APERC believes that an objective reading of the scientific conclusions for nonylphenol contained in the NTP report does not provide adequate evidence of a "low-dose effect" in accordance with the stated definition, namely, "Low-dose effects were considered to be occurring when a non-monotonic dose-response resulted in significant effects below the presumed NOEL expected by the traditional testing paradigm." Clearly, in the absence of any studies showing any effects in the

"low-dose" region, the conclusion of the Subpanel should have been that nonylphenol does not cause low-dose effects.

Please let me know if you have any questions on our comments. Thank you in advance for your consideration.

Sincerely,

Robert J. Fensterheim
Executive Director

From: Lynda J. Green
Sent: Monday, July 16, 2001 7:52 PM
To: liaison@starbase.niehs.nih.gov
Subject: NTP Endocrine Disruptors Low-Dose Peer Review Final Report

CSPA NTP Report Comments - Final.1.doc

Dear Sir/Madam:

In response to the Federal Register Notice on May 16, 2001, page 27152, please find the attached comments from CSPA on the "NTP Endocrine Disruptors Low-Dose Peer Review Final Report." If you have any questions, please feel free to contact me.

Lynda Green, M.P.H.
Assistant Director of Scientific Affairs
CSPA
202-872-8110

CSPA
Consumer Specialty Products Association

July 16, 2001

NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive
P.O. Box 12233
MD A3-01
Research Triangle Park, NC 27709

RE: NTP Endocrine Disruptors Low-Dose Peer Review Final Report, [Federal Register: May 16, 2001 (Vol. 66, No.95, Page 27152)]

Dear Sir/Madam:

This submission by the Consumer Specialty Products Association (CSPA, formerly known as CSMA) is in response to the above referenced notice announcing the opportunity for public comment on the NTP Endocrine Disruptors Low-Dose Peer Review Final Report. CSPA, headquartered in Washington, DC, represents over 200 companies engaged in the manufacture, formulation, distribution and sale of aerosol products, scented candles and air fresheners, disinfectants and sanitizers, automotive products, detergents and cleaning products, home, lawn and garden pesticides, and polishes and floor finishes for home, institutional and industrial use.

The Association has previously provided comments on the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) report that was submitted to the EPA Science Advisory Board (SAB) and FIFRA Scientific Advisory Panel (SAP) for a special consultation. CSPA has supported the EDSTAC process and believes that the group made significant progress on a very complex topic. The release of the NTP Endocrine Disruptors Low-Dose Peer Review Final Report, however, has raised the following concerns within the Association:

1. The various sub-panels of the Organizing Committee did not adhere to NTP's initial definition of "low-dose," and were not consistent in their definition. Prior to the meeting, NTP defined low-dose as those doses that were below the NOAEL for a compound. During and subsequent to the meeting, NTP expanded the working definition to include biological changes that occur at environmentally relevant exposure levels or at doses that are lower than those typically used in the U.S. EPA toxicity testing. This new definition is not relevant if one is attempting to define the shape of a dose-response curve and determine if there are low-dose effects with a given compound. Traditional toxicological practice assumes a threshold with a NOAEL below and LOAEL above, and increasing response with increasing dose above the threshold.

Non-cancer studies have at least one dose in a range to result in a NOAEL. These differences include taking the LOAEL and dividing by a 10x safety factor or defining low-dose as below accepted NOAEL, where there would be no datapoints available to review.

The importance of these matters is that if NTP begins to interpret the breadth or range of standard dosing regimens currently employed for toxicology testing, this would re-define how preceding and subsequent toxicity testing is interpreted or performed. We believe this was not the intent of NTP, but given the current scenarios, this could lead to the conclusion that current toxicology testing schemes are insufficient (i.e., biological effects are detected below the NOAEL or additional endpoints are required to detect adverse effects, both with which we disagree). This is not the case, and would require re-testing or additional testing.

2. We have several concerns relative to the Final Report's statements regarding the limitation of currently accepted toxicology guideline studies. The toxicology study guidelines that have been instituted are performed under rigorous GLP guidelines and have undergone extensive peer-review and validation by numerous international scientific organizations. If EPA believes that these guideline studies should be revised, we strongly urge EPA to follow an established protocol for including comments from international associations, such as the OECD harmonization principles.

3. The Final Report has overstated the evidence in support of low-dose evidence and effects. In particular, the sub-panels relied heavily on a weak study performed in 1997 by vom Saal *et al,* related to bisphenol A (BPA). We consider the study weak because the statistical power was poor in terms of animal numbers, the study did not account for litter effects, the material and dosing solutions were not analyzed, the study only used prostate weights as an end point, and vom Saal did not provide the raw data as requested in all cases. Subsequent to the vom Saal study in 1997, a comprehensive three-generation reproduction study on BPA demonstrated no low-dose effects on multiple sensitive endpoints. This study is more compelling than the non-reproducible study results of vom Saal et al. in that the three generation study has significant statistical power over any other previous study and has numerous relevant endpoints indicating no low dose effects.

   There also are questions regarding interpretation of the data in the vom Saal study on BPA. What was the molecular and biological plausibility of using the prostate weight as an end point? This was never broached or discussed as part of weighing the evidence in the Final Report. NTP did not fully use a weight of the evidence approach since there are several other very strong studies on which NTP could evaluate this effect. If these had truly been evaluated and incorporated into the discussion, the conclusion would have been that there is no credible evidence for the low-dose effect, and that it has not been proven by the current toxicology principles.

CSPA appreciates this opportunity to comment on the Final Report and looks forward to participation in future endocrine disruptor issues. We sincerely hope that NTP will incorporate all comments received since the Final Report does not accurately reflect the outcomes from the October meeting.

If you have any questions or comments please do not hesitate to contact me at (202) 872-8110.

Sincerely,

Lynda Green, M.P.H.
Assistant Director of Scientific Affairs

David Crawford, Ph.D.
Endocrine Issues Task Force, Chairman

American Water Works Association
Government Affairs Office

Dedicated to Safe Drinking Water

1401 New York Avenue, NW
Suite 640, Washington, DC 20005
http:www.awwa.org
(202) 628-8303 Fax (202) 628-2846

July 16, 2001

NTP Office of Liaison and Scientific Review
111 T.W. Alexander Drive
PO Box 12233
MD A3-01
Research Triangle Park, NC 27709

Re: Comments on the NTP Report of the Endocrine Disruptors Low-Dose Peer Review

Gentlemen:

Enclosed are the comments by the American Water Works Association on the NTP Report of the Endocrine Disruptors Low-Dose Peer Review. Dr. Crawford-Brown of the University of North Carolina aided in the preparation of these comments. Dr. Crawford-Brown represented water utilities on the original Endocrine Disruptors Screening and Testing Advisory Committee (EDSTAC) and has assisted AWWA in developing comments on many other complex issues.

The American Water Works Association (AWWA) is an international non-profit, scientific and educational society dedicated to the improvement of drinking water quality and supply. Founded in 1881, the Association is the largest organization of water supply professionals in the world. AWWA's over 56,000 members represent the full spectrum of the "drinking water community": treatment plant operators and managers, public health officials, scientists, academicians, and others who hold a genuine interest in water supply and public health. Our membership includes over 4,000 utilities that supply roughly 75 percent of the nation's drinking water.

If you have any questions on these comments, please feel free to call Jeanne Bailey or me at our Washington Office.

Yours Sincerely,

Thomas W. Curtis

Thomas W. Curtis
Deputy Executive Director

Enclosures

Headquarters Office:
6666 W. Quincy Avenue, Denver, CO 80235
(303) 794-7711 Fax (303) 795-1440

cc: Jeanette Wiltse—USEPA OST
Rob Renner
Ed Baruth
Alan Roberson
Steve Via
Jeanne Bailey
Clare Haas Claveau
Mark Scharfenaker

P:\jeanne\ntp71601edpeerreivewcover

COMMENTS BY THE AMERICAN WATER WORKS ASSOCIATION ON THE REVIEW OF THE NATIONAL TOXICOLOGY PROGRAM'S "ENDOCRINE DISRUPTORS LOW DOSE PEER REVIEW"

Introduction

Under statutory mandates in both the 1996 Amendments to the Food Quality Protection Act (FQPA) and the Safe Drinking Water Act (SDWA), the Environmental Protection Agency (EPA) formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC). The purpose of EDSTAC was to advise the Agency on the design of methods to identify endocrine disruptors and quantify their potential to cause adverse effects. A screen is an assay that simply indicates the potential to cause effect. It is designed to err on the side of precaution; i.e. the screen is significantly more likely to falsely show a non-endocrine disruptor to be a disruptor than it is to show a disruptor to be a non-disruptor. If a substance is shown to have potential to cause effect, it is subjected to further testing. A test is designed to more definitely determine if the substance truly is an endocrine disruptor (hereafter, ED), to define the dose-response characteristics, and to provide insights into the mechanism of action.

Use of both screens and tests requires identifying the doses to be used. Since most effects are relatively rare, particularly at low doses, it is necessary to use a large number of animals, to use high doses (increasing the probability of effect per animal), or some combination of these factors if effects are to be seen. Since increasing the size of the population can be prohibitively expensive, studies usually employ high doses. All substances, however, if given in sufficiently large doses, produce toxic effects simply due to the mass of substance applied (i.e. the effect is not necessarily due to the chemical properties, but rather to the sheer mass applied). As a result, testers usually identify a maximum dose that can be applied without these acute toxic effects (the maximum tolerated dose), and then progressively lower that dose in three or four stages to study the effects at lower doses. These lower doses are some fraction of this maximally tolerated dose, but are still high enough to provide statistically reliable estimates of effect (if these effects occur) in a reasonably small population of animals.

The assumption of such tests is that dose-response curves are monotonic (i.e. the level of effect increases always as the dose is increased). If that assumption is true, then a No-Observed-Effects-Level (NOEL) identified in a test using relatively high doses will be a valid approximation of any threshold dose. In other words, if effects fail to appear at this NOEL value, they also will fail to appear at all doses below this value. This is a mathematical consequence of a monotonic dose-response relationship. Some members of the EDSTAC, however, expressed concern that dose-response curves for EDs might not be monotonic. They pointed out that, unlike other systems of the body, the endocrine system operates at remarkably low concentrations of hormones. They also pointed out increasing evidence that dose-response curves for some substances might not be monotonic, but rather U-shaped or inverse U-shaped. In a classical U-shaped curve, such

as might be found in hormesis, the effect at low doses is greatly over-predicted by linear extrapolation of effects from high doses, and the substance can even produce a reduction in effect (i.e. be beneficial) at low doses. In an inverted U-shaped curve, there can be a region of low dose that produces a greater effect than at higher doses. It is even possible that the effect appears at low doses, disappears at intermediate doses, and then re-appears (perhaps as a different effect) at high doses. It is this latter possibility that concerned the members of EDSTAC mentioned above.

The EPA has taken this concern seriously, with good reasons. There is increasing scientific evidence that dose-response curves may indeed be non-monotonic, including both the U-shaped and inverse U-shaped curves. If these exist, they have significant regulatory and testing implications. In the case of U-shaped curves, risk reduction would be obtained only by reducing the doses down to some threshold level. Further reductions would not only fail to produce a benefit, but would actually introduce a health detriment. In the case of inverse U-shaped curves, or more complex curves in which effects appear at low and high doses, but not at intermediate doses, the traditional approach to selecting doses in a screen or test might lead to results where the lowest dose employed in the study was in the intermediate dose range, completely missing the effects at low doses and falsely identifying a NOEL.

It is important to understand why some members of the EDSTAC felt that an additional program in screening and testing for EDs was not needed. They argued that there already exist numerous screens and tests for reproductive, cancer and developmental health effects (the effects expected from EDs). While these screens and tests were not specifically designed to identify EDs, the adverse health effects they measure are precisely those that an ED might produce. They argued further that these screens and tests would already have indicated if a substance produced adverse effects, and the only information that would be provided by an ED program would be further elucidation of the mechanism of action. But since mechanism of action does not drive identification of any regulatory quantity such as a Reference Dose (RfD), this information would not be needed. All that is needed is a NOEL or Lowest Observed Effects Level (LOEL) and, they argued, this can already be obtained through existing testing methods.

The counter argument raised at the EDSTAC meetings was that these existing screens and tests, while they might measure the same health outcome, used doses in the intermediate to high dose range. Such a protocol might be valid for the mechanisms that were assumed to operate when the screens and tests were developed, but might not be valid for EDs since the endocrine system operates at concentrations well below those of other systems and mechanisms. In addition, it was argued that the endocrine system passes through "windows" of sensitivity in time as an organism develops, and that the existing screens and tests were not designed with these windows in mind. From this perspective, screens and tests for EDs might be significantly different from those for other effects, at least in the sense of requiring a much larger range of doses and greater attention to specific periods of development.

At the close of the EDSTAC process, the available science could not resolve this issue fully. The EPA was encouraged to examine the issue through a larger review of the information, and to determine whether that science indicated a need to take special precautions in the selection of doses and times of exposure for ED screens and tests. As a result, the Agency requested the National Toxicology Program (NTP) to consider the issue and advise them on the proper selection of doses. The result was a peer-review process sponsored jointly by the EPA, NTP and National Institute of Environmental Health Science (NIEHS). The goal was to "establish a sound scientific foundation upon which the U.S. EPA could determine what aspects, if any, of its standard guidelines for reproductive and developmental toxicity testing need to be modified to detect and characterize low dose-effects of endocrine disruptors" (see NTP 2001; page i). For the review, low dose was defined as "the range of human exposures or...doses lower than those typically used in the EPA's standard testing paradigm for evaluating reproductive and developmental effects" (see NTP, 2001; page i). Results of the review are summarized in the Endocrine Disruptors Low Dose Peer Review Report (NTP, 2001).

Comments on the Review Process

At the request of the EPA, the National Toxicology Program (NTP)/National Institute of Environmental Health Sciences (NIEHS) conducted a scientific peer review to evaluate reported low-dose reproductive and developmental effects and dose-response relationships for endocrine disrupting chemicals. AWWA believes that the panel used a very solid process for performing the ED Low Dose Review, at least with respect to collecting and analyzing the primary data on health effects. Rather than relying solely on statistical analyses performed by the original researchers (which would have created difficulty in drawing coherent conclusions given the range of analytic methods employed by those researchers), the panel elected to obtain the raw data and perform a consistent statistical analysis on all sets of data. These analyses were performed by a Statistics Subpanel consisting of individuals with a strong background in the application of statistical methods to the analysis of animal and *in vitro* data. A total of 59 studies were identified as relevant to this low-dose issue; the studies are listed in the NTP report, and include a reasonable range of studies. The authors of the studies were contacted and asked to submit the raw data along with a completed questionairre concerning issues such as dosing regime, treatment of controls, caging methods, feed methods, etc. Of the 59 studies identified, data and questionnaires were supplied by 49 authors; the other 10 could not due so due to constraints on time and/or resources, or due to a lack of electronic formatting of the data. From these 49 studies, 39 studies were identified as particularly germane and the data re-analyzed. These data covered a reasonable range of substances, endpoints and dosing periods (i.e. different ages or developmental periods).

Five other subpanels were formed. Three focused on specific classes of substances: one on Bisphenol A, representing a well-studied substance; one on Other Environmental Estrogens and Estradiol; and one on Androgens and Antiandrogens. Each of these subpanels reviewed the literature on the substance or class of substances within their purview, and utilized the re-analyzed data from the Statistics subpanel. A fourth subpanel on Biological Factors and Study Design considered the role of species, developmental

window, dosing regime, etc, on the expected health outcome. Their focus was on understanding the biological processes involved in ED effects, and how this biology might inform the design of effective testing protocols. They also considered how well previous studies (specifically the 49 used in the peer review) conformed with these biological principles, and the implications of any deviations. The three subpanels on specific substances used the guidance provided by this fourth group to decide whether effects seen in a study might be due to problems with various biological factors in the study design. A fifth subpanel on Dose-Response Modeling considered the implications of alternative dose-response models on interpretation of existing studies, on the design of future studies, and on the issue of whether low-dose effects might be missed under current dosing protocols. In a sense, the three subpanels on Statistics, Biological Factors and Study Design, and Dose-Response Modeling acted in support of the three subpanels considering specific substances or classes of substances, since the analyses of substances/classes constituted the case studies used to address the original issue. They also provided additional summaries that might be used in the design of future studies, including a statistical methodology that might be applied uniformly in the design and interpretation of such future studies.

Specific Comments

1. The panel did not consider the issue of whether an effect observed in a study is adverse. It was decided that this issue cannot be resolved fully at present given the existing state of the science. The report explicitly stated (see page vii of the Executive Summary) that the toxicological significance of the findings is not known. It is possible, therefore, that some (perhaps all) of the effects reported in the 49 studies might not constitute adverse health effects in the sense in which this term is used in regulatory decisions. This is in part disappointing, since the larger science policy issue is not simply whether existing dosing protocols will miss some effects, but whether they will miss effects that are adverse and might drive a regulatory decision. Despite this limitation, the review does provide a valuable first step. Before one can determine whether *adverse* effects might be missed, it is necessary to determine whether *any* effects at all are missed (adverse or not). The review process at least identified that some effects might be missed, opening the way to a further review as to whether such effects also are adverse.

AWWA Response:
The AWWA encourages the Agency to take the analysis of the Peer Review one step further and determine whether current testing methods are missing adverse effects, and not simply effects, at low doses. The EDSTAC clearly considered it important to distinguish between effects and adverse effects, particularly since the goal is protection of public health and welfare.

2. The conclusion of the panel is that effects have been shown in some studies of Bisphenol A, Other Environmental Estrogens and Estradiol, and perhaps Androgens and Antiandrogens at doses near the NOELs as established by existing dosing protocols. These effects, however, have not been fully reproducible, appear to vary significantly

across species and even strains, and result from mechanisms that are not yet identified. There are several possible explanations for such findings:

- The positive studies may simply be random excursions from a mean of no effect. The panel reviewed data from 39 studies, many of which contained measures of effect on more than one endpoint. Given this large number of measurements, a 95% confidence criterion on each individual study is likely to be exceeded several times throughout the set of such studies, even absent a true effect in any one study. The panel analyzed confidence intervals for each individual study, and classified an effect as positive if the excess incidence satisfied the confidence criterion. A follow-up step would be to determine whether the overall body of data examined (i.e. the 39 studies) displays a statistically significant elevation in the number of positive studies expected purely by chance.

- There may truly be an effect that varies with substance tested, species/strain and dosing conditions. If this is the case for EDs, the implication is that it may be necessary to perform significantly more research to identify model species/strains, as well as exposure windows (i.e. sensitive times for exposure during development of an organism), before reliable ED testing can be conducted. This is one of the conclusions of the Biological Factors subpanel.

- The variability of results may indicate problems with inter-laboratory comparability in dosing, controls, animal care, etc. The doses being delivered are small, and the effects measured are slightly above control levels. The study methodologies are pushing the limits of current measurement methods. As greater uniformity in study conduct occurs, and as researchers gain more experience in conducting studies at lower doses, the variability may disappear.

Overall, the panel did a good job of collecting and analyzing the existing data, and of providing a statistical summary. They have correctly hedged their conclusions by noting that the effects have been variable across labs, study protocols, species/strains, etc, and that this complicates interpretation of the results. They also have a reasonable basis for concluding that the data, when taken overall (rather than simply study-by-study), indicate a possibility that effects are being produced at doses near the current NOELs. Still, to raise this conclusion above a "possibility", it will be necessary to determine first what combination of the three explanations provided above is in effect.

AWWA Response:
The AWWA encourages the Agency to take the next step in determining whether existing data can allow selection of one, or a combination, of the three explanations above.

3. Existing modeling efforts clearly cannot resolve the issue of low dose effects. As shown by the Modeling subpanel, current scientifically valid models include linear, threshold and even non-monotonic functions (both U-shaped and inverse U-shaped). The differences between these functions lie largely in differences in parameter values (such as

binding efficiency, number of receptors and Michaelis-Menten kinetics), and these parameter values simply are unknown for essentially all of the studies examined by the other subpanels. As a result, the Modeling subpanel could provide little guidance to the other subpanels, other than to point to the possibility that any of the above dose-response functions might apply.

There is, in fact, a curious disjuncture between the empirical and modeling sides of the panel review. The modeling appears to be almost an afterthought, given in the abstract with no specific applications to the data used in the review. This might simply reflect the existing state of the science; i.e. the modeling is not yet at a point to allow significant use of models in interpreting data and allowing extrapolation. It also might reflect the intent of the peer review, which was to determine whether there are data indicating effects at low doses, rather than to consider models explaining those data or facilitating extrapolation. Still, the review would have been more compelling and complete had the link between data analysis and modeling been stronger. This could have included, for example, an attempt to determine if the studies examined, as well as the larger body of more fundamental biological, biochemical and toxicological data on these substances from which model parameters might be estimated, allow a significant ability to provide parameters that would narrow the range of dose-response shapes to be considered by the other subpanels.

This point was underscored by the Biological Factors and Study Design subpanel, and is mentioned also in the Executive Summary. There is a clear call for collecting basic pharmacokinetic data in future studies so models can be better formulated. At the least, the data should provide insight into which of the general model forms might be most biologically realistic for a substance, which then could guide attempts at extrapolation.

AWWA Response:
The AWWA encourages the Agency to further consider the issue of modeling in support of assessing low dose ED effects. The role of modeling was insufficient in the ED Low Dose Peer Review process and should be improved. The relative biological support for different model formulations should be assessed for specific compounds, and this should temper interpretation of the effects data (perhaps in a Bayesian framework). Testing programs, designed to test alternative hypotheses underlying models, should be developed. At present, there is too large of a separtion between modeling and empirical studies. These two activities must be better integrated.

4. Response bias (in which respondents to a request have a significantly different characteristic than do non-respondents) is always of concern in soliciting and then analyzing study data. Poor response rates may cause an unreliable conclusion to be drawn. For this peer review, however, the response rate was high (49 of 59 studies). In addition, the inability to retrieve 10 sets of data appears to have been related to issues of resources, rather than to any factors that would bias conclusions. It is unlikely, therefore, that response bias played a role in the current review.

AWWA Response:

The AWWA congratulates the Agency, the NTP, and the NIEHS for developing this process of data collection and for following through to such a high response rate. This significantly improves the quality of the final conclusions, and offers a model for how future reviews in a wide range of scientific areas might be conducted.

5. The panel concluded that traditional multi-generation reproductive studies have yet to show a significant ED effect at low doses (i.e. near or below existing NOELs). Such studies provide the most direct analysis of toxicologically significant effects in whole organisms exposed over significant periods of time. The lack of adverse effects in these studies lends at least some support to the claim that the effects found in other kinds of studies (i.e. non-multi-generational) may not be adverse to at least reproduction and development. The panel, however, correctly notes several limitations to such a conclusion: (i) the multi-generation test does not follow the F2 generation past day 21 of development, and so may miss effects that only appear in that generation at later times (or perhaps in higher generations); (ii) the test does not in general examine cancer effects; (iii) the test does not employ a wide range of doses, and may have simply missed a dose point of relevance; and (iv) some known endocrine active substances have produced negative results in these tests, despite the expectation that they should produce positive results (and do produce positive results in many other kinds of studies). Limitation (iv) may indicate either a problem with the multi-generation study, or may indicate that the effects found in other study methods are not adverse effects in the sense of producing discernible reproductive or developmental changes (since compensatory mechanisms in the body allow it to avoid adverse effects at these levels despite the alteration of some functional parameters).

AWWA Response:
The AWWA encourages the Agency to consider whether the multi-generational test must be modified to account for effects appearing in F2 after the 21st day (by which time the reproductive tract is not fully developed), and to account for especially sensitive windows of exposure in F0, F1 and F2.

6. Overall, the review panel selected a reasonable sample of studies on which to base their analyses. The sample includes a good range of substances, endpoints and dosing periods, and includes the major studies likely to form the basis of any scientific analyses in the future. The authors of the studies, for the most part, have established track records in performing these kinds of studies. The choice not to include dioxin and dioxin-like compounds, as well as phthalate esters, was not based on any scientific reasoning, but rather on the fact that other reviews were being conducted for these compounds. It might have been best to include these in the present review so as to not "skew the sample", and also to determine whether this design for the review process produced results for these compounds that are consistent with the more extensive reviews being conducted. At the same time, there is little reason to suspect that the overall conclusions would have changed had these compounds been included. In fact, the conclusions probably would simply have been underscored.

AWWA Response:

The AWWA believes the Peer Review process utilized a sufficient range of compounds to address the question of low dose ED effects, although it is noted that inclusion of dioxin and the phthalate esters might have provided a benchmarking exercise for the review process itself.

7. The document does not mention how the authors of the original studies will play a role in the final review of the report. The authors were asked to complete questionnaires and to provide data, but will they also be asked to review the document to ensure their data are being interpreted properly? That would be an important component of the review process. They certainly have the opportunity to do so as part of the public review and comment period, but they also should be approached directly for comments. This may already be part of the planned review, but is not made clear in the document.

AWWA Response:
The AWWA encourages the Agency to formally elicit review comments from the study authors, and to determine any reasons for differences in opinion between the authors and the Peer Review subpanels.

8. The questions addressed by the subpanels on specific classes of compounds wisely include both positive and negative conclusions. In other words, the subpanels were asked not only to assess the strength of a claim that the substances DOES cause effects at low doses, but also the strength of the claim that it does NOT. This was a good choice of analytic framework, since it forces the panel members to consider both sides of the claim, rather than simply building support for one side or the other. The result is a more balanced assessment of the evidence than would otherwise have taken place.

AWWA Response:
The AWWA encourages the EPA to continue this approach to assessing evidence in the future, both with respect to potential ED substances and in other realms of science-based decisions. Final judgments as to whether effects occur should be based on consideration of three possibilities: the substance does cause an effect (and the evidence provides support for that claim), the substance does not cause an effect (and the evidence provides support for that claim), and the data are insufficient to form any reasonable judgment. The AWWA recognizes that the Agency already does this informally, but the design of the ED Low Dose Peer Review Process made this reflection more formal.

9. There is some confusion left by the document with respect to the issue of interpreting the meaning of "low dose". As mentioned by the BPA subpanel, low dose was defined with respect to intake rate regardless of route of exposure (see page 1-3). The subpanel also notes correctly that target tissue dose can differ significantly when the same intake rate is applied by different routes. The result is that "low dose" means different target doses for different routes of exposure in the peer review. This complicates interpretation.

The reader would be helped by some indication as to whether positive effects at "low doses" might have been restricted to some subset of the data in which "low dose" corresponds to the high end of the target dose range. For example, suppose the positive

effects were seen only in routes of exposure where the ratio of target dose to applied dose is significantly larger than for other routes, and suppose these "other routes" are the ones encountered in the environment. The conclusion that effects were seen at "low doses" might then be an artifact of the inability to use target dose as the dose measure. There is some hint that this might be occurring in the Summary statement on page 1-6.

AWWA Response:
The AWWA encourages the Agency to always consider target dose, rather than applied dose, in determining whether low dose effects occur. At the least, all routes of exposure should be placed on a common basis through scaling factors that account for pharmacokinetic differences between routes of exposure. Only in this way can a proper assessment be made as to whether "low dose" exposures in the environment (where only a subset of the exposure routes is present) are expected to produce effects.

10. The conclusions from the BPA subpanel are highly caveated, and it is not clear how they will be interpreted by the Agency. The subpanel is saying that there are limited studies showing low dose effects, but there also are more plentiful studies showing no such effect. They cannot rule out the possibility that the differences are due solely to study design, but there does seem to be a problem with replicability of the studies showing positive results. This will represent an interesting case of weight of evidence determination for the Agency, and there is a clear need for the Agency to define how it will consider weight of evidence in applying the lessons from the NTP review.

AWWA Response:
The AWWA encourages the Agency to continue along its current path of applying weight of evidence determinations to evidence, rather than adopting the position that a single positive study indicates an effect. The ED Low Dose Peer Review shows clearly that conclusions can be significantly less secure when the full range of data are considered, especially in cases like BPA where the data supporting a claim of low dose effects are largely outweighed by data showing no effect. Given the variability of results provided by different studies, it is important that results of a study be replicated, especially if they have significant regulatory implications (see Comment 2 on page 2-5 of the Other Environmental Estrogens and Estradiol subpanel).

11. Figure 1 on page 2-7 is impossible to interpret, given the lack of explanation in the document. There is some mention of it in the Introduction by that subpanel, but then it is never used. The figure itself has no substantive heading, so it cannot be read on its own terms. It mentions observations in the heading, but then mentions simulation in the footing. The reader is left wondering what this figure represents, how it guided the thinking of the subpanel (if at all), etc. The figure should be given a better explanation in the text, and should be provided a heading or footing giving full detail as to how it was generated and how it should be interpreted.

12. As noted by the Androgens and Antiandrogens subpanel, there remains some confusion in the document as to how "low dose" should be defined (see page 3-2). All of the subpanels seem to have adopted a definition related to doses at or near the NOEL or

LOEL. This is a bit unfortunate, since the goal was to determine whether existing testing methods will miss ED effects that occur only at low doses. By definition, a LOEL is already a dose that has produced effects even in the existing methods. So, if ED effects are found at or near these doses, this would not support the claim that new methods are needed. Even if ED effects appear "at or near" the NOEL from existing methods, this would not necessarily mean new methods are needed for ED effects. It has always been assumed in regulatory science that effects could occur slightly below the NOEL if due to nothing more than study statistics. This is one of the reasons uncertainty factors are applied.

AWWA Response:
The AWWA encourages the Agency to develop a better working definition of "low dose", rooted in the specific decisions that must be made in regulatory affairs.

13. The EDSTAC spent a considerable amount of time discussing how to determine whether an effect truly is an ED effect. Most of the endpoints measured in the studies considered by the ED Low Dose Peer Review Panel can be produced by mechanisms other than endocrine disruption. If the effect occurs at low doses, therefore, this is not direct evidence that ED effects are occurring at these doses, since the primary mechanism of action may not be endocrine disruption (but rather, for example, cytotoxicity, proliferative response, cell cycle delay, etc). This remains a significant source of confusion in identifying an ED effect. It is necessary to obtain data at doses below which the other (non ED) mechanisms should not be operating. The data available to the Peer Review panels do not appear to have allowed this kind of analysis. This strengthens the comment (see item 12 above) that it still is not clear whether existing testing methods, rooted in these non-ED mechanisms, will falsely identify a NOEL and fail to capture ED effects occurring at lower doses (due, for example, to an inverse U-shaped dose-response curve).

AWWA Response:
The AWWA encourages the Agency to conduct research on the relevant biological endpoints at doses below those at which the other mechanisms may be expected to act. The whole animal studies examining reproductive and developmental outcomes are not specific to ED effects, and so it is difficult to determine whether there is any case at present for the claim that ED effects appear at or below the NOEL. It might be possible to examine the whole animal studies to determine whether the effects seem to be a strong function of temporal windows, and to argue that only ED mechanisms can account for these windows, but this has not been done to date.

14. The conditions identified by the Biological Factors and Study Design subpanel as being relevant to design and interpretation of studies represent a good summary of the state of the science. At the same time, these already constitute good laboratory practice, and so it is not clear whether anything new has been added by the report. Perhaps the report could focus on bit more on how the guidelines of good laboratory practice should be assessed when dealing with the special case of ED screens and tests.

15. Finally, it is noted that a significant amount of effort went into collecting, compiling, abstracting and analyzing the data reviewed by the subpanels. It is important that this effort not be lost. The data could be placed into an electronic format and made available to the scientific community over the internet. It would be useful to establish it as a baseline to which future ED studies could be added. This requires that someone, or some group, be assigned the task of maintaining and updating the database regularly.

AWWA Response:
The AWWA encourages the Agency to build on this peer review effort by making this data collection and analysis effort publicly available. The first step has been taken to defining a common format for study design, data collection, data abstraction, etc. Given the finding of the subpanels that even slight differences in study design and data analysis can lead to significant differences in conclusions, standardization of the process of assessment is needed.

Conclusion:
The Peer Review subpanels did an excellent job of analyzing the data available, and drawing scientific conclusions as to whether ED effects might be expected at or near the NOEL/LOEL. There is no strong evidence in the report to suggest that effects will appear below the current NOELs. There continues to be evidence of effects near the NOELs, and in response, the scientific community has applied uncertainty factors. AWWA believes that current testing protocols are sufficient and does not believe that this report justifies changes. AWWA believes that the Agency should undertake additional research to determine whether the doses at which ED effects begin to occur might be significantly below the existing NOEL so as to justify changing the screening protocols or increasing uncertainty factors.