Natural Variability and the Influence of Concurrent Control Values on the Detection and Interpretation of Low-Dose or Weak Endocrine Toxicities

John Ashby, Helen Tinwell, Jenny Odum, and Paul Lefevre

Syngenta Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire, United Kingdom

While defining the no effect level for the 5α-reductase inhibitor finasteride in the Hershberger assay, we encountered an inverted-U low-dose trophic effect on the prostate gland of the rat. Two attempts to confirm this observation were unsuccessful, and we concluded that the positive effect initially observed was associated with normal biologic variability. During the same period we attempted, unsuccessfully, to repeat our own observation of weak uterotrophic activity in the rat for the sunscreen 3-(4-methylbenzylidene)camphor (4MBC). Further evaluation led us to conclude that 4MBC is uterotrophic only when the control uterine weights are at the low end of their normally encountered range. This led us to reevaluate our earlier mouse uterotrophic assay data for bisphenol A (BPA). Originally we had concluded that BPA gave irreproducible evidence of weak uterotrophic activity, but upon ordering the eight experiments we had conducted, according to decreasing control uterine weight, we confirmed reproducible weak uterotrophic activity for BPA when the control uteri were at the low end of their normal range. In this article, we describe these observations, together with a reanalysis of the data associated with several reported instances of weak or low-dose endocrine effects that have proven difficult to confirm in independent laboratories. These include the activity of BPA on the CF1 mouse prostate; the activities of BPA, octylphenol, and nonylphenol on the rat testis; and the effect of polycarbonate caging on control mouse uterine weight. In all of these cases, variability among controls provides a major obstacle to data interpretation and confirmation. Our recommendations on experimental design are also presented, with a view to ending the current impasse on the reality, or otherwise, of low-dose or weak endocrine toxicities. *Key words:* 3-(4-methylbenzylidine)camphor, bisphenol A, endocrine disruption, finasteride, low dose, nonylphenol. *Environ Health Perspect* 112:847–853 (2004). doi:10.1289/ehp.6862 available via *http://dx.doi.org/* [Online 18 February 2004]

Two issues in the field of chemically mediated endocrine disruption (ED) are the possible occurrence of ED effects at dose levels lower than the no effect level defined by classical toxicity evaluations, and how to interpret weak chemically induced ED effects that lie within the range of normally encountered control variability (Ashby 2003). We have studied several reports of low-dose ED effects over the past 6 years, so far without being able to confirm any of them; these chemicals include butylbenzylphthalate (BBP) (Ashby et al. 1997; Sharpe et al. 1995, 1998); bisphenol A (BPA) (Ashby et al. 1999, 2003; Nagel et al. 1997; Sakaue et al. 2001; Talsness et al. 2000; Tinwell et al. 2002a); and nonylphenol (NP) (Colerangle and Roy 1996; Lee 1998; Odum and Ashby 2000; Odum et al. 1999). In contrast, we were able to confirm (Tinwell et al. 2002b) the weak uterotrophic activity of 3-(4-methylbenzylidene)camphor (4MBC) reported by Schlumpf et al. (2001a, 2001b), despite the inability of Bolt et al. (2001) to confirm this activity.

Recently, we encountered two instances that may throw light on the origins of these reproducibility problems. The first of these was the observation of an apparent inverted-U low-dose effect for the 5α-testosterone reductase inhibitor finasteride (FIN) in the Hershberger antiandrogen assay, an effect we were unable to reproduce. The second was our inability (after publication of Tinwell et al. 2002b) to reconfirm our own observation of uterotrophic activity for 4MBC. In each case, further studies revealed that control variability between experiments was probably at the root of these reproducibility problems. This conclusion led us to reevaluate earlier findings for BPA in the mouse uterotrophic assay (Tinwell et al. 2000) and to analyze earlier reports of low-dose effects for BPA (Howdeshell et al. 2003; Nagel et al. 1997; Nonneman et al. 1992; Sakaue et al. 2001) and NP (Lee 1998).

Materials and Methods

Chemicals. We obtained testosterone propionate (TP) and FIN from Sigma Chemical Company (Poole, Dorset, UK). Tocopherol-stripped corn oil (ICN Biomedicals Inc., Basingstoke, Hampshire, UK) was used to suspend FIN and TP by homogenization (Ashby 1987). We obtained 4MBC (CAS Registry No. 38102-62-4; Merck batch TT805785 029) from ChemQuest (Wilmslow, Cheshire, UK) and diethylstilbestrol (DES) from Sigma Chemical Company. Antarelix (ANT) was a gift from Europeptides (Argenteuil, France). We used arachis oil (AO; Sigma Chemical Company) as the vehicle for 4MBC (homogenized as described previously; Tinwell et al. 2002b), DES, and ANT. All chemicals were stored at room temperature except ANT, which was stored at 4°C. All preparations were

shaken vigorously just before and during dosing. All experiments were conducted according to U.K. Home Office guidelines as described in the appropriate Home Office license for animal use.

Hershberger assays. **Animals.** Male Alpk:APfSD rats (animal breeding unit at AstraZeneca Pharmaceuticals, Macclesfield, Cheshire, UK) were castrated at 6 weeks of age via midline incision followed by 8 days of recovery. In each study, the animals were randomized into groups of six rats of approximately the same initial mean body weight and range. They were group housed at a maximum of six per cage in solid-bottomed polypropylene cages containing sawdust (Wood Treatments Ltd., Macclesfield, Cheshire, UK) and shredded paper as bedding. Fun tubes and houses were provided as environmental enrichment. All animals were given RM1 diet (Special Diet Services Ltd., Witham, Essex, UK) and water *ad libitum* for the duration of the experiment.

Study design. The experimental method has been described previously (Ashby and Lefevre 2000). Rats were given 10 consecutive daily treatments consisting of a subcutaneous (sc) injection of either the corn oil vehicle or TP (0.4 mg/kg) and an oral dose of either vehicle or the relevant dose of FIN. We used 5 mL/kg and 2 mL/kg dosing volumes for oral and subcutaneous administrations, respectively.

The animals were killed 24 hr after the last treatment by an overdose of halothane (Concord Pharmaceuticals Ltd., Essex, UK). The following tissues were removed and weighed at necropsy: liver and combined kidneys (to 1 mg), Cowper's glands, levator ani/bulbo cavernosus muscles (LA/BC), seminal vesicles (with coagulating gland), ventral prostate, glans penis, and combined adrenal glands (all to 0.1 mg). Changes in tissue weights were evaluated for statistical significance using analysis of variance (ANOVA) and analyses of covariance (ANCOVA) with final body weight.

Uterotrophic assays. **Animals and dosing.** The experimental methods were as described previously (Odum et al. 2001). Female Alpk:APfSD (Wistar derived) rats were obtained from the barriered animal breeding unit (AstraZeneca). These were 19–20, 20–21, 21–22, or 22–23 days of age at the beginning of the study. The animals were housed up to a maximum of six per cage as described for the Hershberger assay. Each female received an sc injection of the appropriate compound on 3 consecutive days and was killed 24 hr after the last dose. On the basis of previously published data, 4MBC was given at 1 g/kg (Tinwell et al. 2002b) and ANT was given at 300 μg/kg (Odum et al. (2001); 5 μg/kg DES was used as a positive control. We used a dosing volume of 5 mL/kg body weight when animals were exposed to a single compound. However, in studies in which the animals were exposed to test compound and ANT, we used a dosing volume of 2.5 mL/kg body weight. Females were exposed to ANT 15 min before receiving the appropriate test compound (AO, 4MBC, or DES).

Study design. 4MBC had previously been shown to induce a significant increase in uterine weight in immature rats (Tinwell et al. 2002b). Thus, the initial experiment in the present study was designed to repeat those original observations. We also included additional groups of animals exposed to 4MBC together with the gonadotropin-releasing hormone (GnRH) antagonist ANT. In this experiment, 4MBC failed to significantly increase the uterine weight; however, the control uterine weights were greater than usual. Therefore, we repeated the experiment. In the final two experiments, we used animals of different ages at the start of the uterotrophic assays. All animals were killed 24 hr after the last dose by an overdose of fluothane followed by cervical dislocation. The uterus was removed, trimmed free of fat, gently blotted, and weighed. It was then placed in a preweighed vial, dried overnight (24 hr) at 70°C, and reweighed. All data were assessed for statistical significance using a two-sided Student's *t*-test.

Results

Hershberger assays. The assays for FIN described in this article represent a continuation of previous experiments (Ashby and Lefevre 2000; Blohm et al. 1986) and studies conducted as part of the current Organisation for Economic Co-operation and Development (OECD) validation of the Hershberger assay. The OECD studies demonstrated that a dose of FIN as low as 200 μg/kg/day inhibited the growth of the seminal vesicles, ventral prostate, and glans penis of TP-treated rats (J. Ashby, unpublished data). Subsequently, we have shown that doses of FIN as low as 8 μg/kg/day are active in the assay (Ashby

et al. 2004), with the prostate gland being the most sensitive tissue (Figure 1). The present experiments were designed to determine the no effect level for the inhibitory effect on the growth of the prostate in TP-treated animals. In all of the experiments, a dose of 5,000 μg/kg/day FIN significantly inhibited the growth of the seminal vesicles, ventral prostate, Cowper's glands, and glans penis compared with animals administered only TP (Table 1). In one study (experiment 1, Table 1) growth of the LA/BC was also significantly inhibited.

In the first low-dose experiment (experiment 1, Table 1), doses ranging from 5,000 μg/kg/day to 8 μg/kg/day FIN were all clearly active in the assay. The lowest dose tested (8 μg/kg/day FIN) affected all tissues except the LA/BC and the Cowper's glands. In experiment 2 (Table 1), the only tissues significantly inhibited at 8 μg/kg/day FIN were the seminal vesicles and glans penis. The glans penis was reduced in weight at all dose levels of FIN down to 0.1 μg/kg/day (Figure 2). However, the ventral prostate weight was significantly increased ($p < 0.01$) for the 2 μg/kg/day dose of FIN. In attempts to repeat this inverse prostate weight dose response (Figure 1), we conducted two further experiments (experiments 3 and 4, Table 1) using a tighter dose range of FIN around the dose that

had increased prostate weight in experiment 2. In the first of these repeat studies (experiment 3, Table 1), the only significant weight change observed was an increase in the weight of the Cowper's glands at 5 μg/kg/day FIN. This increase is probably associated with biologic variation and is probably not biologically significant (compare values for Cowper's glands weights in experiments 1–4, Table 1). However, seemingly in support of the previous experiment, we observed an inverted-U shaped response for increases in prostate weight around the 2 μg/kg/day dose of FIN, but none of the effects were statistically significant (Figure 1). In experiment 4, we did not observe the same pattern in the increase of the Cowper's glands weight seen in the experiment 3 (Table 1), and no change in prostate weight was observed below the 5,000 μg/kg/day positive control dose of FIN (Figure 1). In that final experiment, we observed a statistically significant reduction in the weight of the seminal vesicles at 5 μg/kg/day FIN. Inhibition of the growth of the glans penis was not consistently observed in these experiments (Figure 2).

The data sets recorded for the prostate gland during the course of these experiments are shown in Figure 1, both in terms of the weight of the prostate gland relative to the concurrent control value (taken as 100%) and in terms of absolute prostate weights.
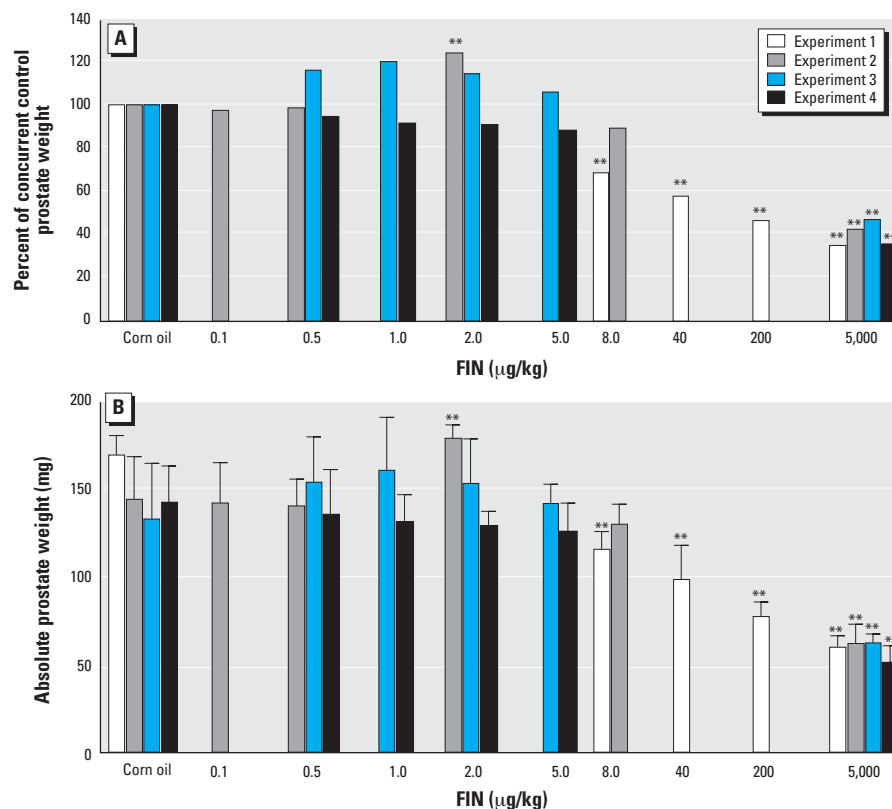

Figure 1. Rat prostate weights from four Hershberger assays of FIN shown as (*A*) a percentage of the concurrent control or as (*B*) absolute weights (mean ± SD) (data from Table 1).
**$p < 0.01$ versus relevant control.

Variability in control tissue and body weights. We calculated a mean of each control tissue weight and of control body weights from the four experiments, yielding a new mean weight value for all controls (corn oil + TP) combined. These means, multiplied by the SD of the mean and divided by 100, yield a percentage of the variation in control weights between experiments. The control prostate weights showed the highest variability between experiments, as follows: prostate, 11%; Cowper's glands, 6%; seminal vesicles, 5%; glans penis, 4%; LA/BC, 4%; liver, 1.8%; kidneys, 1.5%; body weight, 1.1%. The relatively high variability of all of the sex accessory tissues, compared with the liver and kidney weights and body weight, probably relates to the fact that the stimulatory dose of TP used in these Hershberger assays causes submaximal growth of the tissues, being just below the plateau-inducing dose of TP (Ashby and Lefevre 2000). Thus, small pharmacodynamic differences between experiments may yield small differences in the stimulatory dose of TP experienced by the tissues between experiments.

Uterotrophic assays. The data for the uterotrophic studies of 4MBC are presented in Table 2. We have previously shown 4MBC to be reproducibly active in the immature rat uterotrophic assay in this laboratory (Tinwell et al. 2002b) (experiment 1, Table 2). However, in the initial repeat study (experiment 2, Table 2), 4MBC failed to induce a significant increase in uterine weight. In this experiment the control uterine weights were heavier than usual: 35.1 ± 5.9 compared with 23.3 ± 4.3 (experiment 1, Table 2; Tinwell

et al. 2002b); thus, a repeat of this study was undertaken (experiment 3, Table 2). In experiment 3, control uterine weights were within our normal range, and 4MBC induced a weak, although significant ($p < 0.01$), increase in uterine weight. This activity was abolished by the GnRH inhibitor ANT, suggesting that the uterotrophic effects were mediated centrally via a temporal advance in puberty rather than via a direct effect of 4MBC on the uterus. Two further experiments were therefore performed in which we investigated the effect of the initial age of the rat on the activity of 4MBC (experiments 4 and 5, Table 2). Although 4MBC led to an increase in uterine weight in experiments 4 and 5, the significance of the response diminished with the older animals. Thus, the response to 4MBC was highly significant ($p < 0.01$) when 19–20-day-old animals were used; this was reduced to $p < 0.05$ with 20–21-day-old animals, and the

response was not significant when the older animals (21–22 and 22–23 days old) were used. These effects were reproducible between experiments 4 and 5 (Table 2).

Discussion

FIN inhibits the production of dihydrotestosterone from testosterone and consequently shows activity as a male rat developmental toxin with a lowest active dose of 10 µg/kg/day (Bowman et al. 2003). FIN also shows reproducible antiandrogenic activity in the Hershberger assay (Hershberger et al. 1953) at doses of ≥ 8 µg/kg (Figure 1A; Ashby et al. 2004). The present studies initially demonstrated a statistically significant low-dose trophic effect on the prostate gland at 2 µg/kg (experiment 2, Figure 1A), suggesting an androgenic, as opposed to antiandrogenic, effect. In the first attempted repeat of this trophic effect (experiment 3), we observed a
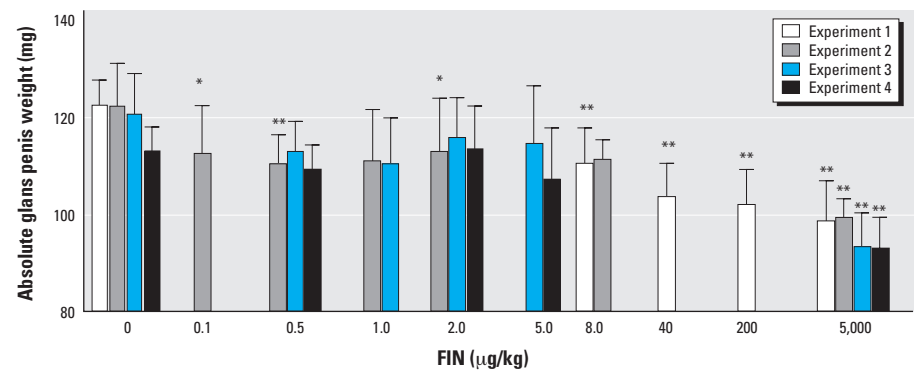

Figure 2. Absolute glans penis weights (mean ± SD) from four Hershberger assays of FIN (data from Table 1). *$p < 0.05$ and ** $p < 0.01$ versus control.

Table 1. Tissue weights from castrated rats administered corn oil (oral), corn oil + TP (0.4 mg/kg sc), or corn oil + TP + FIN (oral), each for 10 consecutive days.

| Experiment no. | Treatment (µg/kg) | Body weight (mean ± SD) | | Tissue weight (mean ± SD) | | | | | | | |
		Initial weight (g)	Final weight (g)	Liver (g)	Kidneys (g)	Adrenals (mg)	LA/BC (mg)	CG (mg)	SV (mg)	Prostate (mg)	Glans penis (mg)
1	Corn oil alone	244.5 ± 13.0	305.5 ± 23.2	13.69 ± 1.53	2.22 ± 0.11	64.2 ± 4.4	148.8 ± 18.2	5.6 ± 0.4	64.8 ± 10.9	24.8 ± 14.0	70.7 ± 6.1
	TP + corn oil	247.2 ± 8.6	342.5 ± 17.1	16.37 ± 1.28	2.45 ± 0.21	52.0 ± 7.4	381.2 ± 35.6	37.7 ± 2.1	506.8 ± 35.1	168.6 ± 13.1	122.4 ± 5.7
	TP + FIN 8	242.8 ± 13.9	335.2 ± 18.9	15.88 ± 1.32	2.47 ± 0.17	49.3 ± 4.2	353.1 ± 37.7	36.9 ± 3.2	405.0 ± 42.4**	114.9 ± 12.4**	110.1 ± 8.2**
	TP + FIN 40	243.7 ± 13.4	335.8 ± 20.4	15.71 ± 1.22	2.34 ± 0.26	53.7 ± 9.7	348.0 ± 20.7	30.4 ± 4.6*	335.5 ± 37.0**	97.6 ± 23.2**	103.4 ± 7.6**
	TP + FIN 200	246.5 ± 10.0	333.8 ± 18.8	15.34 ± 0.50	2.36 ± 0.14	54.8 ± 4.8	317.2 ± 30.9**	32.1 ± 1.3**	295.0 ± 27.8**	76.9 ± 10.1**	101.6 ± 7.2**
	TP + FIN 5,000	248.8 ± 10.2	338.2 ± 11.9	15.73 ± 1.13	2.38 ± 0.20	53.5 ± 3.2	314.3 ± 51.5**	22.0 ± 3.7**	188.9 ± 25.2**	58.0 ± 8.5*	98.7 ± 8.4**
2	TP + Corn oil	264.0 ± 8.6	349.2 ± 11.1	16.63 ± 0.82	2.47 ± 0.22	53.4 ± 8.7	359.4 ± 76.5	39.4 ± 2.4	508.1 ± 71.5	143.3 ± 26.3	121.9 ± 8.9
	TP + FIN 0.1	258.7 ± 7.4	350.2 ± 13.8	16.98 ± 1.12	2.43 ± 0.11	58.1 ± 8.8	366.8 ± 25.0	38.9 ± 2.7	516.7 ± 89.4	141.2 ± 24.3	112.9 ± 9.3*
	TP + FIN 0.5	267.2 ± 9.2	354.0 ± 14.3	17.28 ± 0.98	2.51 ± 0.12	61.4 ± 7.9	358.7 ± 41.7	35.9 ± 3.2	479.5 ± 32.6	140.9 ±16.2	110.1 ± 6.5**
	TP + FIN 2.0	261.5 ± 13.0	348.0 ± 24.9	16.82 ± 1.31	2.42 ± 0.10	55.6 ± 7.2	367.1 ± 35.9	38.6 ± 3.0	528.6 ± 66.3	178.6 ± 8.9**†	112.5 ± 11.5*
	TP + FIN 8.0	263.0 ± 4.6	351.2 ± 8.7	17.75 ± 1.03*	2.53 ± 0.14	50.0 ± 8.3	354.9 ± 45.5	39.6 ± 4.6	423.6 ± 52.5*	129.6 ± 13.3	111.3 ± 4.6*
	TP + FIN 5,000	264.0 ± 8.8	359.2 ± 14.2	17.49 ± 0.64	2.44 ± 0.20	49.9 ± 8.6	333.7 ± 12.6	23.9 ± 3.7**	215.7 ± 42.4**	60.2 ± 13.5**	99.2 ± 4.4**
3	TP + Corn oil	255.7 ± 10.6	340.5 ± 18.5	16.90 ± 1.41	2.40 ± 0.22	56.6 ± 11.1	345.3 ± 41.1	34.3 ± 3.6	466.5 ± 35.6	132.5 ± 32.9	120.8 ± 8.0
	TP + FIN 0.5	256.3 ± 12.6	339.2 ± 12.8	16.44 ± 0.87	2.30 ± 0.06	56.5 ± 4.1	373.3 ± 30.6	36.8 ± 4.5	474.8 ± 36.0	153.3 ± 27.9	112.5 ± 7.1
	TP + FIN 1.0	252.3 ± 8.6	334.3 ± 14.2	16.35 ± 1.35	2.44 ± 0.13	57.8 ± 2.8	333.8 ± 35.9	37.3 ± 1.2	443.9 ± 49.0	159.9 ± 31.7	110.9 ± 11.0
	TP + FIN 2.0	255.2 ± 5.9	342.8 ± 18.8	17.20 ± 2.00	2.49 ± 0.09	54.4 ± 7.0	362.6 ± 29.2	36.6 ± 3.7	474.9 ± 60.6	151.6 ± 29.3	115.9 ± 8.4
	TP + FIN 5.0	256.8 ± 9.1	341.0 ± 22.4	16.63 ± 1.83	2.49 ± 0.21	54.1 ± 5.4	365.0 ± 21.3	38.7 ± 3.7*†	457.9 ± 38.1	140.2 ± 14.0	114.4 ± 12.3
	TP + FIN 5,000	251.2 ± 8.1	332.8 ± 11.7	16.10 ± 0.74	2.34 ± 0.12	53.3 ± 2.9	319.6 ± 32.9	24.6 ± 2.2**	182.2 ± 24.3**	60.8 ± 6.6**	96.6 ± 7.6**
4	TP + Corn oil	252.5 ± 10.5	343.3 ± 14.5	17.05 ± 1.06	2.48 ± 0.08	61.5 ± 7.7	365.3 ± 31.5	37.9 ± 3.4	458.1 ± 36.9	142.0 ± 21.1	112.5 ± 5.3
	TP + FIN 0.5	252.7 ± 7.1	344.7 ± 6.2	17.43 ± 0.70	2.54 ± 0.11	60.5 ± 10.3	360.4 ± 48.1	41.8 ± 2.9	482.8 ± 44.9	134.1 ± 27.8	108.9 ± 5.7
	TP + FIN 1.0	252.8 ± 7.2	341.5 ± 7.1	17.23 ± 0.98	2.44 ± 0.12	60.1 ± 10.9	361.9 ± 34.5	39.2 ± 6.1	432.0 ± 49.2	130.4 ± 17.7	110.2 ± 9.9
	TP + FIN 2.0	250.3 ± 9.3	330.0 ± 18.9	16.27 ± 1.51	2.38 ± 0.09	58.8 ± 6.4	360.5 ± 30.4	37.8 ± 4.3	463.1 ± 45.6	128.6 ± 10.0	113.5 ± 9.2
	TP + FIN 5.0	251.0 ± 10.6	336.5 ± 14.3	16.99 ± 1.16	2.40 ± 0.10	56.5 ± 9.7	376.3 ± 18.7	35.5 ± 4.0	380.7 ± 71.7*	125.5 ± 17.3	107.4 ± 10.9
	TP + FIN 5,000	254.8 ± 8.9	341.7 ± 14.3	17.52 ± 0.86	2.46 ± 0.22	51.4 ± 8.2	340.0 ± 21.9	21.1 ± 4.9**	160.8 ± 23.5**	50.5 ± 9.7**	93.2 ± 6.4**

Abbreviations: CG, Cowper's glands; SV, seminal vesicles; ↑, elevated tissue weight.
*$p < 0.05$ and **$p < 0.01$, FIN + TP versus corn oil + TP by ANOVA and ANCOVA.

statistically nonsignificant increase in prostate weight at the 2-μg/kg dose, and in experiment 4 no effects were seen (Figure 1A). Because reproducibility of an effect is the primary criterion of its validity, we conclude that FIN does not show low-dose endocrine effects. In seeking an explanation for the initially observed low-dose trophic effect of FIN, we noticed that control prostate weights were the most variable of all of the tissue weights recorded in the Hershberger assay: 11% variance for prostate compared with values as low as 1.1% for body weights (see "Variability in control tissue and body weights" in "Results"). This variability is evident in Figure 1B, which shows absolute prostate weights for the series of experiments with FIN. Thus, the control prostate weight in experiment 1 is essentially the same as the statistically increased prostate weight recorded for the FIN group in experiment 2. This variability in control prostate weights raises the strong possibility that the significant effect seen for the low dose of FIN was a chance observation occurring within the normal interexperimental control range. Similarly, although the variability in control glans penis weights (4%) was lower than that for the prostate (11%), it is probable that the

nonreproducible decreases in glans penis weight also induced by low doses of FIN in experiment 2 were associated with normal biologic variability (Figure 2). The key points from these data for FIN are that the increase in prostate weight and the decreases in glans penis weights induced by FIN were all within the control range encountered for these tissues in the present experiments, and the effects induced were not reproducible.

The uterotrophic activity of 4MBC (Table 2, Figure 3) was dependent upon weight of the control uteri in the relevant experiment. There is no way to select for low control uterine weights because there is not a predictive correlation between initial body weight and uterine weights observed 4 days later (Table 2). Thus, it is a matter of chance whether an experiment has control uteri in the low weight range thereby giving rise to a positive response induced by 4MBC or whether the control uteri will be in the high range in which case 4MBC will be negative. The data shown in Figure 3 confirm, as expected, that the uterus grows marginally during postnatal days (PND) 19–25, before the major growth phase starting at puberty (around PND 29–30). By the time the maximum

uterine weight for this prepubertal growth is reached (~40 mg), the assay is no longer able to detect 4MBC as uterotrophic. The fact that the prepubertal growth of the uterus, and the uterotrophic activity of 4MBC, is abolished by GnRH inhibition indicates that both effects are centrally mediated, perhaps via modulation of centrally controlled release of low levels of estradiol from the ovaries or the adrenal glands. These data therefore indicate that there are two mechanisms by which an agent can induce an uterotrophic response. The first mechanism, as illustrated by the reference
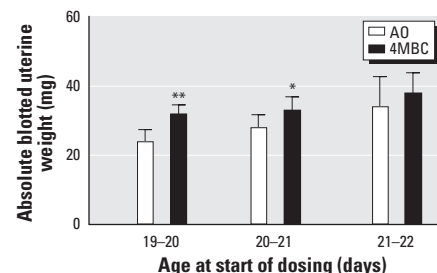

Figure 3. Activity (mean ± SD) of 4MBC (1,000 mg/kg sc) in the immature rat uterotrophic assay using animals of different ages at the start of dosing. $*p < 0.05$; $**p < 0.01$.

Table 2. Activity of 4MBC in the immature rat uterotrophic assay.

Experiment no.	Age at start of dosing (days)	Compound	Dose (per kg)	Uterine weight (mean ± SD) Blotted (mg)	Dry (mg)	Body weight (mean ±SD) Initial (g)	Terminal (g)
1 (Tinwell et al. 2002b)		AO	5 mL	23.3 ± 4.3	4.9 ± 0.8	46.9 ± 2.6	56.4 ± 3.2
		4MBC	500 mg	31.1 ± 4.6**	6.0 ± 0.7**	47.3 ± 3.6	56.9 ± 5.2
		4MBC	1,000 mg	29.4 ± 3.5**	5.6 ± 0.7**	46.4 ± 2.7	56.5 ± 4.1
		DES	5 μg	123.6 ± 8.2**	20.9 ± 1.7**	46.1 ± 5.1	56.4 ± 5.3
2	19–20	AO	5 mL	35.1 ± 5.9	6.6 ± 0.8	46.0 ± 1.7	56.4 ± 2.3
		4MBC	1,000 mg	35.1 ± 4.4	6.3 ± 0.5	46.3 ± 2.4	57.6 ± 4.4
		DES	5 μg	115.5 ± 3.3**	20.0 ± 0.3*	46.2 ± 2.6	56.2 ± 2.9
		ANT	300 μg	21.9 ± 4.0	4.4 ± 0.7	45.1 ± 3.2	57.2 ± 4.2
		4MBC/ANT	1 g/300 μg	20.0 ± 2.2	4.3 ± 0.5	43.8 ± 3.0	57.4 ± 3.7
		DES/ANT	5 μg/300 μg	125.1 ± 5.1**	19.8 ± 2.5**	46.6 ± 1.7	57.7 ± 2.2
3	19–20	AO	5 mL	20.0 ± 1.9	3.8 ± 0.6	38.5 ± 2.6	48.1 ± 2.7
		4MBC	1,000 mg	26.7 ± 2.8**	4.2 ± 1.3	39.4 ± 2.1	49.3 ± 2.0
		DES	5 μg	99.2 ± 6.2**	16.9 ± 1.3**	39.5 ± 1.5	51.2 ± 3.7
		ANT	300 μg	16.2 ± 0.9	3.4 ± 0.3	40.0 ± 2.6	48.8 ± 3.5
		4MBC/ANT	1 g/300 μg	16.8 ± 0.8	3.3 ± 0.4	40.7 ± 2.6	48.9 ± 4.1
		DES/ANT	5 μg/300 μg	113.7 ± 10.9**	19.0 ± 2.4**	40.7 ±2.6	50.8 ± 3.2
4	19–20	AO	5 mL	23.7 ± 3.8	4.8 ± 0.8	43.8 ± 4.3	53.9 ± 5.1
		4MBC	1,000 mg	32.1 ± 2.7**	5.8 ± 0.5**	43.2 ± 3.3	53.1 ± 4.5
		DES	5 μg	108.9 ± 15.6**	19.0 ± 2.5**	42.7 ± 5.2	51.4 ± 6.1
	20–21	AO	5 mL	28.0 ± 4.1	5.3 ± 0.6	49.2 ± 3.1	60.9 ± 4.0
		4MBC	1,000 mg	33.1 ± 3.8*	6.3 ± 0.6	48.6 ± 2.6	60.6 ± 5.0
	21–22	AO	5 mL	33.5 ± 9.8	6.3 ± 1.6	54.9 ± 5.2	65.5 ± 4.7
		4MBC	1,000 mg	38.2 ± 6.0	7.1 ± 1.0	53.3 ± 4.3	66.0 ± 4.9
5	19–20	AO	5 mL	23.0 ± 1.5	4.6 ± 0.4	41.8 ± 1.4	52.1 ± 2.0
		4MBC	1,000 mg	29.1 ± 4.8**	5.4 ± 0.8**	41.3 ± 2.4	50.6 ± 3.6
		DES	5 μg	111.8 ± 11.5**	19.1 ± 1.7**	40.8 ± 2.4	51.7 ± 3.9
	20–21	AO	5 mL	22.5 ± 3.2	4.5 ± 0.7	42.0 ± 3.2	51.1 ± 4.7
		4MBC	1,000 mg	26.8 ± 3.5*	5.0 ± 0.6	42.3 ± 3.6	51.9 ± 4.0
		DES	5 μg	109.9 ± 4.3**	18.1 ± 1.0**	42.7 ± 5.5	53.6 ± 6.9
	21–22	AO	5 mL	26.5 ± 3.2	5.2 ± 0.5	49.6 ± 2.5	62.4 ± 3.5
		4MBC	1,000 mg	28.9 ± 4.2	6.8 ± 3.5	47.9 ± 2.6	60.5 ± 3.6
		DES	5 μg	121.9 ± 28.5**	20.1 ± 2.3**	50.1 ± 1.4	62.7 ± 3.4
	22–23	AO	5 mL	29.8 ± 4.1	5.7 ± 0.8	49.1 ± 2.8	62.2 ± 3.1
		4MBC	1,000 mg	33.4 ± 6.3	6.5 ± 1.0	49.6 ± 2.9	61.4 ± 4.0
		DES	5 μg	108.6 ± 14.4**	19.4 ± 1.9**	52.1 ± 1.9	64.3 ± 3.8

All compounds were given by sc injection.
$*p < 0.05$, and $**p < 0.01$ by two-sided Student's t-test.

estrogen DES, involves direct action of the estrogen on the uterus leading to maximal growth of the tissue (from 20 to > 100 mg; Table 2). This mechanism, as expected, is not influenced by coadministration of the GnRH inhibitor ANT. The second mechanism, as illustrated by 4MBC, involves the test agent influencing the centrally mediated stimulation of prepubertal growth of the uterus. This activity is blocked by GnRH inhibition, and the magnitude of the uterotrophic effect is limited to the maximum control uterine weight that can be encountered in prepubertal animals (~40 mg).

The fact that the activity of 4MBC in the uterotrophic assays was dependent upon the concurrent control uterine weight led us to reevaluate our earlier data for BPA in the mouse uterotrophic assay (Tinwell et al. 2000) in which eight uterotrophic assays were described for the BPA dose range of 0.02 µg/kg to 300 mg/kg. Although we found some evidence of weak uterotrophic activity at some doses, none of the effects was consistently reproducible. The clearest example of this was for the 200-mg/kg dose of BPA, which gave four weak positive responses and four negative responses (Figure 4A). When those data are reordered from chronological order into the order of descending control uterine weights (Figure 4B), the results become consistent with the effects seen for 4MBC. Thus, the positive BPA results were weak and only became significant when the control uterine weights were low. Further, the magnitude of the effects induced by BPA remained constant: the decreasing control values created the statistically significant effects for BPA.

These data for FIN, 4MBC, and BPA indicate that concurrent control values and interexperimental variation in control levels can influence the interpretation of weak endocrine toxicities and thereby affect the qualitative outcome of bioassays. This conclusion is possibly relevant to several other instances of low-dose or weak endocrine effects that have proven resistant to independent confirmation. For example, Sharpe et al. (1995) reported that octylphenol (OP) and BBP were able to reduce rat testes weights. We were unable to confirm those effects for BBP (Ashby et al. 1997), and this led Sharpe et al. to repeat their evaluation of OP (Sharpe et al. 1998). In that repeat study, the control testes weights had fallen and OP was no longer active (Sharpe et al. 1998; Figure 5). Uncertain of the origins of these effects, these researchers ran a further control group and recorded control testes weights similar to their original 1995 study (Figure 5; Sharpe et al. 1998). They were unable to account for this variability in control testes weights, and the activity of BBP and OP in the assay remains unconfirmed. A similar situation exists for the

effects of NP on the rat prostate gland (Lee 1998). Lee described a study in which NP reduced prostate weight and in which access to five identical control groups was possible. In the first four of these, control prostate weights were approximately 60 mg/100 g body weight, but in the fifth study this value was significantly lower (~40 mg/100 g body weight; Figure 6). Lee (1998) did not comment on this unexpected and statistically significant change in control prostate weight, which itself led to the conclusion that NP was uniquely inactive in that fifth experiment (Figure 6). We were subsequently unable to show any effects for NP (Odum and Ashby 2000) when tested under the conditions described by Lee (1998), and the control prostate weights were similar to the first four studies reported by Lee (1998).
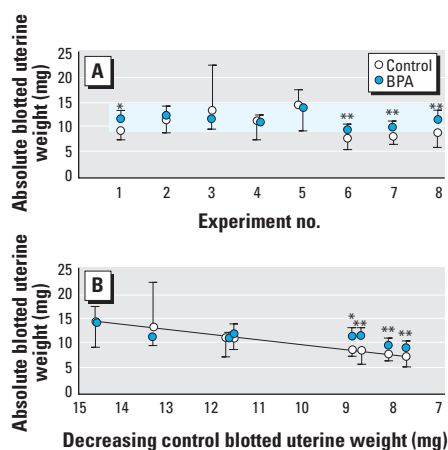
A further example of how the variability of control prostate weights can influence the interpretation and confirmation of chemically induced effects is illustrated by the activity of BPA on the CF1 mouse prostate gland (Figure 7). Nonneman et al. (1992) described the effect of intrauterine position on the subsequent weight of the prostate gland of CF1 mice. Subsequently, Nagel et al. (1997), from the same laboratory, reported the ability of BPA to increase the weight of the mouse prostate gland, with the implication that the estrogenic activity of BPA had simulated the presence of female fetuses being next to male fetuses *in utero*. Cagen et al. (1999) and Ashby et al. (1999) were unable to confirm those effects of BPA, despite the fact that Cagen et al. (1999) recorded control prostate weights as low as those of Nagel et al. (1997).

Figure 4. Activity (mean ± SD) of BPA (200 mg/kg sc) and controls in the immature mouse uterotrophic assay ordered by (*A*) experiment number and (*B*) decreasing control blotted uterine weight. The blue bar in (*A*) indicates the range of BPA values. Data from Tinwell et al. (2000).
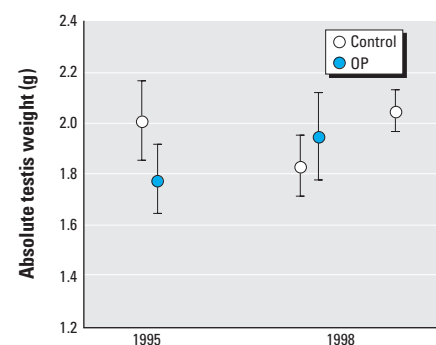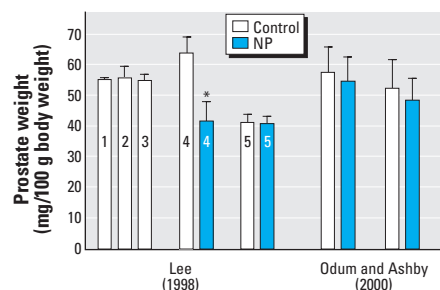*$p < 0.05$; **$p < 0.01$.

Figure 5. Testis weight (mean ± SD) from 90- to 95-day-old control rats and those exposed to 1,000 µg/L OP either *in utero* or during lactation. The control testes weights were reduced in the second study compared to the original control data. An additional control group was assessed in the repeat (1998) study. These weights were similar to the original 1995 data. The authors were unable to account for these differences.

Figure 6. Rat prostate weight (mean ± SD) from five groups of controls reported by Lee (1998) and repeat data from Odum and Ashby (2000). All of Lee's study groups (study number shown within bar) had similar mean body weights at termination on postnatal day (PND) 31. Lee (1998) observed positive effects for NP in study 4 (8 mg/kg NP on PND 6–24) but not in study 5 (8 mg/kg NP on PND 13–30); no account was given of the difference in control prostate weights in the fifth experiment compared with those reported in studies 1–4. In the repeat studies of Odum and Ashby (2000), rats were treated with 8 mg/kg NP on PND 6–24.
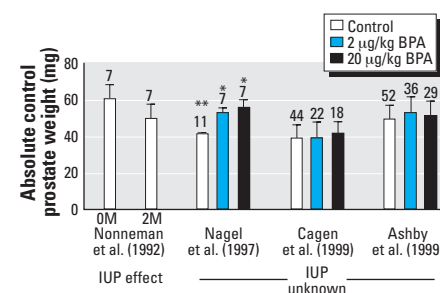*Weight significantly reduced compared to concurrent control.

Figure 7. Absolute weight (mean ± SD) of the prostate from control and BPA-treated CF1 mice. Abbreviations: 0M, male fetus positioned between two females; 2M, male fetus positioned between two males; IUP, intrauterine position. Group sizes are shown above each bar. Animals were killed at 6 months of age, except for those reported by Cagen et al. (1999), which were killed at 3 months of age.
*Statistically increased. **Significantly different ($p < 0.01$) from the 0M mice reported by Nonneman et al. (1992).

Within the present context, the most interesting aspect of the combined data shown in Figure 7 is the variability of the control prostate weights between laboratories and, most particularly, within the laboratory of Nonneman and Nagel. Again, the most interesting questions are what factors influence the unexplained variations in control prostate weights, within and between laboratories, and how these variations influence interpretation of the effects seen for BPA.

A further example of the importance of control variability is illustrated by the reports of weak effects induced by BPA on rat testes (Sakaue et al. 2001; Talsness et al. 2000). We were unable to confirm the studies of either Sakaue et al. (Tinwell et al. 2002) or Talsness et al. (Ashby et al. 2003) after extensive studies. The results reported by Sakaue et al. (2001) and related results from the same laboratory (Ohsako et al. 2001; Sakaue et al. 1999) enable reference to four sets of control observations for rat daily sperm production (DSP) (Figure 8). The most marked reduction in DSP induced by BPA in the two experiments was reported by Sakaue et al. (2001) (Figure 8). However, neither of these effects was significantly different from the control DSP value reported in the same year by Ohsako et al. (2001; bar 2, Figure 8). This therefore provides a further instance of unaccounted control variability enveloping chemically induced effects from another study conducted in the same laboratory. In four independent repeat studies (Ashby et al. 2003), we found that test and control DSP
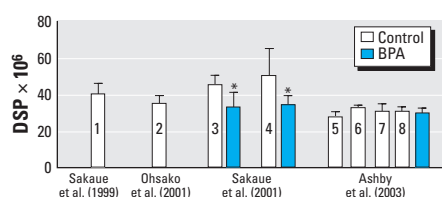

Figure 8. Comparison of control DSP (mean ± SD) reported from the same laboratory [Ohsako et al. (2001) and Sakaue et al. (1999, 2001)] and a different laboratory (Ashby et al. 2003) with the greatest effect induced by BPA (Sakaue et al. 2001). A range of BPA doses was used in these experiments, and only the dose that induced the greatest effect in each experiment is shown: 20 µg/kg (Sakaue et al. 1999); 200 µg/kg (Sakaue et al. 2001); 200 mg/kg (Ashby et al. 2003). The effect of BPA is not significantly different from the control reported by Ohsako et al. (2001; bar 2: one- or two-sided Student's *t*-test). Sakaue et al. (1999) and Ohsako et al. (2001) used Holtzman rats, and Sakaue et al. (2001) and Ashby et al. (2003) used Sprague-Dawley rats. However, the identical control DSP values for Holtzman rats (Sakaue et al. 1999, bar 1) and Sprague-Dawley rats (Sakaue et al. 2001; bar 3) indicate that rat strain is not a key variable on control DSP values and that, consequently, it is possible to compare data across strains and experiments for that laboratory.
*Reported by Sakaue et al. (2001) as statistically different from the concurrent control (bars 3 and 4).

values were not significantly different from each other and were similar to the BPA values reported by Sakaue et al. (2001) (Figure 8). We also have established that diet was not the cause of the different assay outcomes (Ashby et al. 2003; Odum et al. 2001).

The remaining question posed by the low-dose studies on FIN is how to interpret effects that show a nonsignificant change, as illustrated by the nonsignificant increase in prostate weight seen for FIN (experiment 2, Figure 1). If the statistical methods used are appropriate, the absence of significance should indicate the absence of a chemically induced effect. An example of this problem is provided by Howdeshell et al. (2003), who discussed in detail a nonsignificant increase ($p = 0.31$) in uterine weight for CF1 mice housed in worn polycarbonate cages equipped with worn polycarbonate drinking bottles (Figure 9). Howdeshell et al. (2003) associated this effect with the BPA (~300 ng/mL) that was reported to leach from the cages during the course of the experiment, but it may equally have been related to the heavier body weights of the test animals at the start of the experiment ($p = 0.17$). The group sizes in that study were large (57/group), but this represented the pooling of three replicate experiments. It would have been useful to know the results of those three individual experiments to evaluate whether the nonsignificant increase in uterine weight was reproducible between replicates, especially given the problems we encountered (Tinwell et al. 2000) when evaluating the uterotrophic activity of BPA in mice.

Some endocrine-mediated toxicities are highly specific to the conditions of the test, as exemplified by the specificity of the effects induced by BPA in the vagina of F344 rats but not Sprague-Dawley rats (Long et al. 2000) and by the resistance of C57 mice, as opposed to CD1 mice, to the reproductive tract developmental effects of neonatal DES exposure (Couse et al. 2001). Such clear-cut specificities, usually defined in the same laboratory, are
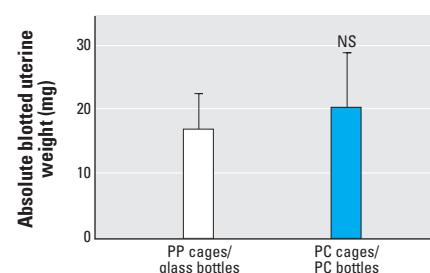

Figure 9. Mean (± SD) uterine weights for mice maintained under different housing conditions (Howdeshell et al. 2003). Abbreviations: NS, not significant; PC, polycarbonate plastic; PP, polypropylene plastic. Values are based on an average of three females from six litters in three replicate experiments; $n = 57$ for each group. Data from individual experiments are not shown.

distinct from the several recent failures to confirm endocrine toxicities between laboratories when using ostensibly the same test protocol. We suggest that the latter instances can only be resolved, and avoided in future, if the following actions are taken:
• For the end points being studied, it is useful to have a historical database with which to compare current assay performance and test results. Inclusion of this database into publications would aid resolution of the problems described herein.
• Most laboratories experience variability in the end points they measure. Understanding the origin of these variations is a necessary precursor to concluding weak endocrine activity for test agents.
• When low-dose or weak endocrine effects are indicated, it is important to confirm the observations before publication.
• Unadjusted data from separate experiments should be shown, as opposed to pooled and adjusted data. Adjusted data are subsidiary to the original data.
• A distinction should be drawn between effects that lie within recent historical control levels and those that exceed those levels. Different mechanisms may operate in these two situations, leading to different approaches to data extrapolation.
• Use of appropriate statistical methods enables objective qualitative judgments to be made. Subjective discussion of statistically nonsignificant effects should be avoided.

REFERENCES

Ashby J. 1987. The efficient preparation of corn oil suspensions. Mutat Res 187:45.
Ashby J. 2003. Endocrine disruption occurring at doses lower than those predicted by classical chemical toxicity evaluations: the case of bisphenol-A (BPA). Pure Appl Chem 75:2167–2180.
Ashby J, Lefevre PA. 2000. Preliminary evaluation of the major protocol variables for the Hershberger castrated male rat assay for the detection of androgens, antiandrogens, and metabolic modulators. Regul Toxicol Pharmacol 31:92–105.
Ashby J, Lefevre PA, Tinwell H, Odum J, Owens W. 2004. Testosterone-stimulated weanlings as an alternative to castrated male rats in the Hershberger anti-androgen assay. Regul Toxicol Pharmacol 39:229–238.
Ashby J, Tinwell H, Haseman J. 1999. 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.
Ashby J, Tinwell H, Lefevre PA, Joiner R, Haseman J. 2003. The effect on sperm production in adult Sprague-Dawley rats exposed by gavage to bisphenol A between postnatal days 91–97. Toxicol Sci 74:129–138.
Ashby J, Tinwell H, Lefevre PA, Odum J, Paton D, Millward SW, et al. 1997. Normal sexual development of rats exposed to butyl benzyl phthalate from conception to weaning. Regul Toxicol Pharmacol 26:102–118.
Blohm TR, Laughlin ME, Benson HD, Johnston JO, Wright CL, Schatzman GL, et al. 1986. Pharmacological induction of 5alpha-reductase deficiency in the rat: separation of testosterone-mediated and 5alpha-dihydrotestosterone-mediated effects. Endocrinology 119:959–966.
Bolt HM, Guhe C, Degen GH. 2001. Comments on "*In vitro* and *in vivo* estrogenicity of UV screens" [Letter]. Environ Health Perspect 109:A358–A359.
Bowman CJ, Barlow NJ, Turner KJ, Wallace DG, Foster PMD.

2003. Effects of in utero exposure to finasteride on andro-gen-dependent reproductive development in the male rat. Toxicol Sci 74:393–406.

Cagen SZ, Waechter JM Jr, Dimond SS, Breslin WJ, Butala JH, Jekat FW, et al. 1999. Normal reproductive organ develop-ment in CF-1 mice following prenatal exposure to bisphenol A. Toxicol Sci 50:36–44.

Colerangle JB, Roy D. 1996. Exposure of environmental estro-genic compound nonylphenol to Noble rats alters cell-cycle kinetics in the mammary gland. Endocrine 4:115–122.

Couse JF, Dixon D, Yates M, Moore AB, Ma L, Maas R, et al. 2001. Estrogen receptor-alpha knockout mice exhibit resistance to the developmental effects of neonatal diethylstilbestrol exposure on the female reproductive tract. Dev Biol 283:224–238.

Hershberger LG, Shipley EG, Meyer RK. 1953. Myotrophic activ-ity of 19-nortestosterone and other steroids determined by the modified levator ani muscle method. Proc Soc Exp Biol Med 83:175–180.

Howdeshell KL, Peterman PH, Judy BM, Taylor JA, Orazio CE, Ruhlen RL, et al. 2003. Bisphenol A is released from used polycarbonate animal cages into water at room tempera-ture. Environ Health Perspect 111:1180–1187.

Lee PC. 1998. Disruption of male rat reproductive tract develop-ment by administration of the xenoestrogen, nonylphenol, to male newborn rats. Endocrine 9:105–111.

Long X, Steinmetz R, Ben-Jonathan N, Caperell-Grant A, Young PCM, Nephew KP, et al. 2000. Strain differences in vaginal responses to the xenoestrogen bisphenol A. Environ Health Perspect 108:243–247.

Nagel SC, vom Saal FS, Thayer KA, Dhar MG, Boechler M, Welshons WV. 1997. Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bio-activity of the xenoestrogens bisphenol A and octylphenol. Environ Health Perspect 105:70–76.

Nonneman DJ, Ganjam VK, Welshons WV, vom Saal FS. 1992. Intrauterine position on steroid metabolism and steroid receptors of reproductive organs in male mice. Biol Reprod 45:723–729.

Odum J, Ashby J. 2000. Neonatal exposure of male rats to nonylphenol has no effect on the reproductive tract. Toxicol Sci 56:400–404.

Odum J, Pyrah IT, Foster JR, Van Miller JP, Joiner RJ, Ashby J. 1999. Comparative activities of p-nonylphenol and diethyl-stilbestrol in Noble rat mammary gland and uterotrophic assays. Regul Toxicol Pharmacol 29:184–195.

Odum J, Tinwell H, Jones K, Van Miller JP, Joiner RL, Tobin G, et al. 2001. Effect of rodent diets on the sexual develop-ment of the rat. Toxicol Sci 61:115–127.

Ohsako S, Miyabara Y, Nishimura N, Kurosawa S, Sakaue M, Ishimura R, et al. 2001. Maternal exposure to a low dose of 2,3,7,8-tetrachlorodibennzo-p-dioxin (TCDD) suppressed the development of reproductive organs of male rats: dose dependent increase of mRNA levels of 5α-reductase type 2 in contrast to decrease of androgen receptor in pubertal ventral prostate. Toxicol Sci 60:132–143.

Sakaue M, Ohsako S, Ishimura R, Kurosawa S, Kurohmaru M, Hayashi Y, et al. 1999. Preliminary BPA experiment. In: 127th Annual Meeting of the Japanese Society of Veterinary Science, 2–4 April, Fujisawa, Japan, 47.

Sakaue M, Ohsako S, Ishimura R, Kurosawa S, Kurohmaru J, Hayashi Y, et al. 2001. Bisphenol A affects spermatogenesis in the adult rat even at a low dose. J Occup Health 43:185–190.

Schlumpf M, Berger L, Cotton B, Conscience-Egli M, Durrer S, Fleischmann I, et al. 2001a. Estrogen active UV screens. SOFW J 127:10–18.

Schlumpf M, Cotton B, Conscience M, Haller V, Steinmann B, Lichtensteiger W. 2001b. In vitro and in vivo estrogenicity of UV screens. Environ Health Perspect 109:239–244.

Sharpe RM, Fisher JS, Millar MM, Jobling S, Sumpter JP. 1995. Gestational and lactational exposure of rats to xeno-estrogens results in reduced testicular size and sperm production. Environ Health Perspect 103:1136–1143.

Sharpe RM, Turner KJ, Sumpter JP. 1998. Endocrine disruptors and testis development [Letter]. Environ Health Perspect 106:A220–A221.

Talsness C, Fialkowski O, Gericke C, Merker H-J, Chahoud I. 2000. The effects of low and high doses of bisphenol A on the reproductive system of female and male rat offspring. Congenital Anom 40:S94–S107.

Tinwell H, Haseman J, Lefevre PA, Wallis N, Ashby J. 2002a. Normal sexual development of two strains of rat exposed in utero to low doses of bisphenol A. Toxicol Sci 68:339–348.

Tinwell H, Joiner R, Pate I, Soames A, Foster J, Ashby J. 2000. Uterotrophic activity of bisphenol A in the immature mouse. Regul Toxicol Pharmacol 32:118–126.

Tinwell H, Lefevre PA, Moffat GJ, Burns A, Odum J, Spurway TD, et al. 2002b. Confirmation of uterotrophic activity of 3-(4-methylbenzylidine)camphor in the immature rat. Environ Health Perspect 110:533–536.