INTEGRATED SUMMARY REPORT

for

Validation of a Test Method for Assessment of Pubertal Development and Thyroid Function in Juvenile Male Rats as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery

September 2007

U.S. Environmental Protection Agency
Office of Science Coordination and Policy
and
Office of Research and Development
Washington, D.C.

Table of Contents

List of Appendices

List of Tables

I. Introduction

A. Purpose of the EDSP

Section 408(p) of the Federal Food Drug and Cosmetic Act (FFDCA) requires the U.S. Environmental Protection Agency (EPA) to

> *develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate [21 U.S.C. 346a(p)].*

Subsequent to passage of the Act, the EPA formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a committee of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement its Endocrine Disruptor Screening Program (EDSP). The EDSP is described in detail at the following website:

http://www.epa.gov/scipoly/oscpendo/

Upon recommendations from the EDSTAC (1998), the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormonal systems as well as wildlife.

B. Tiered approach

The EPA accepted the EDSTAC's recommendations for a two-tier screening program as proposed in a Federal Register Notice in 1998 (USEPA (1998)). The purpose of Tier 1 is to identify the potential of chemicals to interact with the estrogen, androgen, or thyroid (EAT) hormonal systems. A negative result in Tier 1 would be sufficient to put a chemical aside as having low to no potential to cause endocrine disruption, whereas a positive result would require further testing in Tier 2. The purpose of Tier 2 is to more definitively identify and characterize the potential hazard on the endocrine system and to provide risk assessment based, in part, on dose-response

relationships. Tier 2 is expected to comprise multigeneration tests in species representative of various taxa (i.e., mammals, birds, fish, amphibians, and invertebrates).

C. The Tier-1 battery

The EDSTAC (1998) concluded that a Tier-1 battery should be comprised of a suite of complementary screening assays having the following characteristics:

- Maximum sensitivity to minimize false negatives while permitting an as yet undetermined, but acceptable, level of false positives.
- Range of organisms representing known or anticipated differences in metabolic activity and include assays from representative vertebrate classes to reduce the likelihood that important pathways for metabolic activation or detoxification of parent substances or mixtures are not overlooked.
- Capacity to detect all known modes of action (MOAs) for the endocrine endpoints of concern. All chemicals known to affect the action of EAT hormones should be detected.
- Range of taxonomic groups among the test organisms. There are known differences in endogenous ligands, receptors, and response elements among taxa that may affect the endocrine activity of chemical substances or mixtures.
- Diversity among the endpoints and within and among assays to reach conclusions based on "weight-of-evidence" considerations. Decisions based on the screening battery results will require weighing the data from several assays.
- Inexpensive, quick, and easy to perform.

To detect chemicals that may affect the EAT hormonal systems through any one of the known MOAs — interruption of hormone production or metabolism, binding of the hormone with its receptor, interference with hormone transport, etc. — the EDSTAC recommended the *in vitro* and *in vivo* assays shown in Table 1 for inclusion in the Tier-1 screening battery.

Table 1. Tier-1 *in vitro* and *in vivo* screening assays recommended by the EDSTAC

Assays	Reasons for consideration
Estrogen receptor (ER) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the ER.
Androgen receptor (AR) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the AR.
In vitro steroidogenesis	A sensitive *in vitro* test to detect chemicals that interfere with the synthesis of the sex steroid hormones.
Uterotropic (rat)	An *in vivo* assay to detect estrogenic chemicals. It offers the advantage over the binding assay of incorporating absorption, distribution, metabolism, and excretion (ADME)
Hershberger (rat)	An *in vivo* assay to detect androgenic and anti-androgenic chemicals. It offers the advantage over the binding assay of incorporating ADME and differentiating between AR agonists and antagonists.
Pubertal female (rat)	An assay to detect chemicals that act on estrogen or through the hypothalamus-pituitary-gonadal (HPG) axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system.
Frog metamorphosis	A sensitive assay for detection of chemicals that interfere with the thyroid hormone system.
Fish screen	Fish are the furthest removed from mammalians among vertebrates both from the standpoint of evolution—their receptors and metabolism are different from mammals—and exposure/habitat, since they would be subject to exposure through the gills, whole body, and diet. Thus, the fish assay would augment information found in the mammalian assays and would be more relevant than the mammalian assays in triggering concerns for fish.

In addition, the EDSTAC recognized there were other combinations of screening assays that may be suitable and, therefore, recommended that the EPA validate the alternative screening assays shown in Table 2.

Table 2. Alternative *in vitro* and *in vivo* assays recommended for the Tier-1 Screening Battery

Assays	Reasons for consideration
In vitro placental aromatase	The aromatase assay detects chemicals that inhibit aromatase and would be needed if either of the two following assays using males were substituted for the female pubertal assays. The male is not believed to be as sensitive to alterations in aromatase as the female and would not therefore be sufficient to detect interference with aromatase in the screening battery.
Pubertal male (rat)	The assay detects chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.
Adult male (rat)	The assay is also designed to detect chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.

D. Validation

As noted, Section 408(p) of the FFDCA requires the EPA to use validated test systems. Validation has been defined as *"the process by which the reliability and relevance of a test method is evaluated for a particular use"* (OECD (1996); NIEHS (1997)).

Reliability is defined as the reproducibility of results from an assay within and between laboratories.

Relevance describes whether a test is meaningful and useful for a particular purpose (OECD (1996)). For Tier-1 EDSP assays, relevance can be defined as the ability of an assay to detect chemicals with the potential to interact with the EAT hormonal pathways.

Federal agencies are also instructed by the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) Authorization Act of 2000 to ensure that new and revised test methods are valid prior to their use.

In general, the EPA is following a five-part or stage validation process outlined by the ICCVAM (NIEHS (1997)). The EPA believes that it is essential to recognize that this process was specifically developed for *in vitro* assays intended to replace *in vivo* assays. The fundamental problem confronting the EPA is how to adapt and work with this process for rodent and ecological *in vivo* assays in Tiers 1 and 2 that have no suitable *in vitro* substitute.

Nonetheless, the stages of the process outlined by the ICCVAM are as follows:

First Stage - *Test Development*, an applied research function which culminates in an initial protocol. As part of this phase, the EPA prepares a Detailed Review Paper (DRP) to explain the purpose of the assay, the context in which it will be used, and the scientific basis upon which the assay's protocol, endpoints, and relevance rest. The DRP reviews the scientific literature for candidate protocols and evaluates them with respect to a number of considerations, such as whether the candidate protocols meet the assay's intended purpose, the costs and other practical considerations. The DRP also identifies the developmental status and questions related to each protocol; the information needed answer the questions; and, when possible, recommends an initial protocol for the initiation of the second stage of validation.

Second Stage - *Standardization and Optimization,* in which the protocol is refined, optimized, standardized and initially assessed for transferability and performance. Several different types of studies are conducted during this second phase depending upon the state of development of the method and the nature of the questions that the protocol raises. The initial assessment of transferability is generally a trial in a second laboratory to determine that another laboratory besides the lead laboratory can follow the protocol and execute the study.

Third Stage - *Inter-laboratory Validation* studies are conducted in independent laboratories with the optimized protocol. The results of these studies are used to determine inter-laboratory variability and to set or cross-check performance criteria.

Fourth Stage - *Peer Review*, an independent scientific review by qualified experts.

Fifth Stage - *Regulatory Acceptance*, adoption for regulatory use by an agency. The EPA has developed extensive guidance on the conduct of peer reviews because the Agency believes that peer review is an important step in ensuring the quality of science that underlies its regulatory decisions (USEPA (2007)).

Criteria for the validation of alternative test methods *(in vitro* methods designed to replace animal tests in whole or in part) have generally been agreed upon in the United States by the ICCVAM, in Europe by the European Centre for the Validation of Alternative Methods (ECVAM), and internationally by the Organisation for Economic Co-Operation and Development (OECD). These criteria as stated by ICCVAM (NIEHS (1997)) are as follows:

1. The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.
2. The relationship of the endpoints determined by the test method to the *in vivo* biologic effect and toxicity of interest must be addressed.
3. A formal detailed protocol must be provided and must be available in the public domain. It should be sufficiently detailed to enable the user to adhere to it and should include data analysis and decision criteria.
4. Within-test, intra-laboratory and inter-laboratory variability and how these parameters vary with time should have been evaluated.
5. The test method's performance must have been demonstrated using a series of reference chemicals preferably coded to exclude bias.
6. Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test to that of the test it is designed to replace.
7. The limitations of the test method must be described (e.g., metabolic capability).

8. The data should be obtained in accordance with Good Laboratory Practices (GLPs).
9. All data supporting the assessment of the validity of the test methods including the full data set collected during the validation studies must be publicly available and, preferably, published in an independent, peer-reviewed publication.

The EPA has adopted these various validation criteria for the EDSP as described (USEPA (2007)). Although attempts have been made to thoroughly comply with all validation criteria, the various *in vitro* and *in vivo* screening assays are not replacement assays (Validation Criterion No. 6). Many of them are novel assays; consequently, large data bases do not exist as a reference to establish their predictive capacity (e.g., determination of false positive and false negative rates). It is expected that the review of results from the testing of the first group of 50 to 100 chemicals that was recommended by the Scientific Advisory Panel (SAP) (USEPA (1999)) will allow a more complete assessment of the performance of the Tier-1 screening battery in time.

For technical guidance in developing and validating the various Tier-1 screens and Tier-2 tests, the EPA chartered two federal advisory committees: the Endocrine Disruptor Methods Validation Subcommittee, or EDMVS (from 2001 to 2003), and the Endocrine Disruptor Methods Validation Advisory Committee, or EDMVAC (from 2004 to 2006). These committees, composed of scientists from government, academia, industry, and various interest groups, were charged to provide expert advice to the EPA on protocol development and validation. The EPA also cooperates with member countries of the OECD to develop and validate assays of mutual interest to screen and test for endocrine effects.

It should be remembered that even though assays are being developed and validated individually and peer reviewed on an individual basis (i.e., their strengths and limitations are being evaluated as stand-alone assays), the Tier-1 assays will be used in a battery of complementary screens. An individual assay may serve to strengthen the weight of evidence in a determination (e.g., positive results in an ER binding assay in conjunction with positive results in the uterotropic and pubertal female assays would provide a consistent signal for estrogenicity) or to provide coverage of MOAs not

addressed by other assays in the battery. Information supporting the validation of an individual assay may be used at a later date by the Federal Insecticide, Fungicide, Rodenticide Act (FIFRA) SAP for peer review of the EPA's recommendations for a Tier-1 battery. The Tier-1 battery peer review will focus, in part, on the extent of coverage and overlap the suite of assays will have with one another in detecting endocrine-related effects associated with the EAT hormonal systems.

II. Purpose of this report

The purpose of this Integrated Summary Report is to provide a historical summary of the development and validation of a standardized protocol for the male pubertal rat assay proposed as an alternative *in vivo* assay for the Tier-1 screening battery. The reasoning and judgments leading to the various studies, and conclusions concerning the strengths and weaknesses of the assay in its current form, are presented.

III.Purpose of the assay

The purpose of the male pubertal assay is to provide information obtained from an *in vivo* mammalian system that will be useful in assessing the potential of a chemical substance or mixture to interact with the endocrine system. This assay is capable of detecting chemicals with antithyroid, androgenic, or antiandrogenic [androgen receptor (AR) or steroid-enzyme-mediated] activity or agents which alter pubertal development via changes in gonadotropins, prolactin, or hypothalamic function.

Weanling rats, standardized to 8 - 10 per litter at post-natal day (PND) 3-5, are housed 2 to 3 per cage. The test chemical is administered in corn oil by oral gavage (2.5 to 5.0 ml/kg) between 0700 and 0900 (lights 14:10, on 0500h) from PND 23 - 53 (31 days) to 15 males per dose level. The endpoints are growth (body weight); age at preputial separation; serum testosterone, thyroxine (T_4) and thyroid stimulating hormone (TSH); weights of reproductive organs (seminal vesicle plus coagulating gland (with and without fluid), ventral prostate, dorsolateral prostate, levator ani plus bulbocavernosus muscle complex, epididymis, testis); histology of epididymis, testis, thyroid, and kidney; and weights of thyroid, liver, kidney, adrenal, and pituitary. Other endpoints were

included in some of the preliminary studies (serum T_3, estradiol, luteinizing hormone, prolactin, *ex vivo* testis and pituitary hormone production and hypothalamic neurotransmitter concentrations) but these were later removed from the protocol as being relatively uninformative due to wide variation in levels.

The protocol is included as Appendix 1.

IV. Relevance of the assay

As noted above in Section I.D, relevance describes whether a test is meaningful and useful for a particular purpose. For Tier 1 of the EDSP, the purpose is to identify chemicals with the potential to interact with the endocrine system.

An extensive review of the basis for selecting the endpoints and the conditions of the assay was published (Stoker *et al.* (2000b), attached to this Integrated Summary Report as Appendix 2) prior to the validation effort described here, and should be regarded as the primary discussion of the relevance of the endpoints and thus the assay. Because that document, which serves as the Detailed Review Paper (DRP) for the male pubertal protocol, describes in detail the biological relevance of the endpoints, the information presented here will be brief.

Serum androgens in male rats change dramatically during puberty and reproductive organ weights grow rapidly during puberty (Stoker *et al.* (2000b)). This makes the prepubertal period a very sensitive age for exposure to agents which alter the endocrine system. Preputial separation is an apical measure of the progression of puberty and it has been used as the primary biomarker of puberty onset in the rat. It is an androgen dependent event.

V. Overview of studies relevant to validation of the assay

The validation process involved 13 positive test chemicals of various modes of action and strengths, and a test chemical which had not previously been tested for endocrine activity but which had been shown to be negative for reproductive and developmental toxicity. These studies involved five different contract research laboratories working under Good Laboratory Practices (GLP) (Table 3). The contract

studies on the known-positive compounds tested the transferability of the protocol from the developers of the assay (EPA) to contract laboratories.

Following a description of the preliminary contract studies which tested transferability of the assay and examined the applicability across various modes of endocrine action, the results of an interlaboratory comparison study are summarized. This study provided information on the reliability and reproducibility of the assay when conducted in different laboratories.

Results of several in house studies run by EPA's Office of Research and Development (ORD) laboratories to address specific questions will also be presented in this report. Seven chemicals were tested by an ORD laboratory to further assess compounds which alter the hypothalamic-pituitary-gonadal axis and thyroid hormone homeostasis (Table 4). ORD also conducted a feed restriction study to examine to what extent changes in body weight alone might interfere with interpretation of the endpoints of the assay.

There have been numerous publications in the scientific literature which used the male pubertal assay, as shown in Appendix 3. These studies, while in some cases deviating somewhat from the recommended protocol (e.g., omitting certain endpoints, starting dosing slightly later) provide valuable information on transferability, applicability to a range of endocrine interactions, and consistency of results when using this protocol.

Table 3. Chemicals examined during prevalidation and validation of the male pubertal protocol (Dose levels in mg/kg/day. Terminal bodyweight as % of controls shown in parentheses.)

Chemicals	TherImmune 1 (Transferability) block (1,3) (2,4)	TherImmune 2 (Sensitivity)	RTI (Sensitivity)	Interlab Comparison (Reproducibility) (Argus,WIL,Hunt)
Methyl Testosterone	80 (94,89) (88,90)	-	-	-
PTU	240 (44,46) (38,38)	-	2 (92) 25 (54)	-
Ketoconazole	100 (94,90) (87,93)	-	50 (88) 100 (94)	-
Pimozide	30 (85,79) (75,77)	-	-	-
Dibutylphthalate	1000 (95,93) (89,84)	-	-	500 (100,95,100) 1000 (92, 93, 99)
Atrazine	-	-	75 (88) 150 (81)	-
p,p'-DDE	-	-	50 (100) 100 (98)	-
Vinclozolin	-	10 (100) 30 (100) 100 (97)	30 (100) 100 (96)	30 (100,99,93) 100 (94,100,91)
Methoxychlor	-	-	25 (95) 50 (94)	-
Linuron	-	-	50 (93) 100 (84)	-
Phenobarbital	-	25 (97) 50 (99) 100 (92)	50 (97.5) 100 (92.5)	-
Flutamide	50 (98,93) (89,91)	25 (97) 50 (95)	-	-
DE-71	-	-	-	30 (96,98,93) 60 (97,96,95)
Chloronitrobenzene	-	-	-	25 (98,99,100) 100 (94,95,100)

TherImmune-1 blocks 1 and 3 are Sprague-Dawleys; blocks 2 and 4 are Long Evans.

Table 4. Chemicals tested by ORD laboratories to test chemicals that alter the HPG or thyroid axis

Chemicals	Dose level (mg/kg)
Atrazine	6.25, 12.5, 25, 50, 75, 100, 150, 200
Diaminochlorotriazine	Equimolar to atrazine
Desethylatrazine	Equimolar to atrazine
Desisopropylatrazine	Equimolar to atrazine
DE-71	3, 30, 60, 120, 240
Perchlorate	125, 250, 500

(Stoker *et al.* (2000a); Stoker *et al.* (2002); Stoker *et al.* (2004);Stoker *et al.* (2005) Stoker *et al.* (2006))

VI. Transferability of the protocol (TherImmune 1)

A. Purpose

The purpose of this study was to determine if the protocol is transferable. Transferability is the ability of the protocol to be accurately conducted in another laboratory by following the guidance of the assay protocol. The laboratory is assumed to have a reasonable amount of familiarity with reproductive and developmental toxicological techniques, but any specialized techniques necessary to the conduct of the assay must be described sufficiently in the guidance that no further instruction is needed.

The initial study which examined transferability of the assay from the developers to an outside laboratory was the TherImmune single-dose-level study (also referred to as "TherImmune 1" since there was a separate pubertal study, the TherImmune multi-dose-level study, discussed below, which is referred to as "TherImmune 2"). The detailed report from the TherImmune 1 study is attached as Appendix 4. Specific goals of the study were: (1) to assess the transferability of the male protocol (as it existed in 1999), (2) to assess the intra-laboratory variability in endpoint values that might be encountered in a laboratory new to the protocol, and (3) to examine the influence of two different rat strains on the sensitivity of the assay (Long-Evans Hooded and Sprague Dawley). The study was conducted under GLP by an independent, commercial laboratory (TherImmune).

Test chemicals and the dose level of each were selected by the U.S. EPA staff based upon published data demonstrating their ability to alter endocrine function (receptor agonist/antagonist, alter HPG and thyroid homeostasis). In the male, methyl testosterone, flutamide, propylthiouracil, ketoconazole, pimozide and dibutylphthalate were tested at a single high dose level based on well-known effects in historical studies (Table 3). The dose levels were selected to maximize the likelihood of demonstrating transferability: if transferability could not be demonstrated with the dose levels chosen, it was thought unlikely that the protocol could be considered transferable for any compound. (See Section XII.D for further discussion on dose selection issues.)

The TherImmune 1 study was conducted in two blocks using both Sprague-Dawley and Long Evans rats with six animals/treatment group/block. Separate vehicle controls (corn oil) were included for each block and strain. Note: this study did not measure thyroid weight or serum testosterone, as they were optional endpoints at that time.

B. Results

The most relevant results of the TherImmune 1 (single dose-level) study are provided in Table 5; the complete results are provided in Appendix 5[1].

Table 5. Summary of significant effects on major endpoints in male Sprague-Dawley and Long Evans Rats

Treatment	Mode of Action	Age at Preputial Separation	Histopath-ology	TSH	T_4
Flutamide (50 mg/kg/d)	AR antagonist	↑	✔	-	-
Methyl Testosterone (80 mg/kg/d)	AR agonist	↓	✔	-	-
Propylthiouracil (240 mg/kg/d)	Inhibitor of T_4 synthesis	↑	✔	↑	↓
Ketoconazole (100 mg/kg/d)	Inhibits steroidogen-esis	↑	-	-	-
Pimozide (30 mg/kg/d)	Dopamine receptor antagonist	↑	✔ LE	-	-
Dibutylphthalate (1000 mg/kg/d)	Anti-androgenic (not AR mediated)	↑LE	✔ LE	-	↓LE

Key: ↓= Significantly decreased compared to control
↑= Significantly increased compared to control
LE = Long Evans rats only
✔ = Affected histopathology in the appropriate organ(s) (i.e., thyroid for thyroid-active agents, epididymis/testis for androgen (ant)agonists)
AR = Androgen receptor

In general, the data obtained using the protocols successfully demonstrated transferability of the protocol: they identified the expected endocrine-mediated effects on male pubertal development following exposure to chemicals with androgenic or anti-androgenic activity, inhibitors of steroid and thyroid hormone synthesis, and a dopamine antagonist. The measure of the onset of puberty in the male rat, the age of preputial separation, was advanced following exposure to methyl testosterone and delayed by

[1] Throughout this document, an attempt was made to reduce visual clutter by presenting only the most relevant information. The complete tables with all of the endpoints are included in the Appendices.

flutamide (an anti-androgen), propylthiouracil, ketoconazole, pimozide or dibutylphthalate (Table 5).

There was, however, a difference between strains in this study. The response to DBP was less clear in the Sprague-Dawley strain than in the Long Evans strain. Strain differences are discussed further in Section XII.B. It is relevant to note here, however, that subsequent studies showed that the SD strain does respond clearly to DBP: the male pubertal assay identified DBP as interacting with the endocrine system in three independent laboratories at the same dose as used in this transferability study (as well as at a lower dose) in SD rats in the interlaboratory comparison study discussed in Section IX.D.2 (Table 22 and Table 23). The difference in the results for DBP between the transferability study and the interlaboratory comparison study in SD rats is unexplained but may be due to improvements made to the protocol between the two studies, which may have led to decreased coefficients of variation for relevant endpoints and thus greater sensitivity in the more-recent study.

There was initially thought to be a marked difference between the strains for the in-life measurement indicative of the onset of puberty in males (viz., age at preputial separation (PPS)), but this discrepancy was resolved. Although the age at PPS in the control Sprague-Dawley males was within the range expected by comparison to historical control data, the age of PPS occurred 2 - 7 days later in the Long Evans controls (Appendix 5). The age reported for PPS in Block 2 (50.2 ± 2.9 days) was of particular concern, since this advanced age at PPS has never been reported for control males in any strain. Additionally, the coefficients of variation (CV) for the mean PPS in Blocks 1 and 2 were more than 2 -fold higher in the Long Evans rats as compared with the Sprague-Dawley. Discussions with the contractor indicated that for any given day in the study, the same technician recorded the observations in both strains of rats. In addition, the contractor provided the daily observation data along with photographs describing their methods. In several of the control males that were older at PPS, the contractor found an uncommon persistent thread of tissue between the glans penis and prepuce. The age of PPS was not recorded until the thread of tissue disappeared. To determine whether or not this might be more prevalent in the Long Evans males, the contractor subsequently submitted PPS data from four additional control groups of Long

Evans rats. When the PPS data from all the Long Evans males were combined, the mean (44.3 ± 3.64) was closer to the age of PPS that was observed in the Sprague-Dawley 43.0, but the CV associated with this mean (viz., 8.23%) remained greater. Therefore, it appears that the day of onset of PPS is more variable for Long Evans rats than for Sprague Dawleys, but is still within the performance criteria that are discussed in Section VIII. What was thought at first to be a major difference between strains is apparently of little importance when analysis is based on initiation of PPS. The results in Table 5 are based on the initiation of PPS, and the protocol was modified as a result of this study to focus on initiation of PPS if persistent threads are noted.

The results of the single-dose TherImmune study were as follows:

1. Flutamide

Flutamide (4'-nitro-3'-trifluoromethyl-isobutyranilide) is a potent non-steroidal androgen receptor antagonist that has been used therapeutically to treat androgen-dependent prostate cancer (Delaere *et al.* (1991); Murphy *et al.* (1991)) and as a tool to study male reproductive development. Studies in rats have demonstrated that pre- or postnatal flutamide exposure (6.25 to 50 mg/kg) alters androgen-dependent reproductive development (Imperato-McGinley *et al.* (1992); Kassim *et al.* (1997);O'Connor *et al.* (2002)). Flutamide has been shown to decrease reproductive organ weights, feminize male external genitalia, alter androgen-dependent testicular descent, and cause retention of nipples when male offspring are exposed *in utero* (Imperato-McGinley *et al.* (1992)).

Flutamide treatment had a dramatic effect on PPS in the TherImmune 1 study, with a delay of 10 to 11 days. The dose level used (50 mg/kg/day) was a potent concentration of flutamide, but the results demonstrate that PPS can be extensively delayed. Treatment with flutamide produced weight reductions, compared to controls, in the seminal vesicles, coagulating glands, ventral prostate, dorsolateral prostate, epididymides, and/or levator ani plus bulbocavernosus muscles, and an increase in testis size.

The results of this study showed that the male pubertal assay is transferable for androgen-receptor antagonists.

2. Methyltestosterone

Methyltestosterone (17-alpha-methyl testosterone, MT) is an androgen agonist which has been employed in other *in vivo* assays to demonstrate the action of an androgen agonist. Owens *et al.* (2007) showed that 50 mg/kg of MT resulted in an increase in androgen-dependent tissue weights in the castrated male rat. In the TherImmune study, the dose of 80 mg/kg MT significantly advanced puberty 6-7 days, depending on the rat strain. It also significantly increased the weight of the ventral prostate and seminal vesicle, and decreased the weight of the testes and epididymides, likely by down-regulation of LH at the level of the hypothalamus. Therefore, this exercise demonstrated that the male pubertal assay is transferable for androgenic chemicals.

3. Propylthiouracil

6-Propylthiouracil (PTU) produces hypothyroidism by decreasing thyroid hormone synthesis via direct effects on the thyroid gland (Capen (1997), Shiroozu *et al.* (1983)). This TherImmune single-dose-level pubertal study demonstrated that the pubertal male assay is transferable with respect to this kind of thyroid-system interactivity; that is, a contract laboratory was able to detect a decrease in T_4 and increase in TSH with the high dose of 240 mg/kg by following the written protocol.

The TherImmune study also found a 9-day delay in PPS in the SD rat, which is likely due to the extreme weight loss of more than 50% below controls. PTU significantly delayed maturation of the male rats, evidenced by age at puberty, testes weights, and epididymal weights.

4. Ketoconazole

Ketoconazole is a fungicide which is well known for inhibiting steroidogenesis in both sexes. Following an exposure of 100 mg/kg of ketoconazole in the LE rats, there was an increase in adrenal weights and a decrease in epididymal, seminal vesicle and ventral prostate weights, with a delay of PPS in the first block (second had control which had persistent thread preventing complete separation of the prepuce). In the SD rats, the same androgen-dependent tissues displayed decreased weights and delayed

preputial separation in both blocks. In both the SD and LE (one block) rats, there was a consistent three day delay in PPS. Therefore, this steroidogenesis inhibitor was detected in the male pubertal protocol in both strains and in many of the androgen dependent endpoints. Transferability of the protocol for detecting interference with steroidogenesis was demonstrated.

5. Pimozide

Pimozide, a dopamine antagonist, was administered to examine the ability of the protocol to detect compounds which alter hypothalamic regulation of prolactin secretion. Pimozide should induce prepubertal elevation in prolactin secretion from the pituitary (hyperprolactinemia), which has been shown to result in decreased testicular growth and reduced or delayed androgen secretion (Maric *et al.* (1982)). In the TherImmune 1 study, there was a decrease in testicular, epididymal, seminal vesicle and levator ani/bulbocavernosus (LABC) weights, and a significant delay in preputial separation in both strains. These effects were consistent with the Maric *et al*. study and a mode of action of hyperprolactinemia by alteration of the hypothalamic control of prolactin secretion. Transferability of the male pubertal protocol for this type of interaction with the endocrine system was demonstrated.

6. Dibutylphthalate

Dibutylphthalate is a plasticizer reported to cause adverse effects on the developing male reproductive tract when administered late in gestation to pregnant rats. Administration of DBP by gavage to Sprague-Dawley rats during late pregnancy (gestation day 12–21) causes adverse effects on the developing male reproductive tract (Mylchreest *et al.* (1999)). In the high-dose group administered 500 mg/kg/day, hypospadias, cryptorchidism, agenesis of the prostate, epididymis, and vas deferens, degeneration of the seminiferous epithelium, interstitial cell hyperplasia of the testis, thoracic nipples, and decreased anogenital distance were observed. In the intermediate-dose group administered 250 mg/kg/day, agenesis of the epididymis was observed. In the low-dose group administered 100 mg/kg/day, the only effect observed was delayed preputial separation. DBP has antiandrogenic activity but is not an

androgen receptor blocker. It is believed to either inhibit steroidogenesis or directly target the Leydig cells of the testes.

In the TherImmune 1 (transferability, single-dose-level) study, dibutylphthalate produced responses in both strains of rat. The response was greater in the LE rat as compared to the SD. In the LE rat, there were decreases in all androgen-dependent tissue weights (testes, epididymides, LABC, SV and VP) and a delay in PPS while in the SD there was only decreased seminal vesicle weight in one block and increased liver weights in both blocks.

However, if one examines the individual SD data for all the reproductive endpoints (Table 6 through Table 10), it appears that appropriate changes were seen even though they did not reach statistical significance at the $p<0.05$ level. In particular, there was an anomaly in the data for the ventral prostate weights. There were three large weights in the DBP blocks, which affected the outcome. Without these animals, there would have been a 70 mg difference in the mean and the change would have been significant.

In the tables that follow, weights are in grams and the last row shows mean and standard deviation. The asterisk indicates statistical significance.

Table 6. TherImmune single dose study, seminal vesicles, SD, Blocks 1 and 2

Controls Block 1	DBP Block 1	Controls Block 2	DBP Block 2
0.465	0.465	0.540	0.383
0.621	0.559	0.272	0.490
0.441	0.277	0.337	0.277
0.327	0.287	0.395	0.357
0.775	0.339	0.387	0.366
0.656	0.428	0.503	0.350
0.548	0.393	0.406	0.371
0.164	0.110 *	0.100	0.69

Table 7. TherImmune single dose study, ventral prostate, SD, blocks 1 and 2

Controls Block 1	DBP Block 1	Controls Block 2	DBP Block 2
0.221	0.374	0.208	0.363
0.248	0.413	0.253	0.171
0.234	0.121	0.245	0.122
0.200	0.196	0.209	0.199
0.248	0.139	0.231	0.122
0.188	0.157	0.168	0.180
0.223	0.233	0.219	0.193
0.024	0.127	0.031	0.088

Table 8. TherImmune single dose study, testes (paired), SD, blocks 1 and 2

Controls Block 1	DBP Block 1	Controls Block 2	DBP Block 2
2.973	3.295	3.236	2.469
3.107	2.773	2.750	2.538
3.181	2.618	3.226	2.771
3.079	2.830	3.161	1.991
3.033	2.558	3.017	0.714
3.181	2.089	3.182	2.599
3.092	2.694	3.095	2.180
0.083	0.394	0.186	0.764

Table 9. TherImmune single dose study, LABC, SD, blocks 1 and 2

Controls Block 1	DBP Block 1	Controls Block 2	DBP Block 2
0.578	0.678	0.645	0.522
0.760	0.680	0.486	0.530
0.999	0.389	0.577	0.293
0.630	0.587	0.575	0.436
0.440	0.528	0.585	0.395
0.590	0.539	0.600	0.556
0.661	0.567	0.578	0.460
0.191	0.110	0.051	0.105

In addition, the LE rats showed a significant delay in PPS (5 days in one block and 2.5 in the other), while the SD blocks had no significant change. The first block of SD dibutylphthalate rats did have a non-significant delay in the mean age of PPS of 2 days later than the controls, but the second block showed a mean of only a half a day difference from controls (Table 10).

Table 10. TherImmune single dose study, PPS, SD, blocks 1 and 2

Controls Block 1	DBP Block 1	Controls Block 2	DBP Block 2
44	44	43	43
43	44	42	43
44	46	42	43
41	44	43	45
43	46	44	44
43	46	44	43
43.0	45.0	43.0	43.5
1.1	1.1	0.91	0.83

Since DBP produced clear effects in the LE strain and responses, though less clear, in the SD strain, this study demonstrated the transferability of the male pubertal protocol for the mechanism by which phthalates interact with the endocrine system. The subsequent interlaboratory comparison study which tested DBP in three different laboratories in the SD strain at lower dose levels than used in this study (see Section IX.D.2) confirmed the transferability of the protocol for this mechanism of action.

C. Discussion

This study demonstrated that the male pubertal assay is transferable to laboratories which have not had previous experience performing this assay. Results were generally as expected for all the test chemicals, although equivocal for the phthalate ester in one strain (a problem which was resolved in a later study). Intra-laboratory variability of organ weights was higher than expected but did not preclude identification of these relatively strong endocrine-active agents as positives. There was generally no outstanding difference between the two rat strains when all compounds (and thus mechanisms of action) were considered as a whole. Finally, the study demonstrated the relevance of the assay to the detection of interaction with the endocrine system through several different mechanisms of action.

This study also raised important questions and issues, including specific areas of concern with the execution of the protocols such as 1) the discrepancy between the ages of preputial separation identified in the two strains of rats; 2) the large degree of variation associated with the means of the fluid-filled and small tissue weights. Investigation of the causes of these problems led to improvement of the protocol for

subsequent studies -- in the first example by including instructions to use the day of initiation of PPS if separation is incomplete within three days, and in the second example by including specific instructions to avoid dessication of small organs between necropsy and weighing. Feedback from the contractor indicated other areas in the protocols where clarity and more detailed technical direction would have been helpful. The protocol was reviewed to insure that each key step was clearly described.

This transferability study was conducted using high doses of potent chemicals. Following this study it was recommended that the sensitivity of the pubertal protocols for weaker endocrine-active chemicals and lower doses be explored, to define how robust the protocols will be as a screen for identifying endocrine mediated effects.

Whether or not a particular strain of rat should be recommended for testing in the protocol remained an open question at the end of this study. Strain differences are discussed further in Section XII.B.

Following the success of the transferability study, other studies were recommended (in addition to the sensitivity study):

- A study to characterize the effects of reduced food intake and body weight on the endpoints of the male pubertal protocol.
- A study to determine whether the pubertal male protocol is capable of detecting compounds which directly target the hypothalamic control of pituitary hormone release.

VII. Sensitivity of the protocol

Two studies were undertaken to demonstrate the sensitivity of the male pubertal protocol. One study, referred to as the "multi-chemical study", focused on further examining the ability of the assay to respond to chemicals with different modes of interaction with the endocrine system, using weaker chemicals or lower dose levels than were used in the initial transferability study. The other study, referred to as the "multi-dose study", examined the response of the assay to dose levels near the lowest observed adverse effect level for a weak antiandrogen and a weak thyroid-active agent. The studies overlapped on some chemicals and dose levels both between themselves

and with the transferability study (Table 3), to provide an initial look at reproducibility of results across laboratories.

A. Multi-chemical study (RTI)

1. Purpose

The purpose of this study, conducted by RTI International, was to examine the response of the pubertal assays to the effects of a wide variety of chemicals that are known to affect the endocrine system through different pathways and/or mechanisms of action, and to obtain a sense of the ability of the assay to detect weaker chemicals. The following eight chemicals with various modes of action were tested at two dose levels: atrazine, p,p'-dichlorodiphenyldichloroethylene (p,p'-DDE), vinclozolin, methoxychlor, propylthiouracil, ketoconazole, linuron and phenobarbital.

The doses were selected in an attempt to approximate the Maximum Tolerated Dose (MTD, dose causing about a 10% decrease in body weight compared to controls by the end of dosing) and ½ MTD. Due to the size of the study it was conducted in two separate but complete blocks, each with its own control. In the first block, weanling male rats received atrazine (75 or 150 mg/kg/day), p,p'-DDE (50 or 100 mg/kg/day), vinclozolin (30 or 100 mg/kg/day), or methoxychlor (25 or 50 mg/kg/day) in corn oil. In the second block, animals received propylthiouracil (2 or 25 mg/kg/day), ketoconazole (50 or 100 mg/kg/day), linuron (50 or 100 mg/kg/day), or phenobarbital (50 or 100 mg/kg/day) in corn oil. In several cases, notably PTU and phenobarbital, MTD for juvenile/pubertal animals was not known and dose level selection was based on the best available information. For example, the dose level for PTU in the transferability study (240 mg/kg/day) led to terminal body weights that were only 40% of controls so the dose levels in this study were stepped back drastically. For phenobarbital, dose levels were based on the finding of an effect by O'Connor *et al.* (1999) at 100 mg/kg/day in adult male rats and the expectation that juvenile/pubertal animals would be at least as sensitive, or more so, to thyroid-active agents than adults.

2. Results

The summary report from this study is attached as Appendix 6 and results of this study are shown in detail in the Appendix 7 data table (means, SD and CVs). The salient results are provided in summary form at the end of this section (Table 11).

a. Atrazine

Atrazine is a chlorinated triazine herbicide used on grasses and weeds. Although the primary cellular mechanism for this compound's effects on endocrine function are not characterized, it is well established that atrazine disrupts the hypothalamic (central nervous system, CNS) control of pituitary function by suppressing the gonadotropin releasing hormone (GnRH) stimulation. In females, the ovulatory surge of luteinizing hormone is decreased in a dose dependent manner resulting in a similar dose-dependent disruption of the female estrous cycle (Cooper *et al.* (1999), Cooper *et al.* (2000), McMullin *et al.* (2003)). This mode of action was also implicated in a delay in puberty following exposure to atrazine and its primary metabolites using Wistar rats (Stoker *et al.* (2000a); delay in PPS seen at 12.5, 50, 100, 150, and 200 mg/kg/day). Atrazine was used in the present study employing Sprague-Dawley rats to confirm that this compound, purported to alter CNS control of pituitary-gonadal function, would alter pubertal development in this strain when tested in a contract laboratory.

Treatment with atrazine up to 150 mg/kg/day did not significantly alter the day of acquisition of preputial separation. However, the weights (adjusted for covariance with weaning weight) of the seminal vesicles, epididymides, and LABC muscle complex decreased at the low dose; and the same organs plus the ventral and dorsolateral prostates decreased at the high dose. No differences were noted in T_4 or TSH levels, and no treatment-related histopathological changes were observed in the thyroid, testes, or epididymides.

Although other studies have found a delay in PPS (coincident with a decrease in reproductive tract development) with atrazine with doses of 12.5 to 200 mg/kg (Stoker *et al.* (2000a), Wistar rats; Trentacoste *et al.* (2001), SD rats), this study did not repeat the effect. There was a non-significant delay in puberty at the 150 mg/kg dose of 1.5 days (41.4 in the control vs. 42.9 in the 150 mg/kg group), which is exactly the same delay as

shown in the previous study with 150 mg/kg. However, one of the control males had a late onset of PPS (PND 48 vs. mean of 41.4). When this data point was removed, the age of preputial separation exhibited an increasing trend, with a significant delay in the day of acquisition of preputial separation at the high dose (42.9 days vs. 40.8 for the control group). The low dose group achieved preputial separation at a mean age of 42.0 days, which was also delayed, but did not reach statistical significance. The average body weight on the day of acquisition of preputial separation still exhibited a decreasing trend and was significantly reduced at the high dose.

Other studies have also found a decrease in serum testosterone and intratesticular testosterone following a similar exposure (Friedmann (2002);Stoker *et al.* (2000a);Trentacoste *et al.* (2001)), but at this point testosterone was not a required endpoint in the male pubertal protocol.

The conclusion from this study was that the pubertal male assay clearly identified atrazine as interacting with the endocrine system at both dose levels, thus showing that the assay is sensitive to chemicals that affect the HPG axis when conducted in a contract laboratory. However, the lack of statistical significance for the delay in preputial separation is unexplained. The finding of an effect on PPS when the unusual control value for PPS in this study was removed, and the positive result in the Trentacoste *et al*. study in SD rats for preputial separation at 100 mg/kg and 200 mg/kg suggest that rat strain was not the problem. (The Trentacoste *et al*. study is further discussed in Section X.E.) It is also relevant to note that due to dosing mistakes, sample size in this study was only 12 animals rather than the 15 required by the protocol and the 20 used at several of the doses in the Stoker et al. paper, and this reduced sample size may have affected perceived sensitivity of the PPS endpoint.

b. p,p'-DDE

p,p'-DDE is the stable metabolite of DDT and has been shown to be an anti-androgen through competitive binding to the androgen receptor (Kelce *et al.* (1995)). In addition, Kelce *et al.* (1997) and Gray, Jr. *et al.* (1999) showed that p,p'-DDE at 100 mg/kg/day delayed preputial separation.

Treatment with 50 or 100 mg/kg/day of p,p'-DDE significantly delayed preputial separation in this study. Adjusted thyroid, liver, and paired kidney weights were significantly increased at both doses of p,p'-DDE. With respect to reproductive tissues, adjusted paired epididymides weight and LABC weight exhibited a significant decrease at the high dose.

Decreased circulating T_4 levels were observed at the high dose, whereas TSH levels exhibited no effect of treatment. No treatment-related histopathological changes were observed in the thyroid, testes, or epididymides.

The conclusion from this part of the study was that the male pubertal assay successfully detected interaction with the endocrine system for this androgen-receptor antagonist in a contract laboratory at a dose that did not cause significant bodyweight changes, suggesting that the transferability study's result for flutamide was not an isolated case and that the male pubertal assay is fairly sensitive for this kind of interaction with the endocrine system.

c. Vinclozolin

Vinclozolin, 3-(3,5-dichlorophenyl)-5-methyl-5-vinylox-azolidine-2,4-dione, is a fungicide used on fruits, vegetables, turf grass, and ornamental plants. *In vivo*, vinclozolin inhibits androgen receptor (AR)-dependent gene expression (Kelce *et al.* (1997)) and produces a spectrum of anatomical defects. Administration of vinclozolin (400 mg/kg) to rats on Gestation Day (GD) 14 through PND 3 resulted in effects similar to those caused by flutamide, a well-known AR antagonist. These effects included reduced anogenital distance (AGD); persistent nipples; cleft phallus; hypospadias; reduced weights of the ventral prostate, seminal vesicles, and epididymis; and reduced sperm count (Gray, Jr. *et al.* (1999);Kelce *et al.* (1997)). This suggests a clear interaction with the endocrine system that would be expected to be detected by the male pubertal assay. Indeed, exposing weanling male rats to vinclozolin delayed pubertal development as indicated by delayed preputial separation and increased body weight at puberty.

Vinclozolin replaced procymidone in the study design at the suggestion of an Advisory Committee, in order to allow comparison to the "multi-dose" study (discussed

below). Procymidone had originally been chosen to provide a wider variety of anti-androgenic compounds.

The day of acquisition of preputial separation exhibited significant delays at both the 30 and 100 mg/kg/day doses of vinclozolin in this study. In addition, three males in the high-dose group failed to achieve preputial separation prior to scheduled necropsy. Adjusted testes (increase) and paired seminal vesicles with coagulating glands (decrease) weights exhibited significant treatment-related effects at both dose levels. Adjusted LABC weight, adjusted paired epididymal weight, and adjusted dorsolateral prostate weight was significantly decreased at the high dose, whereas adjusted ventral prostate weight exhibited a significant treatment effect, but no pair wise differences from the control group.

Circulating T_4 levels were significantly decreased at both doses of vinclozolin, while no effect was observed on TSH. This effect on thyroid hormone homeostasis has been previously observed with vinclozolin (O'Connor *et al.* (2002);Shin *et al.* (2006)) and was expected.

No treatment related histopathological changes were observed in the thyroid, testes, or epididymides.

The conclusion from this part of the study was that vinclozolin, which works through AR antagonism, was detectable by the pubertal male assay in a contract laboratory at a dose that did not affect body weight, suggesting (as did the p,p'-DDE results discussed above) that the transferability study's results on flutamide was not an isolated result and that the male pubertal assay is fairly sensitive for this type of interaction with the endocrine system.

d. Methoxychlor

Methoxychlor is an agonist of the estrogen receptor alpha, an anti-estrogen through estrogen receptor beta, and an anti-androgen through an androgen receptor mediated mechanism (Gaido *et al.* (2000)). The overall effect from this chemical is usually associated with estrogen agonism; there were no previous *in vivo* results on which to base expectations of what might happen in pubertal males. The EDSP's Advisory Subcommittee recommended inclusion of this chemical in the male pubertal

multi-chemical study in order to determine whether similar conclusions could be reached about this chemical from both the male pubertal and the female pubertal assays.

No effect on preputial separation was noted at 25 or 50 mg/kg/day methoxychlor. Adjusted seminal vesicle with coagulating glands weight and adjusted LABC weight exhibited a significant decrease at the high dose. Thyroid hormone levels were unaffected, and no treatment-related histopathology was observed.

The conclusion from this part of the study was that methoxychlor provided evidence of interaction with the endocrine system that is consistent with estrogenicity. However, the signal was not strong. While the male pubertal assay appears to be able to detect an estrogenic chemical, it may not be appropriate to make a decision without supporting evidence from other assays.

e. Propylthiouracil (PTU)

PTU affects the thyroid directly by inhibition of thyroperoxidase in the thyroid cell, causing hypothyroidism. With exposure to 2 and 25 mg/kg of PTU, there was a decrease in terminal body weight of 8 and 45 % as compared to controls, respectively. Therefore, it is likely that the 25 mg/kg effects observed are confounded with effects from the severe inhibition of growth.

As expected, both 2 and 25 mg/kg propylthiouracil produced a decrease in the circulating T_4 and an increase in the circulating level of TSH. These doses also significantly increased adjusted thyroid weight and resulted in thyroid follicular cell hypertrophy/hyperplasia. Preputial separation was significantly delayed at the high dose, which was shown in the previous Therimmune-1 study. All of the androgen-dependent tissue weights (ventral prostate, dorsolateral prostate, seminal vesicles, epididymides, LABC complex, and testes) decreased significantly at the high dose, but the severe body weight loss compared to controls at this level (46%) may interfere with interpretation of these results.

The conclusion of this part of the study was that the male pubertal assay identified PTU as interacting with the thyroid system at a dose level which was not confounded by severe inhibition of growth (2 mg/kg/day). This provided stronger

evidence of the transferability of the assay for the detection of thyroid-active compounds than the previous study. In addition, it showed that the male pubertal assay is sensitive to low doses of strong thyroid-active agents. It is noteworthy that Yamasaki *et al.* (2002) using the male pubertal protocol and PTU at 1 mg/kg/day published results consistent with these conclusions: he reported a reduction in T_4, an increase in TSH, and increases in thyroid and pituitary weights. This is further evidence of transferability of the protocol, reproducibility of results, and sensitivity of the assay to a non-toxic level of a thyroid-active agent.

f. Ketoconazole

Ketoconazole is a broad-spectrum inhibitor of steroidogenesis in both sexes.

The postnatal day of acquisition of preputial separation was delayed at both 50 and 100 mg/kg of ketoconazole, as was shown in the TherImmune 1 study at 100 mg/kg (the only dose level tested). At the high dose, all androgen-dependent tissue weights (ventral prostate, dorsolateral prostate, seminal vesicles, epididymides, LABC complex, and testes) exhibited a significant decrease. At the low dose, the seminal vesicles and the LABC complex decreased significantly. No effect of treatment was observed histologically in the testes, epididymides, or thyroid samples. Thyroid hormone levels were not affected by either dose of ketoconazole in this study.

This part of the study demonstrated that the male pubertal assay is sensitive to a lower dose than the dose used in the transferability study for this steroidogenesis inhibitor. The results on age at preputial separation and on seminal vesicle, epididymal, ventral prostate, and adrenal weights were consistent between the two studies, thus confirming transferability. It is particularly important to note that the adrenals responded to this steroidogenesis inhibitor, as expected, and did not respond to any of the other test chemicals in this study, thus suggesting that this endpoint is indeed useful as an indicator of involvement of the steroidogenesis pathway. It is also noteworthy that in this study, all of the androgen-dependent tissues responded with a significant decrease at the high dose, while in the previous study a few of the tissues (dorsolateral prostate, LABC complex, and testes) were not shown to have statistically significant decreases. While the reason for the difference is not known, it may be due to the improvements in

the wording of the protocol drawing attention to the need to take precautions against dessication of small organs on the day of necropsy, which were inserted as a result of the experience with the first study. In any case, the results in both studies correctly identified this compound as interacting with the endocrine system, thus confirming reproducibility in two independent contract laboratories for this steroidogenesis inhibitor.[2]

g. Linuron

Linuron is an herbicide which has been shown to be antiandrogenic by competitively binding to the androgen receptor (Cook *et al.* (1993);Lambright *et al.* (2000)). Linuron was tested because it is known to be difficult to detect by other assays and was therefore thought to be a good indicator of the sensitivity of the male pubertal assay.

As expected, linuron delayed puberty at both the 50 and the 100 mg/kg/day dose levels, as evidenced by delayed acquisition of preputial separation. All of the androgen-dependent tissue weights showed statistically significant decreases at both doses with the exception of the testes, which was statistically significant only at the high dose. T_4 decreased at the high dose only, and TSH decreased at the low dose only. Pituitary weight decreased significantly at both doses, and liver and kidney decreased significantly at the high dose only. The body weight decrease (compared to controls) of 16% at the high dose may suggest that MTD was breached, but the 7% decrease at the low dose suggests strongly that interaction of linuron with the endocrine system was correctly identified without interference from body weight effects. No treatment-related histopathological changes were observed in the testes, epididymides, or thyroid.

This part of the study demonstrated that the male pubertal assay is effective in identifying antiandrogens that are difficult to detect by other *in vivo* assays.

---

[2] It should be noted that the original study plan called for additional steroidogenesis inhibitors to be tested. Attempts were made to obtain either fadrozole or letrozole, both of which are more specific to aromatase than ketoconazole is; and finasteride, which is specific to 5-α-reductase. These compounds could not be obtained for this study.

h. Phenobarbital

Phenobarbital is a commonly prescribed antiepileptic barbiturate whose hepatotoxicity and effects on the thyroid have been established. It induces the hepatic microsomal enzymes to accelerate the metabolism of endogenous hormones and exogenous xenobiotics. It has been shown to induce hypothyroidism by enhancing the clearance of thyroid hormone (McClain *et al.* (1989)). Endpoints of reproductive and developmental toxicity have not been as well explored, although some data are available. Gupta *et al.* (1980) evaluated the reproductive and developmental toxicity of phenobarbital in male rats. Males exposed *in utero* to phenobarbital from GD 12-19 (40 mg/kg/day administered to the dam) had reduced anogenital distance; delayed testes descent; decreased seminal vesicles weight; and reduced fertility. Both serum testosterone and luteinizing hormone were decreased when these males reached young adulthood. The age at preputial separation was not affected by gestational phenobarbital exposure. Wani *et al.* (1996) reported that phenobarbital inhibits steroidogenesis and causes a rapid decrease in serum testosterone.

In the current study, phenobarbital treatment delayed acquisition of preputial separation at both 50 and 100 mg/kg and the body weight at PPS was significantly increased in both dose groups. There were no effects on body weight gain, but treatment-related effects were detected at both dose levels as an increase in the adjusted weight of the liver. Thyroid weight was increased at both doses, though the increases were not statistically significant (27.3 mg in controls, 31.9 and 32.5 mg at the low and high dose levels, respectively). Testis, seminal vesicle, epididymal, and ventral prostate weights were also decreased by phenobarbital exposure at the high dose.

Serum T_4 was not different from control (5.78 ng/ml) in either the 50 mg/kg group (5.66 ng/ml) or the 100 mg/kg group (5.73 ng/ml). Serum TSH was increased in this study, rising from 11.53 ng/ml in the control rats to 16.37 ng/ml in the 50 mg/kg group and 14.47 ng/ml in the 100 mg/kg/ group, although these changes were not statistically significant at the $p<0.05$ level.

The lack of effect on terminal body weight and the increase in liver (and thyroid) weights are similar to previous reports (De Sandro *et al.* (1991);Hood *et al.* (1999);McClain *et al.* (1989)). but the lack of statistically significant effects on thyroid

hormones is not. This lack of effect appears to be due to the dose of phenobarbital selected. Marty *et al.* (2001a), using the same dose levels as were used in this study (50 and 100 mg/kg) reported a relatively small suppression of T_4, but no effect on TSH or thyroid weights in the peripubertal male. O'Connor *et al.* (1999), studying the effect of phenobarbital in the intact adult male assay, did report a decrease in T_4 (3.8 ± 0.2 ug/dl in control vs. 1.2 ± 0.1 ug/dl in the 100 mg/kg group) but a small but statistically significant increase in TSH (18.2 ± 1.2 ng/ml in control vs. 24.9 ± 1.8 ng/ml in the 100 mg/kg group). Yamada *et al.* (2004) found that a dose level of 125 mg/kg to male rats for ten days suppressed T_4 (4.3 ng/ml in controls to 2.7 ng/ml in treated) in one experiment, but the same dose was without effect in another experiment. In neither experiment was terminal body weight affected. As in the current study, Yamada *et al.* (2004) reported TSH was increased slightly (9.6 to 13.3 ng/ml) but not significantly. These observations suggest that the doses of 50 to 125 mg/kg may yield variable results. This conclusion is supported by the observation that a more consistent decrease in serum T_4 and rise in TSH follows exposure to higher doses of phenobarbital (e.g., 300 mg/kg or greater, Hood *et al.* (1999)). No histopathological changes were reported for the thyroid gland in the current study. Yamada *et al.* (2004), using a dose level of 125 mg/kg, did observe mild follicular and colloid changes in the thyroid.

The decrease in adjusted organ weight observed at the high dose for paired testes, epididymis, ventral prostate, seminal vesicles, and LABC, is consistent with the observations of Marty *et al.* (2001a) at the same 100 mg/kg/day dose level in a male pubertal study. They observed decreases in testes, prostate, and seminal vesicle weights, though not epididymides. (LABC weight was not measured in that study.) Decreases in androgen-dependent endpoints are likely due to decreases in luteinizing hormone secretion following PB exposure (Beattie *et al.* (1973)), so these changes were expected.

The conclusion of this part of the study was that the male pubertal protocol appears to be sensitive to the thyroid-related and gonadal effects of phenobarbital even though the thyroid-related responses were not significant at the $p < 0.05$ level. Since neither of the doses reached the MTD, the experiment was not a fully adequate test of the capability of the assay, but the correct trends were observed. The data from this

study are generally consistent with other published pubertal male assay studies on this compound.

The overall conclusion from this study is that the male pubertal assay is reliably sensitive to androgen-receptor-mediated anti-androgens, steroidogenesis inhibitors, and compounds which affect the HPG axis. The assay can also detect thyroid-active agents. Not all of the mechanisms of thyroid system activity were examined in this study, but the two most environmentally-relevant ones were. The results of this study for phenobarbital were not statistically significant for thyroid activity but did indicate the correct responses, and the fact that MTD was not reached suggests that the dose levels tested were not an adequate challenge for the assay.

Table 11. Organ weights and hormone results from multi-chemical study (RTI)

Chemical and dose level		PPS	VP	SV	DLP	Testes	Epi	LABC	T_4	TSH	Thyroid	Other
Atrazine	75			↓			↓	↓				↓liver
	150		↓	↓	↓		↓	↓				↓liver
p,p'-DDE	50	↑									↑	↑ liver
	100	↑					↓	↓	↓		↑	↑ liver
Vinclozolin	30	↑		↓		↑			↓			
	100	↑		↓	↓	↑	↓	↓	↓			
Methoxychlor	25											
	50			↓				↓				
PTU	2								↓	↑	↑	↓liver
	25	↑	↓	↓	↓	↓	↓	↓	↓	↑	↑	↓liver
Ketoconazole	50	↑		↓				↓				↑ adren.
	100	↑	↓	↓	↓	↓	↓	↓				↑ adren.
Linuron	50	↑	↓	↓	↓		↓	↓		↓		
	100	↑	↓	↓	↓	↓	↓	↓	↓			↓liver
Phenobarbital	50	↑										↑ liver
	100	↑	↓	↓		↓	↓	↓				↑ liver

B. Multi-dose-level study (TherImmune 2)

1. Purpose

The purpose of this study was to examine the sensitivity of the pubertal assay to a weak anti-androgen (vinclozolin) and a weak thyroid-active agent (phenobarbital) in a contract laboratory. Flutamide was chosen as the positive control for the anti-androgen.

The lowest dose for vinclozolin was chosen to be between the No Observed Adverse Effect Level in a developmental toxicity study (3 mg/kg) and the next dose level

tested, at which several effects were seen (12 mg/kg). The dose levels for phenobarbital were the same as for the multi-chemical (RTI) study, based on the adult male assay for endocrine effects (O'Connor *et al.* (1999) as discussed in Section VII.A.1)[3], plus an additional lower dose. Males were dosed with corn oil (0 mg/kg), or the test articles: phenobarbital (25, 50, and 100 mg/kg/day), vinclozolin (10, 30, and 100 mg/kg/day), or flutamide (25 and 50 mg/kg/day). Two of the chemicals overlapped with the multi-chemical study (RTI) and the other (flutamide) was run in the initial TherImmune study. The data is for this study is shown in Appendix 10 (means, SDs, CVs).

Due to an oversight, serum hormone levels (T_4, TSH, testosterone) were not obtained in this study.

2. Results

The observations made in this study were as follows and are summarized in Table 12. The full report is included as Appendix 9 and a more detailed summary including all the endpoints is available as Appendix 10.

a. Vinclozolin

Background information on this AR antagonist was presented in Section VII.A.2.c.

In the multi-dose-level study, the age of PPS increased significantly with each dose of vinclozolin in a dose dependent manner – PND 42.6, 43.9, and 47.7 for 10, 30, and 100 mg/kg/day males, respectively (1.5, 2.8 and 6.6 days delayed). Age at PPS and epididymal weight were the most sensitive endpoints for this antiandrogenic compound as they were the only effects shown with 10 mg/kg. Mean body weight at PPS was increased in the vinclozolin-treated males in direct relationship to their increased age. Vinclozolin treatment resulted in lower weights of seminal vesicles, ventral prostate, dorsolateral prostate, and levator ani plus bulbocavernosus muscles,

[3] Note: The multi-chemical and multi-dose studies were performed simultaneously so the ambiguity of results at 100 mg/kg/day, which was discussed in the multi-chemical section of this report, was not known at the time.

and higher adrenal weights at 100 mg/kg/day. Testis weight was higher than control for 30 and 100 mg/kg/day males. Thyroid weight was not changed.

No treatment-related histopathological changes were observed in the thyroid, testes, or epididymides.

This part of the study showed that the male pubertal assay can be sensitive to dose levels that are near the LOEL in a developmental toxicity study when run in a contract laboratory. In addition, this study showed consistency with the multi-chemical study at the two dose levels which overlapped: at the high dose level, seminal vesicles, dorsolateral prostate, epididymides, and LABC weights decreased and testis weights increased, at both contract laboratories; at the 30 mg/kg/day level, testis weight increased at both laboratories. Both laboratories found delays in PPS at both the 100 and the 30 mg/kg/day level. The only differences between the two labs were a finding of decreased ventral prostate weight in this study at the high dose level, and decreased seminal vesicle weight at the 30 mg/kg/day level in the multi-chemical study. Thus the identification of the chemical as interacting with the endocrine system was reproducible at both dose levels across the two contract laboratories.

b. Flutamide

Background information on this non-steroidal AR antagonist is given in Section VI.B.1.

Flutamide treatment had a clear effect on PPS. Only six of the 15 males in the 25 mg/kg/day dose group and one of the 15 males in the 50 mg/kg/day dose group achieved PPS prior to necropsy on PND 54 (a 12.2 and 12.7 day delay for the 25 and 50 mg/kg dose respectively). This effect is consistent with the results of the transferability study, in which flutamide at a 50 mg/kg/day dose produced a similar (10 to 11 day) delay in PPS. Mean body weight at PPS was increased in the both flutamide-treated groups. Both doses of flutamide also reduced the androgen dependent tissues (i.e., seminal vesicles plus coagulating glands, ventral prostate, dorsolateral prostate, epididymides, and levator ani plus bulbocavernosus muscles) and produced an increase in testis size at 25 and 50 mg/kg/day. Again, the effects in the reproductive organs at 50 mg/kg in this study were consistent with the results at 50

mg/kg in the transferability study. Liver weights were higher and kidney weights were lower at 25 and 50 mg/kg/day. Adrenal weight was higher than control in 50 mg/kg/day flutamide-treated males; however, this apparent effect is not supported by other studies in the literature. Evaluation of the individual adrenal weights revealed that the control animals in the present study had a low mean (47.87 ± 2.98 mg) with two undersized glands weighing 18.5 and 34.1 mg. Thus, the apparent difference in adrenal weights in the flutamide treated groups is likely more the result of the smaller control mean.

This part of the study confirmed that the effects of this well-characterized antiandrogen were reproducible when compared to the results from the same dose in the transferability study (performed by the same contract laboratory). It also served its purpose as the positive control for the weak antiandrogen. The lower dose of this compound also produced extreme effects on PPS.

c. Phenobarbital

Background information on this antithyroid agent is given in Section VII.A.2.h.

Phenobarbital treatment resulted in a dose-related increase in liver weight at all doses and lower pituitary, kidney, ventral and dorsolateral prostate, and seminal vesicle weight at 100 mg/kg/day. Ventral and dorsolateral prostate weights decreased at the 25 but not at the 50 mg/kg/day level.

Phenobarbital did not change thyroid weight in this study and the potential effect of phenobarbital on T_4 and TSH was not evaluated in this study. Thus, an effect of this barbiturate on the thyroid axis was not detected in the present study. Again, as in the multichemical study, the dose levels of phenobarbital used in the present experiment were not high enough to induce a change in the thyroid measures included. However, as in the multichemical study, phenobarbital was detected as interacting with the endocrine system through its effects on age at PPS. There were, however, some differences. The RTI study found a delay in PPS at both 50 and 100 mg/kg, while this study found a delay at only 100. In addition, only the RTI study found the increase in thyroid weight, and this increase was statistically significant only if unadjusted for covariance with weaning weight. Both studies found increased liver weight at all doses and a decrease in reproductive tissues, testes and epididymides at 100 in the RTI study

and in the ventral prostate, dorsolateral prostate, and seminal vesicles at 100 in this study. Body weight change at 100 was similar in both studies, with a decrease of 6 % in the RTI study (not statistically significant) and 8% (statistically significant) in this study. Others have found similar results in adults exposed for similar durations (De Sandro *et al.* (1991);Hood *et al.* (1999);McClain *et al.* (1989)) finding no change in body weight and an increase in liver and thyroid weights. Marty *et al.* (2001a) found a suppression of T_4, but no effect on TSH or thyroid weights at the same doses as these studies.

3. Discussion

This study tested the ability of the assay to detect endocrine-related effects of a weak anti-androgen and a weak thyroid-active agent at levels expected to be near the lowest observable effect level for relevant effects. The study also supplied information about the repeatability of the assay across laboratories inasmuch as each of the chemicals was tested in another laboratory (in a different study) with at least one dose in common.

The study showed that the male pubertal assay is quite sensitive for the weak anti-androgen tested, finding an effect even at a dose near the LOEL for relevant effects from a developmental toxicity study. The study also showed that effects were reproducible across laboratories for this chemical when the doses tested were the same. It was noted, however, that this reproducibility holds only for the general conclusion of interactivity with the endocrine system, not for each endpoint individually.

The results for the weak thyroid-active agent were generally consistent with the results from the multi-chemical study at the same doses in a different laboratory. However, both studies were run at dose levels that may not have reached the MTD. The lack of T_4 and TSH measurements in this multi-dose study also prevent using this study to make firm conclusions about the sensitivity of the male pubertal assay for thyroid-active agents.

Table 12. Multi-dose-level study (TherImmune 2) summary

Test article	Dose level	Body wt	PPS	Testis	Epi	VP	DLP	SV	LABC	Other wts	Histo
Vinclozolin	10	-	↑	-	↑	-	-	-	-	-	-
	30	↑	↑	↑	-	-	-	-	-	-	-
	100	-	↑	↑	↓	↓	↓	↓	↓	-	-
Flutamide	25	-	↑	↑	↓	↓	↓	↓	↓	↓kidney	-
	50	↓	↑	↑	↓	↓	↓	↓	↓	↓kidney ↑adrenal	testis
Phenobarbital	25	-	-	-	-	↓	↓	-	-	↑ liver	-
	50	-	-	-	-	-	-	-	-	↑ liver	-
	100	↓	↑	-	-	↓	↓	↓	-	↑ liver ↓pit	-

- = no effect, ↑ = higher than control means, ↓ = lower than control means

C. Special studies

EPA's Office of Research and Development (ORD) performed several additional studies to investigate the range of applicability of the pubertal assays across different modes of endocrine activity. These studies demonstrated that the pubertal assays will identify not only chemicals which alter sex steroid function but also compounds that alter hypothalamic-pituitary and thyroid function. In addition, ORD also addressed another technical concern which has generated much debate: whether or not a reduction in body weight gain during the study has a confounding effect on the endpoints associated with pubertal development. To address this concern, ORD evaluated the effects of food restriction on the endpoints in the pubertal male protocol.

1. Hypothalamic-pituitary-gonadal axis studies

Prior to the multi-chemical study, one of the concerns was whether or not the male pubertal protocol would be able to detect compounds which targeted the hypothalamic regulation of pituitary hormones. To assist with this concern, ORD scientists performed a study using atrazine, a chlorotriazine herbicide which had recently been shown to suppress hypothalamic GnRH regulation of luteinizing hormone (Table 6, Stoker *et al.* (2000a)). Atrazine, at doses of 12.5 to 200 mg/kg in the male pubertal protocol, delayed preputial separation by 1.5 to 2 days and suppressed the growth of androgen dependent tissues at 50 mg/kg and up, which has also been shown by other researchers (Trentacoste *et al.* (2001)). The primary metabolites of atrazine were also evaluated in this protocol and were also detected in the male pubertal assay with delays in puberty, decreased testosterone and decreased reproductive tract development (Stoker *et al.* (2002)). Although the male protocol appeared to be more sensitive to lower doses of atrazine than the female (Laws *et al.* (2000)), the female protocol showed more robust changes in the age of puberty with this chemical (5 day delay in vaginal opening following exposure to 200 mg/kg of atrazine as compared to 2 days in the male). Also, note that these effects were observed in two strains of male rats (Table 13).

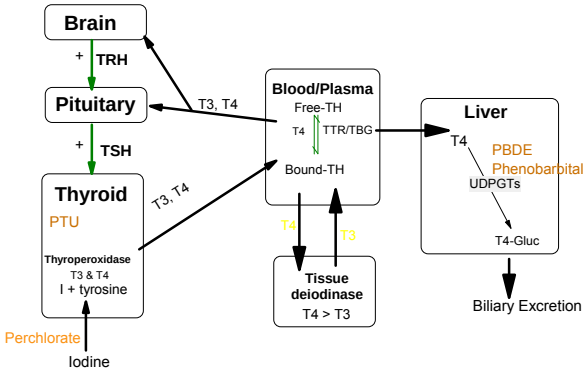
Table 13. Strain and dose comparison in male rats

Exposure	Strain	Effect	NOAEL/LOAEL	Reference
PND 23 to 53	Wistar	Delayed PPS	6.25/12.5	Stoker et al., 2000
PND 22 to 47	SD	Delayed PPS	50/100	Trentacoste et al., 2001
PND 23 to 53	Wistar	Decreased VP and SV	25/50	Stoker et al., 2000
PND 22 to 47	SD	Decreased VP and SV	50/100	Trentacoste et al., 2001
PND 23 to 53	Wistar	Non-significant decrease in T on PND 53	150/200	Stoker et al., 2000
PND 22 to 47	SD	Decreased T on PND 47	ND/50	Trentacoste et al., 2001
PND 22 to 47	SD	Decreased T on PND 47	ND/50	Friedmann, 2002

2. Thyroid axis studies

Alteration of thyroid homeostasis (thyroid hormone and TSH) can result from environmental exposures to compounds which act via a variety of mechanisms including 1) direct action on the thyroid gland by inhibition of synthesis (e.g., altered thyroperoxidase or altered uptake of iodine into the thyroid cell) or secretion of thyroid hormone (increased TSH release from the pituitary, increased TRH from hypothalamus); or 2) indirect action by inhibiting 5' deiodinase (T_4> T_3) or increased thyroid hormone clearance (by either induction of hepatic microsomal enzymes or altered TTR/TBG) (for review, see Capen (1997);McClain *et al.* (1989);Pazos-Moura *et al.* (1991) and refer to Figure 1). At this point, no environmental chemicals have been found to bind to the thyroid receptor. (See Stoker *et al.* (2000b) for review of toxicant effects on thyroid function.)

Figure 1. Mechanisms of thyroid disruption



When the Endocrine Disruptor Methods Validation Subcommittee (EDMVS) was established in 2001, one of the recommendations was to expand the number of "thyroid active" compounds that were tested in the male and female pubertal protocols. At the time, only PTU, which inhibits thyroperoxidase activity, had been evaluated in the pubertal male assay.

EPA/ORD labs evaluated thyroid compounds which alter thyroid hormones by various mechanisms in the pubertal protocols. They first added phenobarbital to the multidose and multichemical contract studies to examine the ability of the assay to

detect compounds which act by microsomal hepatic enzyme induction (increased clearance of thyroid hormone). ORD itself focused on evaluating full dose responses of several thyrotoxicants to assess the sensitivity of the assays for weaker thyroid active chemicals. The first compounds examined were DE-71 and perchlorate. Other studies, still in progress, include thiram, triclosan and 2,4-D (unpublished data).

a. DE-71

One of the compounds tested was DE-71, a polybrominated diphenyl ether (PBDE) mixture, which has been shown to induce hepatic microsomal enzymes and increase the clearance of thyroid hormone. This study was published in 2004 and another in 2005 (Stoker *et al.* (2004); Stoker *et al.* (2005)). This study showed that the male pubertal protocol was sensitive for the detection of thyroid active compounds which act by inducing hepatic microsomal enzymes and thereby increasing the clearance of thyroid hormone. The male assay was more sensitive than the female assay with this chemical, with a LOEL of 3 in the male and a LOEL of 30 in the female. Although both of the pubertal protocols identified an effect of DE-71 on the thyroid hormone concentrations and thyroid histopathology, the dose required to bring about a significant change in these hormones in the male were clearly lower than that observed in the female suggesting that the male protocol may represent a more sensitive estimate of potential thyrotoxic agents. Whether this difference was the result of differences in the two sexes or duration of testing remains to be determined. The presence of a sex difference in these measures in the prepubertal animal is in agreement with previous reports (Kieffer *et al.* (1976); Fukuda *et al.* (1975), Craft *et al.* (2002)). However, the male pubertal assay, employing a longer exposure period, did appear to enhance the ability of the assay to detect an effect at lower doses. Again, it was the male pubertal assay that appeared the most sensitive based upon the thyroid hormone data, with a LOEL of 3 mg/kg for serum T_4 concentrations and the significant change in serum T_3 and TSH concentrations following exposure to 30 and 60 mg/kg DE-71.

The results on DE-71 must be considered together with the results in the multi-chemical and multi-dose studies on phenobarbital in contract laboratories, discussed in

previous sections. The effects of phenobarbital, which acts via a mechanism similar to DE-71, was not detected by the male pubertal assay at the dose levels tested. However, it is unclear that the MTD was reached in those studies: one study reported a significant decrease in terminal body weight and the other did not, at the same dose level. The thyroid indicators trended in the expected direction, and the lack of T_4 and TSH measurements in the multi-dose study make it an inadequate basis for conclusions. The repeatability of the DE-71 results in three independent contract laboratories (see Section IX.D.3) provides further assurance that the pubertal male assay is sensitive for this mechanism of interaction with the thyroid system.

There are reasons to believe that the male may be, in general, more sensitive to thyrotoxicant exposure. In other studies, using the same dosing duration in males as in females, it has been proposed that the male rat is more sensitive than the female to perturbations of thyroid homeostasis, as indicated by an increased incidence of thyroid tumors following sub-chronic (Siglin *et al.* (2000)) or chronic exposures to environmental chemicals (Capen (1997)). Siglin *et al.* (2000) also found that adult male rats exposed to perchlorate for two weeks showed a hypothyroid hormonal profile (decreased T_4 and T_3 with an increased TSH), while the adult female cohort altered T_4, but not TSH or T_3. The basis for this difference has not been elucidated. In general the serum T_4 and TSH concentrations in the adult rat differ between the two sexes (male higher TSH and T_4) (Kieffer *et al.* (1976); Fukuda *et al.* (1975)), although in the present study only the T_4 levels were higher in the male controls. Thus, the initial difference in T_4 levels between the two sexes may contribute to the differences observed in sensitivity to DE-71.

b. Perchlorate

In another study, the ability of this EDSP male pubertal protocol to detect the known thyrotoxicant ammonium perchlorate as an endocrine disruptor was evaluated (Stoker *et al.* (2006)). Ammonium perchlorate is a primary ingredient in rocket fuel, fertilizers, paints, and lubricants. Over the past 50 years, potassium perchlorate has been used to treat hyperthyroidism in humans. Perchlorate alters thyroid hormone secretion by competitively inhibiting iodide uptake by the thyroid gland.

In this study, ammonium perchlorate was administered at 62.5, 125, 250, and 500 mg/kg to male Wistar rats. Doses of 125 to 500 mg/kg perchlorate decreased T_4 in a dose dependent manner. TSH was significantly increased in a dose dependent manner at the same doses. Thyroid histology was significantly altered at all doses, even at the 62.5 mg/kg dose, with a clear dose-dependent decrease in colloid area and increase in follicular cell height. No effects on preputial separation, a marker of pubertal progression, or reproductive tract development were observed at any dose. These results demonstrate that the male pubertal protocol is useful for detecting thyrotoxicants which target the thyroid axis by this mechanism (altered uptake of iodide). This study also found that perchlorate exposure during this period did not alter any of the reproductive developmental endpoints.

Other studies published in the literature investigating the ability of the male pubertal assay to detect thyroid-system-interactive agents are discussed below in Section X.F. The compounds tested and dose levels used are summarized in Table 14.

Table 14. Studies evaluating the effects of thyrotoxicants in the male pubertal assay

Contract studies	TherImmune 2000	PTU 240 mg/kg	
	TherImmune 2003	PB 25, 50 and 100 mg/kg	
	RTI 2003	PTU 2 and 25 mg/kg	PB 50 and 100 mg/kg
EPA studies	ORD in-house	PBDE 3, 30 and 60 mg/kg	Perchlorate 60, 120, 240 mg/kg
Other published studies	Marty *et al.* (2001a)	PB 50 and 100 mg/kg	PTU 240 mg/kg
	Yamasaki *et al.* (2002)	PTU 0.01 and 1 mg/kg	

While it may be inappropriate to make final conclusions about the strengths and limitations of the male pubertal assay to detect thyroid-system-interactive agents given that several additional studies are underway, the interim conclusion that can be made from the PTU, phenobarbital, perchlorate, and DE-71 studies is that the male pubertal

assay is sensitive to most mechanisms of thyroid-system interaction that have been tested.

3. Dose selection and body weight issues

The current protocol requires a high dose level at or just below the maximum tolerated dose (MTD), and a lower dose level at half the high dose level. As there is likely to be a minimal amount of information concerning the toxicity of many of the compounds that will be examined in the Tier 1 battery, the basis for the MTD will often be a decrease in body weight. The MTD for body weight is defined as the dose that produces a 10% reduction in body weight as compared with the appropriate control group (Hodgson (1987)). This approach assumes that a 10% reduction in body weight alone would not alter the endpoints in the pubertal assays. However, one concern with the male pubertal protocol is that many of the endpoints included may be sensitive to alterations in body weight *per se* and thus changes in body weight associated with exposure to the test chemical may confound the interpretation of the data.

There is little doubt that rigorous food restriction regimens resulting in body weight decreases of greater then 50% vs. control will produce moderate to severe reproductive alterations in organ weights, fertility, and reproductive development (Merry *et al.* (1979); Merry *et al.* (1985); Bronson *et al.* (1990); Hamilton *et al.* (1986); Widdowson *et al.* (1960); KENNEDY *et al.* (1963); Perheentupa *et al.* (1995)). However, these studies provide limited insight into how relatively small but statistically significant losses of body weight may influence reproductive and thyroid endpoints. In this regard, several studies have shown that minor reductions in daily food consumption (e.g., approximately 10% of pair-fed controls) and associated weight loss of approximately 10% or less do not alter the endpoints that are included in the proposed pubertal assays (Aguilar *et al.* (1984);Chapin *et al.* (1993);Engelbregt *et al.* (2001);O'Connor *et al.* (2000);Ronnekleiv *et al.* (1978)). These studies support the practice of using body weight as a part of the dose-setting process. However, as this assumption had not been fully examined for animals of the specific sex and age used in the pubertal protocols, EPA/ORD conducted a special study to evaluate the effect of

food restriction and body weight loss on the parameters measured in the pubertal assays (Laws et al. 2007).

Weanling male rats were provided a daily food supply that was 10, 20, 30 or 40% less than the intake of controls (determined by actually measuring the control food intake on each test day) beginning on PND 23 and continuing until PND 53. This regimen of food restriction led to weight loss (when compared to controls) of 2, 6, 9, and 19% respectively at necropsy. Importantly, there was no effect on preputial separation, even with a 19% decrease in body weight. Also, the thyroid hormone and TSH concentrations in the serum of the underfed males were not different from controls at 19% body weight loss. There was, however, a decrease in T_3 and T_4 at 9%. There were no alterations in reproductive organ weights (the androgen dependent tissues[4]) following restrictions in food intake that resulted in a 9% reduction in terminal body weight (Table 15 and Table 16).

These observations are similar to those reported by O'Connor *et al.* (2000) in the adult male: they, also, reported no significant impact of a 10% decrease in body weight on the reproductive endpoints (organ weights and hormones), although it should be noted that these results were obtained in the adult male rather than the pubertal (growing) male, and over a 15-day exposure period rather than a 30-day exposure period.

Trentacoste *et al.* (2001) reported a significant decrease in seminal vesicle weights and ventral prostate weights in a feed-restriction study using the pubertal male protocol but terminating after 25 days rather than the protocol's 30 days.

There was no change in the serum testosterone level in any of the feed restricted groups in the ORD study. This is consistent with the result in the pair-fed (restricted feed) group of Stoker *et al.* (2000a)'s atrazine study using the pubertal assay, where a body weight decrease of approximately 15% compared to controls did not result in a significant change in serum testosterone level. It is also consistent with O'Connor *et al.* (2000)'s adult male assay, where feed restriction leading to body weight decreases of

[4] Of the reproductive organs examined in the male pubertal assay, only testis, epididymis, seminal vesicle with coagulating gland, and ventral prostate were examined in this study. LABC complex and dorsolateral prostate weights were not measured.

10, 15, 21, and 26% compared to controls resulted in no significant change in serum testosterone. (These results were, however, in adult animals and from a shorter duration of exposure.) It is not consistent with Trentacoste *et al.* (2001)'s feed-restriction study in a foreshortened pubertal male assay in which a significant reduction in serum testosterone was seen when body weight reduction compared to controls was "about 10%" but the study was terminated after 25 days rather than 30 days.

Table 15. Tissue weights in male rats at necropsy in ORD food restriction study

Level of dietary restriction	Body weight at necropsy	Pituitary (mg)	Adrenal (mg)	Liver (g)	Kidney (g)
control	296.9 ± 8.3	10.84 ± 0.35	56.7 ± 1.8	13.84 ± 0.52	2.62 ± 0.08
10%[b]	290.4 ± 3.9	10.28 ± 0.36	46.7 ± 2.3 *	12.62 ± 0.35 *	2.44 ± 0.05
20%[c]	283.8 ± 8.4	9.45 ± 0.36 *	50.1 ± 1.9	11.83 ± 0.25 *[r]	2.32 ± 0.07 *
30%[d]	259.8 ± 5.1*	8.95 ± 0.34 *	46.2 ± 1.7 *	10.80 ± 0.16 *[r]	2.21 ± 0.07 *
40%[e]	235.4 ± 5.1*	8.30 ± 0.35 *	41.5 ± 1.7 *	9.30 ± 0.20 *[r]	1.91 ± 0.04 *[r]

[a] Mean ± SEM (n=13)
[b] Terminal mean body weight 2% lower than control; [c] Terminal mean body weight 6% lower than control
[d] Terminal mean body weight 9% lower than control; [e] Terminal mean body weight 19% lower than control
* Significantly different from the control ($p<0.05$).
[r] Significant treatment effect (decrease) using relative tissue weight at necropsy (<0.05).

Table 16. Hormone concentrations and reproductive tissue weights in males at PND 53

	Level of dietary restriction				
Parameter	Control	10 %	20%	30%	40%
Body weight loss relative to controls	-	2%	6%	9%	19%
T_4 (ug/dl)	4.57 ± 0.31[a]	4.42 ± 0.24	3.91 ± 0.20	3.51 ± 0.17*	3.39 ± 0.25*
T_3 (ng/dl)	99.41 ± 4.02	94.38 ± 3.85	88.37 ± 2.59	82.78 ± 3.23*	80.28 ± 2.95*
TSH (ng/ml)	1.38 ± 0.11	1.23 ± 0.13	1.40 ± 0.13	1.17 ± 0.07	0.95 ± 0.12*
T (ng/ml)	1.61 ± 0.36	2.30 ± 0.40	1.11 ± 0.18	2.31 ± 0.56	1.53 ± 0.56

	Level of dietary restriction				
Parameter	Control	10 %	20%	30%	40%
Body weight loss relative to controls	-	2%	6%	9%	19%
Ventral prostate (mg)	0.253 ± 0.014	0.280 ± 0.013	0.236 ± 0.017	0.221 ± 0.015	0.194 ± 0.010*
Left epididymis (mg)	0.250 ± 0.010	0.251 ± 0.008	0.241 ± 0.007	0.242 ± 0.008	0.224 ± 0.008*[r]
Left testis (g)	1.42 ± 0.04	1.45 ± 0.04	1.40 ± 0.05	1.39 ± 0.05[r]	1.35 ± 0.03[r]
Seminal vesicle (mg)	0.554 ± 0.041	0.650 ± 0.039	0.539 ± 0.042	0.491 ± 0.037	0.384 ± 0.031*

[a] Mean ± SEM (n=13)
* Significantly different from the control ($p<0.05$).
[r] Significant treatment effect (increase) using relative weight at necropsy ($p<0.05$)

As noted by O'Connor *et al.* (2000), a reduction in body weight can lead to a decrease in some organ weights such as the liver. In the ORD study, a statistically significant decrease was observed in the absolute pituitary, adrenal, liver and kidney weights in those groups in which the terminal body weight was 4, 6 and 9% less than controls. However, most of these significant effects disappeared when the tissues were evaluated on a relative weight basis. Importantly, it is generally accepted that certain organ weights (e.g., liver) are body-weight-dependent, and that expression on a relative-to-body-weight basis will correct for body weight decrements (Feron *et al.* (1973)).

In summary, even though the ORD study did not achieve an exact 10% loss in body weight, the results of this study do support the concept that a 10% reduction in body weight is a reasonable basis for setting the maximum dose for the male pubertal assay. More importantly, a reduction in body weight of greater than 10% should be considered too extreme. For thyroid endpoints, a body weight loss of approximately

6% relative to controls is the most that should be confidently tolerated. As body weight loss approaches 9 or 10% relative to controls, additional studies and/or a weight-of-evidence approach should be used when interpreting data for the thyroid endpoints.

VIII. Establishing performance criteria

Prior to the data analysis of the interlaboratory validation study (described below), performance criteria were established for the endpoints in the assay. These criteria provide a reference for the determining the quality of the data submitted by the participating laboratories and a means to evaluate the variability and efficacy of each endpoint.

In most contexts, "performance criteria" refer to standards of accuracy and/or precision for a positive control chemical, often a weak one, run simultaneously with the test chemical. Such criteria assure that the performing lab can detect at least a minimum signal. In the case of an apical *in vivo* assay such as the male pubertal assay, however, such an approach to performance criteria is impractical. It is not likely that any one chemical will initiate a response in all the endpoints evaluated (e.g., antiandrogens will not necessarily alter thyroid function and thyroid toxicants will not necessarily alter testicular function). Thus, testing only one mode may be insufficient to prove sensitivity to another mode. In addition, one chemical may not stimulate all of the endpoints within a single mode of action. The result is that several positive chemicals would have to be tested each time a test chemical (or set of test chemicals) is run. Considering the animal welfare concerns and other expenses associated with an *in vivo* assay of the size and duration of the male pubertal assay, it was deemed inappropriate to require such multiple weak positive controls.

As an alternative, therefore, accuracy and precision performance criteria for vehicle control animals are being used for the pubertal assay. The precision criteria are particularly important inasmuch as they help to ensure that differences between treated groups and controls can be discerned. Criteria on simultaneous controls were also considered more relevant than a periodic, non-concurrent certification program using weak positive controls.

For these reasons, we established performance criteria using historical control data. Before being considered capable of producing high quality data, the performing laboratory must demonstrate that the personnel can conduct all technical aspects of the protocol and provide control data that meet acceptable standards consistent with their own historical control data bases, as well as data published by other laboratories. The means and CVs for all endpoints should fall within the range specified by the performance criteria. Providing such information allows an independent reviewer to determine whether or not the study was conducted correctly and a basis for comparison of results across laboratories.

As a significant number of studies using the male pubertal protocol were completed prior to the time that the multi-laboratory comparison study was completed, it was possible to establish performance criteria for each endpoint within the assay. Data from the previous EPA-sponsored contract studies, data from peer-reviewed published experiments conducted by other government agencies, commercial and academic laboratories and data from unpublished studies conducted within the National Health and Environmental Effects Research Laboratory (EPA/ORD) -- the developer of the assay and the laboratory most experienced with the pubertal assay -- were included in this analysis.

Of the studies examined, approximately half of these studies employed Sprague-Dawley rats while the other half of the studies employed Wistar rats. For each strain, the mean and standard deviation for each endpoint were calculated, and the performance criterion for the mean body weight of control animals at termination was established as the interval covered by the mean ± two standard deviations in the historical controls, for the applicable strain. This is intended to cover approximately 95% of the values likely to be encountered from acceptable laboratories.

To create a performance criterion for the variability associated with the measurement of each endpoint, the coefficient of variation (CV) was also determined for each study, and the mean and standard deviation of the CVs for each endpoint were calculated. For this measure, the data from both strains was included because this measure reflects the ability of the individual laboratory (prosector) to perform each measure and is thus independent of the strain. The maximum acceptable value (i.e.,

performance criterion) for the CVs for each endpoint was set at the mean plus 1½ standard deviation. The minimum acceptable value was set at zero rather than the mean minus 1 standard deviation, since less variability is always desirable. This criterion, too, is intended to cover approximately 95% of the values likely to be encountered from acceptable laboratories.

Table 17. Performance criteria for controls

Endpoint	Rat strain	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
Ventral prostate (grams)							
	Wistar	0.223	0.072	0.151 to 0.295	16.67	5.65	22.32
	SD	0.246	0.086	0.160 to 0.332			
LABC (grams)							
	Wistar				15.77	11.33	27.10
	SD	0.651	0.204	0.447 to 0.855			
Epididymis (grams)							
	Wistar	0.474	0.124	0.350 to 0.598	10.94	5.45	16.39
	SD	0.446	0.082	0.364 to 0.528			
Seminal vesicle (grams)							
	Wistar	0.576	0.234	0.342 to 0.810	20.61	0.45	21.06
	SD	0.507	0.212	0.295 to 0.719			
Testis (grams)							
	Wistar	1.341	0.250	1.091 to 1.591	9.27	8.35	17.62
	SD						
T4 (ug/dl)							
	Wistar	5.478	2.164	3.314 to 7.642	18.27	9.20	27.46
	SD	5.716	1.660	4.056 to 7.376			
Thyroid weight (milligrams)							
	Wistar				15.39	8.24	23.63
	SD	20	6	14 to 26			

Endpoint	Rat strain	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
TSH (ng/ml)[5]							
	Wistar				34.04	24.26	58.29
	SD	14.162	9.950	4.212 to 24.112			
Age at PPS (postnatal day, where day of birth = PND 0)							
	Wistar	43.124	2.948	40.176 to 46.072	3.64	2.03	5.67
	SD	43.147	3.366	39.781 to 46.513			
Weight at PPS (grams)							
	Wistar	209.142	31.850	177.292 to 240.992	7.54	0.03	7.57
	SD	222.223	33.946	188.277 to 256.169			
Testosterone (ng/ml)							
	Wistar	2.118	2.540	0 to 4.658	58.82	30.88	89.70
	SD	2.110	1.850	0.260 to 3.960			
Final body weight (grams)							
	Wistar	291.818	41.578	250.24 to 333.396	6.62	0.85	7.47
	SD	295.647	36.412	259.235 to 332.059			
Adrenals (milligrams)							
	Wistar	54.597	13.768	40.829 to 68.365	15.42	7.34	22.77
	SD	46.478	14.636	31.842 to 61.114			
Kidneys (grams)							
	Wistar	2.516	0.550	1.966 to 3.066	9.56	5.20	14.76
	SD	2.646	0.404	2.242 to 3.050			
Liver (grams)							
	Wistar	14.070	2.874	11.196 to 16.944	10.24	4.69	14.93
	SD	12.670	2.680	9.990 to 15.350			
Pituitary (milligrams)							
	Wistar	8.051	1.934	6.117 to 9.985	12.14	3.83	15.98
	SD	10.354	2.544	7.810 to 12.898			

[5] TSH values were derived using the kit from Amersham.

Endpoint	Rat strain	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
Weaning weight (grams)							
	Wistar	58.238	11.058	47.180 to 69.296	8.04	2.21	10.25
	SD	52.642	7.170	45.472 to 59.812			

[a] Bottom of the acceptable range for coefficient of variation is zero.

IX. Interlaboratory study to examine reproducibility of the male pubertal protocol (Charles River/Argus, Huntingdon, WIL)

A. Purpose

The main purpose of the interlaboratory comparison study was to evaluate the reproducibility of a chemical's effects on the endpoints included in the male pubertal protocol. Specifically, this exercise examined the ability of three contract laboratories to produce similar results when testing the same chemicals at the same two dose levels using the written male pubertal assay protocol. Although the reproducibility of each endpoint was of interest, the main comparison of concern was whether the weight of evidence of the effects leads to the same conclusion from each laboratory concerning interaction with the estrogen, androgen, and/or thyroid systems. For the weight of evidence, EPA is not requiring that the assay consistently display a pattern of endpoint responses diagnostic for a particular mode or mechanism of action, but only that thyroid-associated responses not be used to claim consistency with sex-steroid-associated responses or vice versa. The final decision about whether there is the potential for interaction of a test chemical with the endocrine system (the goal of Tier 1 screening) is likely to be based on results from a battery of assays, not on any assay in isolation.

Another purpose was to establish whether laboratories with no previous experience with this protocol could reasonably be expected to meet performance criteria based on historical data generated, for the most part, in experienced labs. Based on

this information, another component of this exercise was to determine if the existing protocol required further optimization.

B. Dose-setting for the interlaboratory study

The EDMVS raised significant issues concerning dose selection while reviewing EPA's plans for the interlaboratory validation study. Some members of EDMVS suggested that it is necessary to see how closely different labs can determine appropriate dose levels at which to test an unknown chemical; other members suggested that the focus of a validation study should be on the capabilities of the assay itself, not on the difficulties of dose-setting. EPA decided that in the validation study, dose levels would be the same across laboratories. The reasoning was that it matters less whether laboratories come up with similar high-dose levels in an interlaboratory comparison study than whether the high dose run for a test chemical is shown in the assay itself to be the maximum tolerated dose.

This interlaboratory comparison study was conducted in three laboratories, Argus, Huntingdon and WIL labs.

C. Chemicals tested

The chemicals which were tested were the following:

Dibutylphthalate (DBP) at doses of 500 and 1000 mg/kg

Vinclozolin (Vin) at 30 and 60 mg/kg

DE-71 at doses of 30 and 60 mg/kg and

2-Chloronitrobenzene (2-CNB) at doses of 25 and 100 mg/kg.

Due to the expense of conducting large interlaboratory validation studies in a relatively long *in vivo* assay, no attempt was made in this interlaboratory validation study to test all of the various modes of action that might be detectable by the male pubertal assay, nor to establish quantitatively the limits of detection that are reproducible across laboratories for each mode. Instead, in accordance with the strategy announced soon after the EDSTAC report, this interlaboratory study focused only on demonstrating that the assay produces similar results across laboratories for a limited number of chemicals. The two chemicals DBP and vinclozolin were chosen to demonstrate

reproducibility for anti-androgens. Since no single chemical is currently known which will give a positive response in all of the male pubertal assay's endpoints simultaneously, it was necessary to test two chemicals to cover all of the endpoints. DE-71 was chosen to test the thyroid-related endpoints, and 2-CNB was chosen in an attempt to challenge the assay with a toxic but endocrinologically inactive chemical (see below). A conscious effort was made to use relatively weak chemicals that had been tested before. In the case of DBP, usually considered a strong anti-androgen, a response had not been detected in the Sprague-Dawley rat in the TherImmune transferability study so it was considered an acceptable challenge in this study, which also used SD rats.

The choice of 2-CNB deserves particular explanation. Several members of the EDMVS had raised concerns that the specificity of the male pubertal assay had not been tested -- that is, that no compound had been shown to be negative in the assay at the MTD. Without such proof of specificity, there was concern that the pubertal assay might respond to endocrine responses secondary to other toxicities in addition to direct interaction with the endocrine system. The difficulty in testing specificity, however, is that no chemicals have been tested and shown to be negative for the endpoints used in the pubertal assay -- that is, there is no standard against which the specificity of the pubertal assay could be evaluated.

In a good-faith effort to identify a chemical which caused toxicity but no endocrine effects, EPA searched the literature for reproductive and developmental toxicity studies in rats, reasoning that if other toxicities but no reproductive or developmental effects were seen in such a study, the chemical might not be interacting with the endocrine system. Few candidate chemicals were identified. In most cases, chemicals had been tested in reproductive toxicity studies because related chemicals were known to have such toxicity and not surprisingly, the test chemicals tested positive as well.

2-CNB appeared to be the best, though not ideal, candidate. It caused methemoglobinemia but only decreased spermatogenesis among all of the reproductive system endpoints evaluated in the National Toxicology Program's 13-week rat reproductive toxicity study (Bucher (1993)). This study was an inhalation study. No relevant studies in rats by the oral route were identified. In discussions with the

EDMVAC after the interlaboratory validation study of the male pubertal assay was complete, one member noted that the similarity of 2-CNB to dinitrobenzene, a known testicular toxicant, was troubling and should have suggested that 2-CNB would not be an appropriate compound to test. However, it should be noted that the isomer of dinitrobenzene that is toxic to the testis is the 1,3- isomer, and that the 1,2- isomer appears to be inactive (Blackburn *et al.* (1988)).

D. Results of the interlaboratory validation study

A detailed summary of the data from the interlaboratory validation study of the male pubertal assay is shown in Appendix 18 and the salient points are as follows:

1. Ability to meet the performance criteria

The mean and CV for each endpoint for each of the three labs in the interlaboratory study were examined to determine if they met the performance criteria (See Table 18 and Table 19). The mean CV for each endpoint by laboratory is also shown in Table 20 to allow a more detailed comparison of CVs across laboratories. Also see Figure 2 which contains 17 graphs of historical Wistar and SD data for each endpoint and compares to the three contract studies in the interlaboratory study. These data can also be referred to for the section on strain differences to demonstrate the similarities between these two strains of rats for each endpoint. The interlaboratory study used Sprague-Dawley rats in all three laboratories.

An analysis of the laboratories' ability to meet the performance criteria (accuracy and precision for each endpoint) is shown in the following Tables:

Table 18. Number of endpoints within each lab that met performance criteria for controls for coefficients of variation in the interlaboratory comparison study

Lab	Argus	Huntingdon 1	Huntingdon 2	WIL
Number of endpoints within the acceptable range for CV (0 to [Mean + 1.5 SD of historical controls])	13/17	12/17	11/17	12/17
Endpoints with high CVs	Age @ PPS Wt. @ PPS Pit VP - - -	Age @ PPS Wt. @ PPS Pit VP Adrenal - -	- Wt. @ PPS Pit VP SV Thyroid final BW	- Wt. @ PPS - VP SV Thyroid final BW

Table 19. Ability of the laboratories to meet the performance criteria in the interlaboratory study for each individual endpoint

	Argus		Huntingdon 1		Huntingdon 2		WIL	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Weaning weight	no	yes	yes	yes	yes	yes	yes	yes
Age at PPS	yes	no	yes	no	yes	yes	yes	yes
Weight at PPS	yes	no	yes	no	yes	no	yes	no
Adrenals	yes	yes	yes	no	yes	yes	yes	yes
Kidneys	yes	yes	yes	yes	no	yes	yes	yes
Liver	no	yes	yes	yes	yes	yes	yes	yes
Pituitary	no	no	yes	no	yes	no	yes	yes
Ventral prostate	yes	no	yes	no	yes	no	yes	no
Seminal vesicle	no	yes	no	yes	yes	no	no	no
Left testis[6]	yes	yes	yes	yes	yes	yes	yes	yes
LABC	no	yes	yes	yes	yes	yes	yes	yes
Epididymis	no	yes	no	yes	yes	yes	yes	yes
Thyroid weight	yes	yes	no	yes	yes	no	yes	no
Final body weight	yes	yes	yes	yes	yes	no	yes	no
T4	no	yes	yes	yes	yes	yes	yes	yes
TSH	yes	yes	yes	yes	yes	yes	yes	yes
Testosterone	yes	yes	yes	yes	yes	yes	yes	yes

Mean = the average +/- 2 times the standard deviation of the historical means.
CV = average coefficient of variation + 1.5 times the standard deviation of the CV from historical data, down to zero.
Yes = the lab met the performance requirement for that endpoint
No = the lab did not meet the performance criteria.

[6] Note: The performance criteria for left testis are based on data from Wistar rats only as the data for Sprague-Dawleys were only for paired testes.

Table 20. CVs in the interlaboratory study for each endpoint (%)

	Argus	Huntingdon		WIL
		Block 1	Block 2	
Weaning weight	9	9	8	7
Age at PPS	9	6	5	4
Weight at PPS	17	9	9	8
Adrenals	19	25	13	13
Kidneys	12	6	12	8
Liver	9	11	13	12
Pituitary	38	25	18	11
Ventral prostate	29	42	29	34
Seminal vesicle	19	17	28	22
Left testis	12	8	8	14
LABC	12	12	17	12
Epididymis	13	13	12	11
Thyroid weight	21	12	24	35
Final body weight	6	7	10	8
T4	8	8	14	20
TSH	53	44	43	28
Testosterone	54	57	65	63

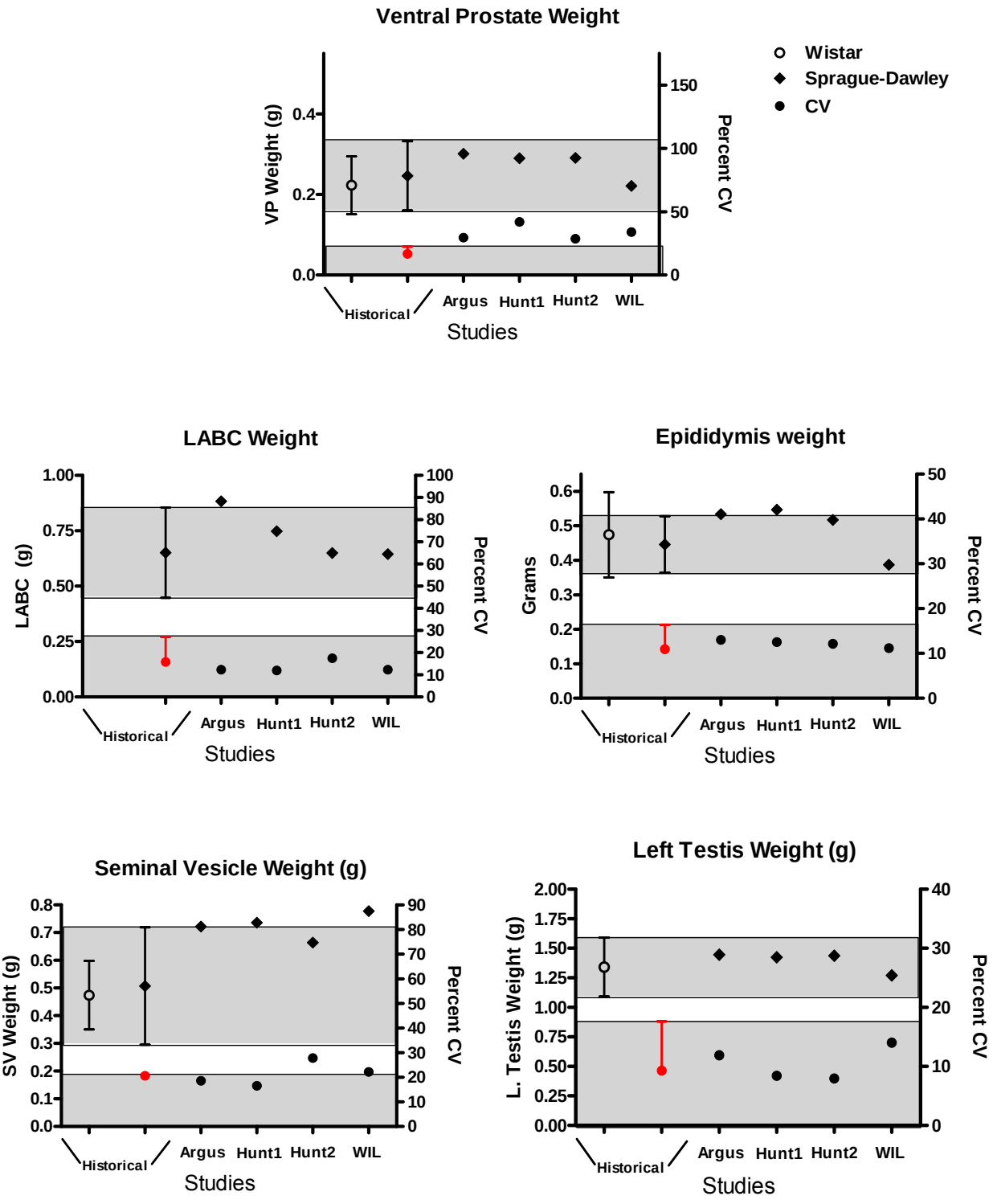
The diagrams below show how well the individual laboratories met the performance criteria. They plot the historical means +/- 2 standard deviations for each endpoint, separately for Wistar rats and Sprague-Dawley rats. To the right of the historical control data, the mean for each lab in this interlaboratory validation study is plotted. Since the rats in this study were Sprague-Dawleys, the acceptable range based on Sprague-Dawley historical controls has been shaded in.

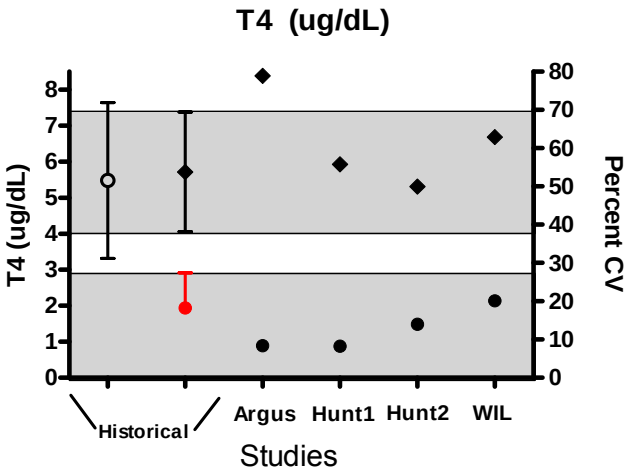
On the same graphs, the coefficients of variation for each endpoint have also been plotted. As explained in Section VIII, data from Wistars and Sprague-Dawleys were combined when setting the CV performance criteria so there is only one value for historical controls. The CVs for each laboratory for that endpoint are also plotted. The acceptable range, based on the mean CV + 1.5 standard deviations, is also shown. The acceptable range always reaches to zero since less variability than the historical controls showed is acceptable.

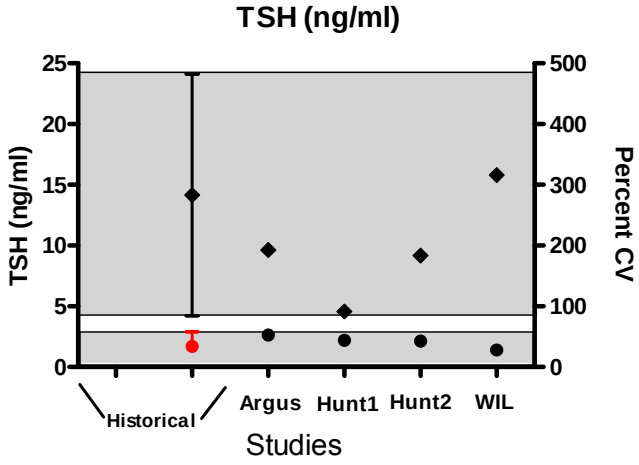
Note that the CVs (i.e., the bottom set of points) are scaled to the right axis, not the left.

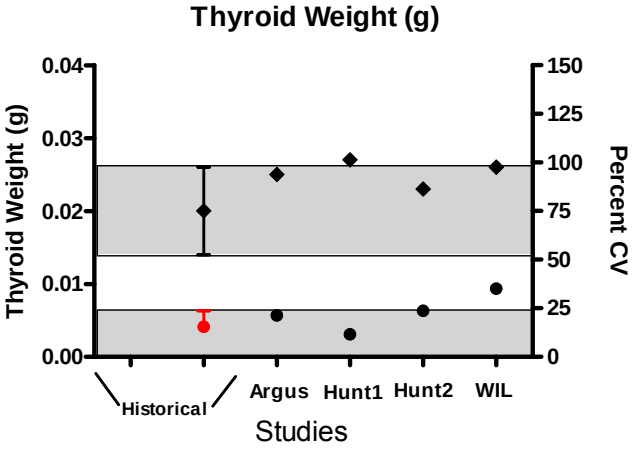
Figure 2. Performance criteria

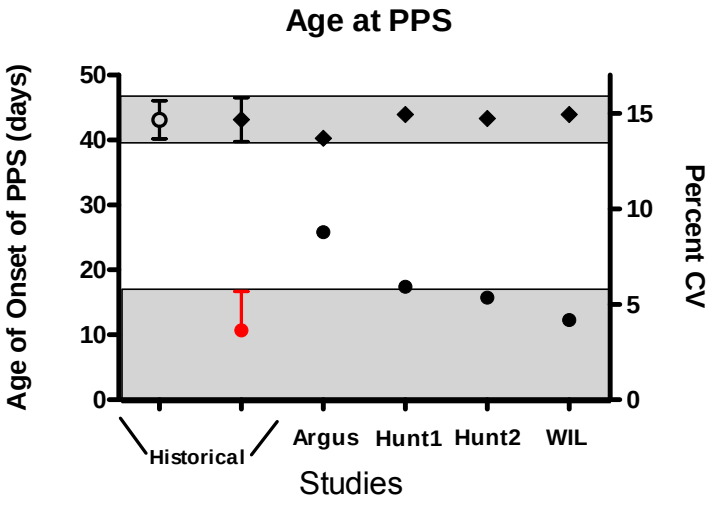
.

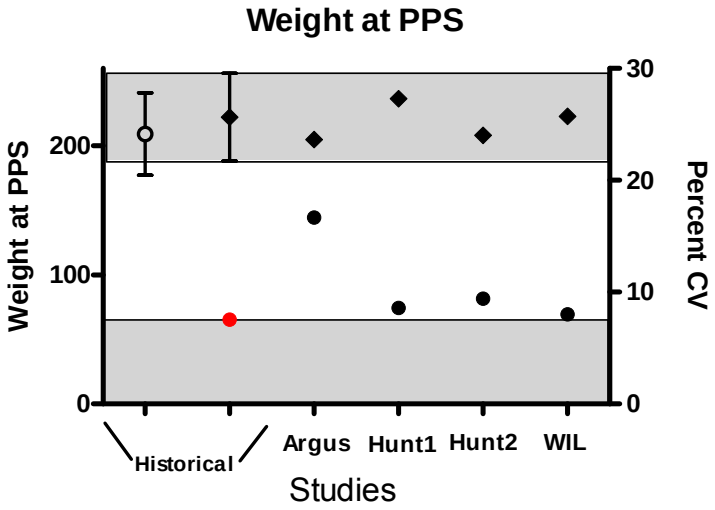


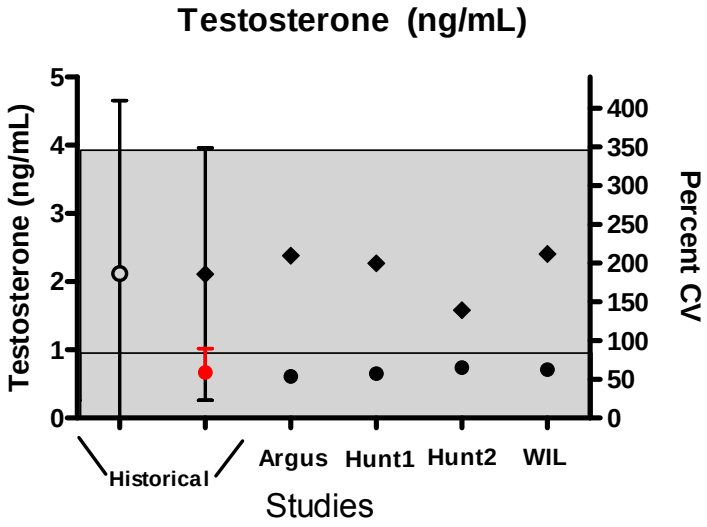

T4 (ug/dL)
T4 (ug/dL)
Percent CV
0
1
2
3
4
5
6
7
8
0
10
20
30
40
50
60
70
80
Historical
Argus
Hunt1
Hunt2
WIL
Studies

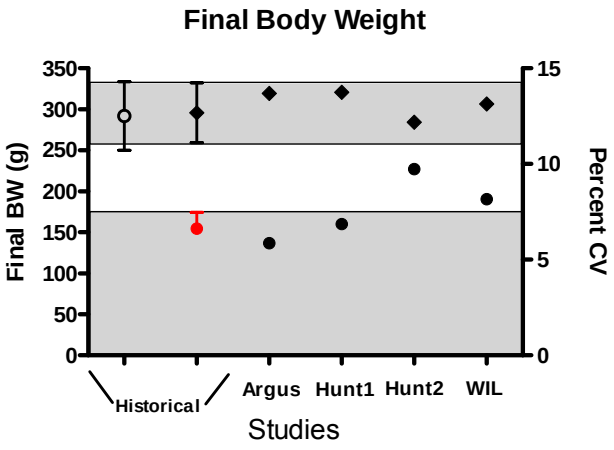

TSH (ng/ml)
TSH (ng/ml)
Percent CV
0
5
10
15
20
25
0
100
200
300
400
500
Historical
Argus
Hunt1
Hunt2
WIL
Studies

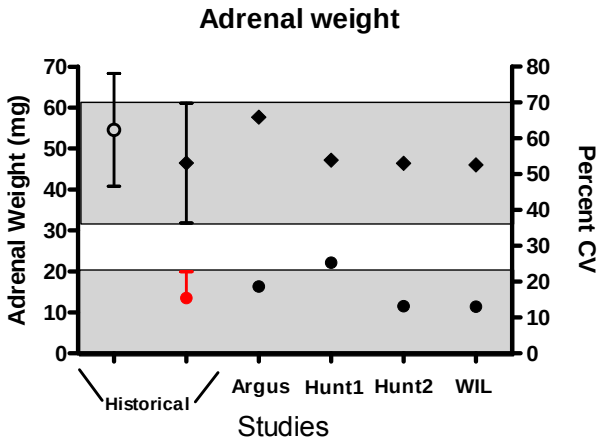

Thyroid Weight (g)
Thyroid Weight (g)
Percent CV
0.00
0.01
0.02
0.03
0.04
0
25
50
75
100
125
150
Historical
Argus
Hunt1
Hunt2
WIL
Studies

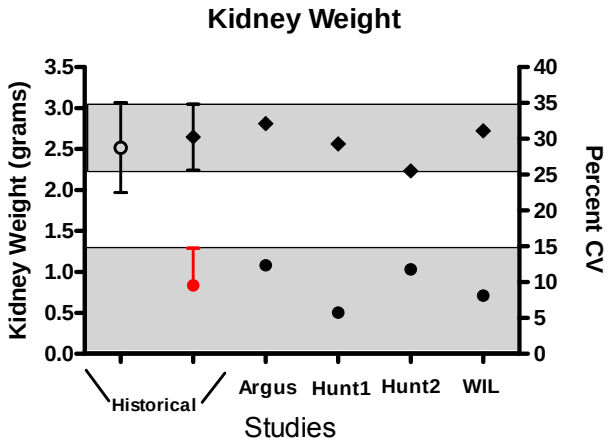

Age at PPS
Age of Onset of PPS (days)
50
40
30
20
10
0
15
10
5
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies

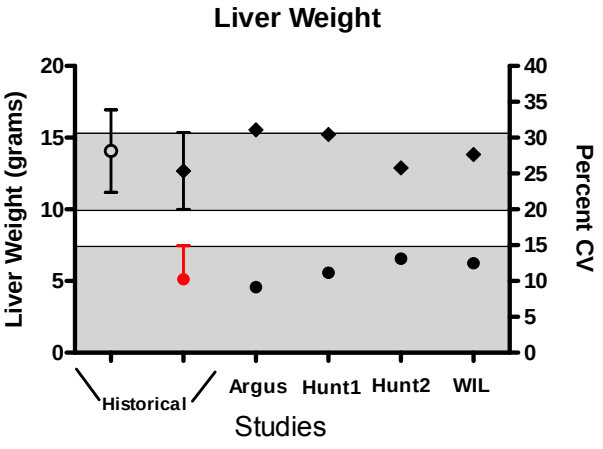

Weight at PPS
Weight at PPS
200
100
0
30
20
10
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies

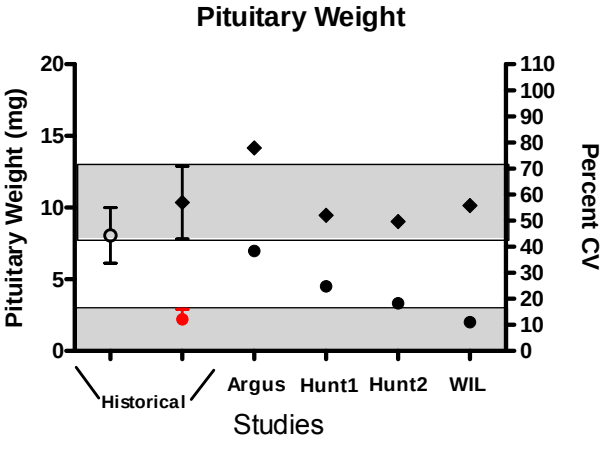

Testosterone (ng/mL)
Testosterone (ng/mL)
5
4
3
2
1
0
400
350
300
250
200
150
100
50
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies

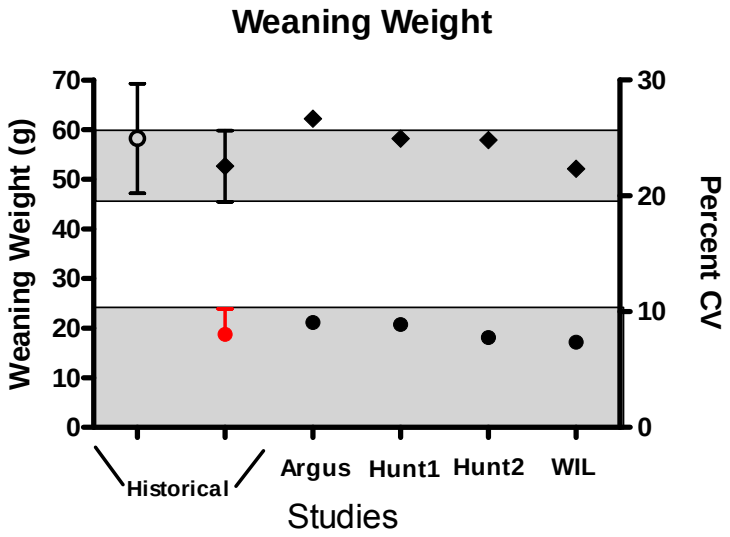

Final Body Weight
Final BW (g)
350
300
250
200
150
100
50
0
15
10
5
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies


Adrenal weight
Adrenal Weight (mg)
70
60
50
40
30
20
10
0
80
70
60
50
40
30
20
10
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies


Kidney Weight
Kidney Weight (grams)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
40
35
30
25
20
15
10
5
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies


Liver Weight
Liver Weight (grams)
20
15
10
5
0
40
35
30
25
20
15
10
5
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies


Pituitary Weight
Pituitary Weight (mg)
20
15
10
5
0
110
100
90
80
70
60
50
40
30
20
10
0
Percent CV
Historical
Argus
Hunt1
Hunt2
WIL
Studies



In general, the laboratories met most of the performance criteria but each missed certain important endpoints. None of the datasets appeared to be so compromised that it would have to be totally disqualified.

2. Dibutylphthalate

Historically, DBP has been shown to result in delayed PPS, decreased growth of androgen dependent tissues, decreased testes weights and decreased testosterone at similar doses when administered *in utero* (Foster (2006)).

In this study, two of the laboratories found significant but not excessive body weight losses compared to controls at both doses, indicating that MTD was reached but not exceeded. The third lab found no significant change in terminal body weight compared to controls, suggesting that MTD might not have been reached. This third lab still found clear evidence of interaction of DBP with the endocrine system.

Table 22 and Table 23 show the consistency of the effect of exposure to DBP on the male pubertal endpoints. The ventral prostate was not as consistent with one lab finding no change, in one lab both doses decreased ventral prostate weight and another lab found only a low dose effect on this endpoint. Testosterone was very similar with two labs finding a decrease at both doses, while the other found decreased T only at

highest dose. For SV with fluid, two labs found a significant decrease at both doses and the third found an effect only at the highest dose. Epididymis weight was highly consistent with a significantly decrease at both doses in all three labs. Also, all three labs found decreased testes weight (both absolute and adjusted) at both doses. Delayed onset of puberty was found in one lab at both doses and at only the high dose by another lab. The third lab did not detect a difference in age at PPS at either dose. T_4 concentration was significantly suppressed in all three labs at both doses. TSH was not very consistent, with one lab finding an increase at both doses, one lab only at the high dose and the other finding no change. The overall comparison of DBP across all three laboratories was good, with all three labs detecting both doses of DBP for decreased LABC weight, subsequent delayed puberty and altered reproductive tract growth.

For histopathology data on testicular and epididymal changes, the three labs were also consistent as shown in Table 21, with all labs finding degenerating seminiferous tubules at both doses and hypospermia in the epididymal sections. All three labs were consistent in finding no histopathological changes in thyroid.

Table 21. Comparison of histology results between laboratories in the interlaboratory comparison study

	Argus	WIL	Huntingdon
Testis	500 mg/kg Degeneration of seminiferous tubules, presence of intraluminal giant cells	500 mg/kg Degeneration of seminiferous tubules	500 mg/kg Increased number of sloughed germ cells
	1000 mg/kg Degeneration of seminiferous tubules, presence of intraluminal giant cells	1000 mg/kg Degeneration of seminiferous tubules	1000 mg/kg Degeneration of seminiferous tubules
Epididymis	500 mg/kg Hypospermia	500 mg/kg Hypospermia	500 mg/kg Hypospermia
	1000 mg/kg Hypospermia	1000 mg/kg Hypospermia	1000 mg/kg Hypospermia

Therefore, for DBP there was good reproducibility across the laboratories for delayed puberty and decreased reproductive tract growth. All labs detected the highest dose of DBP for decreased testosterone. Thyroid hormone effects were also reproducible between laboratories.

Table 22. Interlaboratory reproducibility of DBP, 500 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	96%*	95%*	102%
Age of PPS	↑	↑	-	-
Epididymides	↓	↓	↓	↓
Ventral Prostate	↓	↓	↓	-
SV/CG wf	↓	↓	↓	-
LABC	↓	↓	↓	↓
Testes	↓	↓	↓	↓
Testosterone	↓	↓	↓	-

Table 23. Interlaboratory reproducibility of DBP, 1000 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	92%*	93%*	99%
Age of PPS	↑	↑	-	↑
Epididymides	↓	↓	↓	↓
Ventral Prostate	↓	-	↓	-
SV/CG wf	↓	↓	↓	↓
LABC	↓	↓	↓	↓
Testes	↓	↓	↓	↓
Testosterone	↓	↓	↓	↓

The following figure (Figure 3) shows the results for all of the endpoints at both doses for all three laboratories. It demonstrates at a glance the consistency of results across laboratories. (A larger version can be found in Appendix 17 as Figure D-5.)

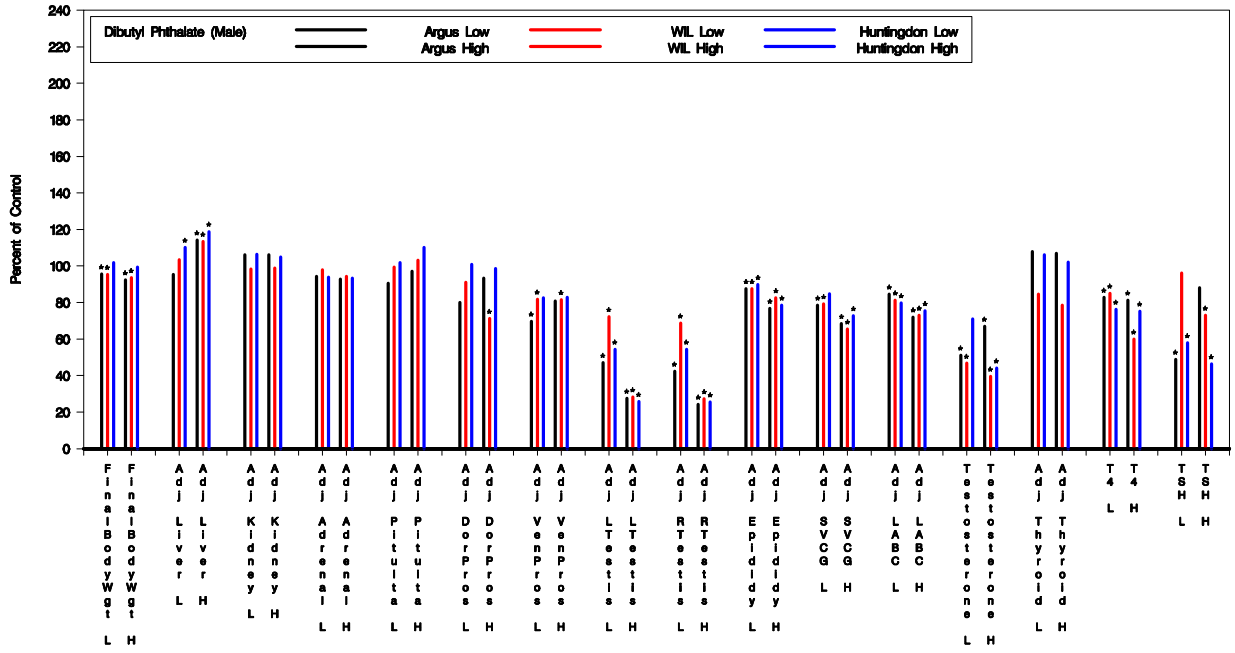


Figure 3 Male Pubertal Assay Percent of Controls for Dibutyl Phthalate versus Endpoints By Laboratory at the Low (500 mg/kg/day) and High (1000 mg/kg/day) Dose Levels (Significant Differences from Vehicle Controls the 0.05 Level are Marked by "*').

3. DE-71

Historically, DE-71 has been shown to alter thyroid hormone homeostasis in the male rat at doses similar to this study (30 and 60 mg/kg) (Stoker *et al.* (2004);Stoker *et al.* (2005)). Those studies found decreased T_4, increased TSH and thyroid histology changes at both doses and delayed PPS at both doses with decreased SV and VP at the high dose. In the interlaboratory study, the labs detected decreased T_4 and increased liver weight in all labs at both doses (Table 24 and Table 25). One of the laboratories found the delay in PPS at the high dose and the other two laboratories found decreased LABC weight (one at high dose and other at both doses). The lab that observed the decreased LABC at both doses also reported decreased testosterone at both doses. Therefore, the decreased T_4 concentrations and the increased liver weights were most consistent at both doses (all labs the same) and the TSH and thyroid histology were generally consistent between labs at the highest dose (one lab missed the change in TSH). Although the DE-71 anti-androgenic effect was not consistent between the labs, each lab did detect at least one androgen dependent change (PPS or LABC weight) at the high dose. The three laboratories were consistent in the measure

of thyroid histology, which evaluated thyroid follicular cell height and colloid area which are sensitive measures for changes in thyroid hormone homeostasis. At 30 mg/kg of DE-71, only one lab observed increased follicular height and decreased colloid area. However, at 60 mg/kg, all three laboratories observed these effects in the thyroid histology. No differences were reported in any laboratory for epididymal or testicular histopathology.

Figure 4 shows the results for all of the endpoints at both doses for all three laboratories.

Table 24. Interlaboratory reproducibility of DE-71, 30 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	97%	98%	92%*
Age of PPS	↑	↑	-	-
LABC	n/a	↓	-	↓
Ventral Prostate	-	-	-	-
SV/CG wf	-	-	-	-
Testosterone	-	-	-	↓
T_4	↓	↓	↓	↓
TSH	↑	-	↑	-
Thyroid Histo *Follicular* *Colloid*	 ↑ ↓	 - -	 ↑ ↓	 - -
Liver Wt.	↑	↑	↑	↑

Table 25. Interlaboratory reproducibility of DE-71, 60 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	97%	96%	95%
Age of PPS	↑	↑	-	-
LABC	n/a	↓	↓	↓
Ventral Prostate	↓	-	-	-
SV/CG wf	↓	-	-	-
Testosterone	-	-	-	↓
T_4	↓	↓	↓	↓
TSH	↑	-	↑	↑
Thyroid Histo *Follicular* *Colloid*	 ↑ ↓	 ↑ ↓	 ↑ ↓	 ↑ ↓
Liver	↑	↑	↑	↑

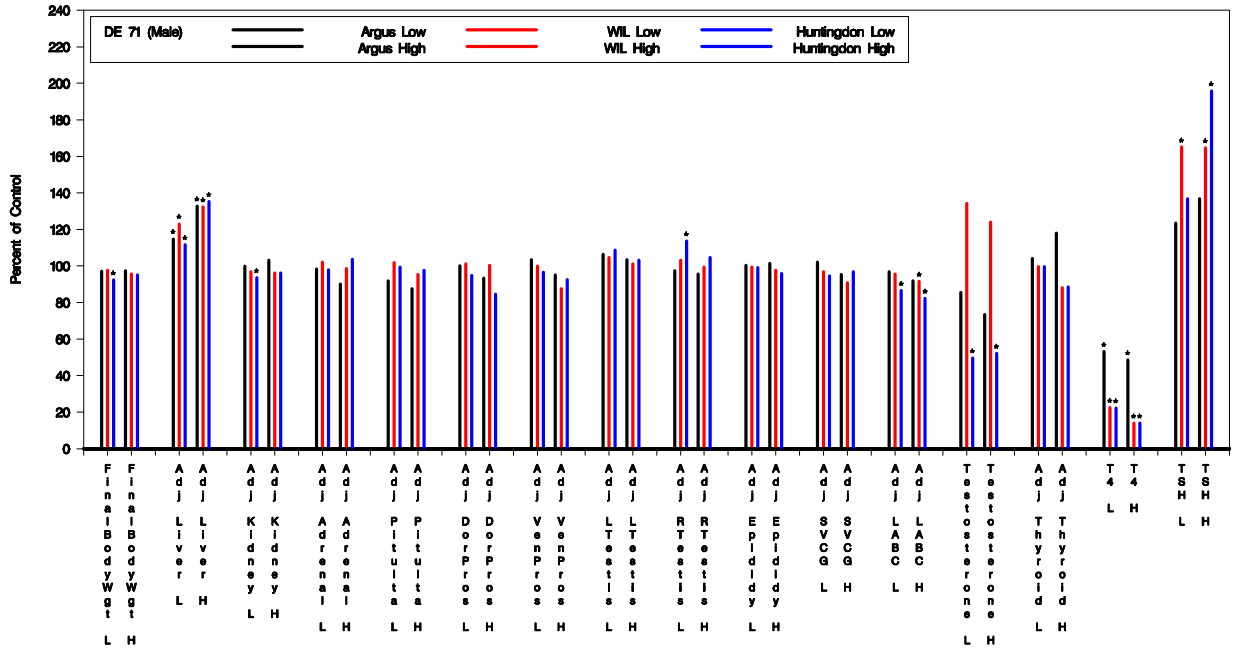


Figure 4 Male Pubertal Assay Percent of Controls for DE-71 versus Endpoints By Laboratory at the Low (30 mg/kg/day) and High (60 mg/kg/day) Dose Levels (Significant Differences from Vehicle Controls at the 0.05 Level are Marked by "*').

4. 2-Chloronitrobenzene

As explained earlier, 2-chloronitrobenzene's effects on the endocrine system were unknown but it had minimal if any endocrine-related reproductive effects in the Reproductive Assessment by Continuous Breeding program at the National Institutes of Health (Bucher (1993)). It is an oily yellow solid that is used primarily as a chemical intermediate in the production of dyes, lumber preservatives, drugs, and photographic chemicals.

In the interlaboratory study, the labs observed delayed PPS, decreased serum testosterone and decreased growth of androgen dependent tissues at the highest dose. The comparison of the responses across the three laboratories was very consistent for effects on PPS and androgen function (Table 27 and Table 28). There were also consistent effects on the liver weight (in all labs both doses increased) and fairly consistent for effects on kidney weights, with two labs observing an increase at both doses and one lab with no effect.

For histopathology, all three laboratories found degeneration of spermatids at the highest dose and two of three labs found hypospermia at the high dose (Table 26).

Figure 5 shows the results for all of the endpoints at both doses for all three laboratories.

Table 26. A summary of effects of 2-CNB on histology in the interlaboratory comparison study

	Argus	WIL	Huntingdon
Testis	100 mg/ kg/ d Slight spermatid retention, increased tubular vacuolation, degeneration of elongating spermatid population	100 mg/ kg/ d Degeneration of spermatids, Degeneration of elongating spermatids	100 mg/ kg/ d Degeneration of germinal epithelium, presence of intraluminal multinucleated giant cells
Epididymis	100 mg/ kg/ d Hypospermia and sloughing of epithelium in lumen of caput and cauda	No effects	100 mg/ kg/ d Hypospermia and sloughing of epithelium in lumen of caput and cauda

Table 27. Interlaboratory reproducibility of 2-CNB, 25 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	99%	99%	105%
Age of PPS	n/a	-	-	-
LABC	n/a	-	-	-
Ventral Prostate	n/a	-	-	-
SV/CG wf	n/a	-	-	-
Testosterone	n/a	-	-	-
T_4	n/a	↑	-	↑

Table 28. Interlaboratory reproducibility of 2-CNB, 100 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	94%*	95%*	100%
Age of PPS	n/a	↑	↑	↑
LABC	n/a	↓	↓	↓
Ventral Prostate	n/a	↓	↓	↓
SV/CG wf	n/a	-	↓	↓
Testosterone	n/a	↓	↓	-
T_4	n/a	↑	-	-

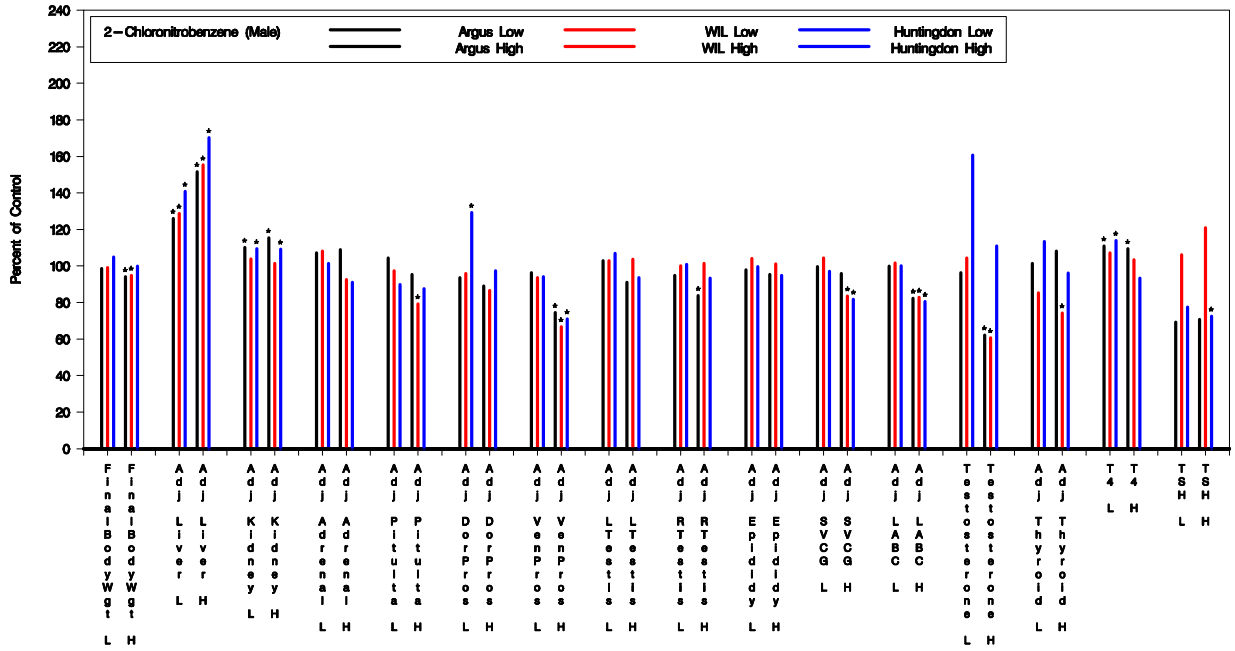


Figure 5 Male Pubertal Assay Percent of Controls for 2-Chloronitrobenzene versus Endpoints By Laboratory at the Low (25 mg/kg/day) and High (100 mg/kg/day) Dose Levels (Significant Differences from Vehicle Controls at the 0.05 Level are Marked by "*').

Therefore, for 2-CNB there was good agreement between labs on high dose effects on PPS, testosterone and androgen responsive tissues weights. The testicular and epididymal histopathology was also fairly consistent between the three labs. Also, there was good agreement on increased liver and kidney weight. The effect on the thyroid axis was inconsistent and the data are difficult to interpret.

5. Vinclozolin

Historically, doses of 30 and 60 mg/kg of vinclozolin have been shown to result in delayed puberty, decreased development of androgen dependent tissues (SV at lower doses and VP, SV, Epi, LABC, DLP at higher doses), increased testosterone at the high dose, and decreased thyroxine at both doses (Monosson *et al.* (1999).

In this study, the laboratories found delayed puberty and decreased reproductive tract development. All three labs found delayed PPS at both doses. Two labs found decreased epididymal and one found decreased seminal vesicle weights at the lowest dose (30 mg/kg), and all three labs found decreased seminal vesicles, levator ani/bulbocavernosus (LABC) and epididymal weight at the highest dose (60 mg/kg). Only one laboratory observed the expected decreased VP at the high dose. Two of the

three labs found an increase in testosterone at the high dose. Also, all laboratories found a decrease in thyroxine (T_4) serum concentrations at both doses. So overall, there was good agreement across the three laboratories. All labs detected vinclozolin at both doses for anti-androgenic effects and found the effects on thyroid hormone, and all labs found no significant changes in thyroid histology.

Figure 6 shows the results for all of the endpoints at both doses for all three laboratories.

Table 29. Interlaboratory reproducibility of vinclozolin, 30 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	99%	99	93%*
Age of PPS	↑	↑	↑	↑
LABC	-	-	-	↓
Ventral Prostate	-	-	-	-
SV/CG wf	↓	-	↓	-
Testosterone	-	-	-	-
Epididymis	↓	↓	↓	↓

Table 30. Interlaboratory reproducibility of vinclozolin, 60 mg/kg

Endpoint	Expected Change	Argus	WIL	Huntingdon
BW as % of control	-	100%	100%	91%*
Age of PPS	↑	↑	↑	↑
LABC	↓	↓	↓	↓
Ventral Prostate	↓	-	↓	-
SV/CG wf	↓	↓	↓	↓
Testosterone	↑	↑	-	↑
Epididymis	↓	↓	↓	↓

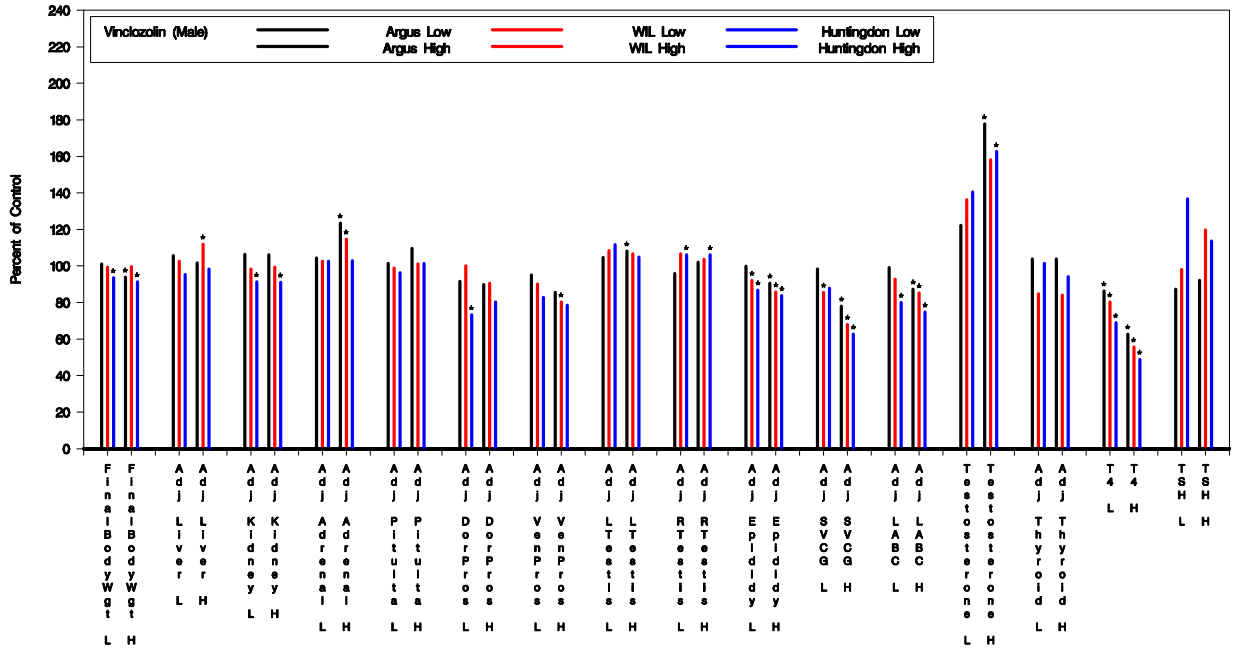


Figure 6 Male Pubertal Assay Percent of Controls for Vinclozolin versus Endpoints By Laboratory at the Low (30 mg/kg/day) and High (100 mg/kg/day) Dose Levels (Significant Differences from Vehicle Controls at the 0.05 Level are Marked by "*').

E. Conclusions of Interlaboratory Comparison Study

The current study demonstrates that the male pubertal protocol is sensitive, transferable and reproducible in contract laboratories. Results among the three labs were reproducible and using a weight of evidence approach, all of the test chemicals would have been flagged as having the potential to interact with the endocrine system in Tier 1 had they been tested as unknowns (assuming no other assays were in Tier 1). (Dose administration and data acquisition, including histopathological evaluation, were performed blind to chemical identity in this study so they were tested as unknowns, but the compounds themselves are well-characterized for endocrine interaction, with the notable exception of 2-CNB.) The SV, LABC and epididymis weights were sensitive to the altered androgenic activity following vinclozolin and DBP exposure and the effects were reproducible between labs. PPS was sensitive and reproducible following exposure to compounds which alter androgen production or activity.

All of the thyroid endpoints in this assay were reproducible following exposure to DE-71. The thyroid hormones and increased liver weight was a consistent change

following DE-71 exposure. Only one laboratory found a significant change in the thyroid histology at both the low and high dose, but none of the labs found a significant change in thyroid weight.

The thyroid histology data showed very good agreement and reproducibility across the three laboratories. Detection appears to be sensitive for hypothyroidism with compounds which induce clearance of thyroid hormone (DE-71) by induction of hepatic enzymes. The results of the testicular histology were useful for detecting DBP, which has been shown to be a Sertoli cell toxicant. In addition, the epididymal histology was useful in combination with testicular assessment to confirm the pattern of changes following exposure to DBP.

The effects of 2-CNB on the male endpoints were reproducible between laboratories, with the most consistent effects on PPS, LABC, ventral prostate, liver and testis histopathology. This study did not resolve the issue of specificity of the male pubertal assay, as had been hoped. If the results had been clearly negative, it would have been possible to suggest that the male pubertal assay can distinguish between direct interaction with the endocrine system and interaction secondary to non-endocrine toxicity. With the results clearly positive, however, and the details of the mechanisms of 2-CNB toxicity unknown, it is not possible to determine whether 2-CNB is interacting with the endocrine system directly or responding to stresses induced by other toxicities.

Each laboratory had CVs for certain endpoints which were not within the range specified by our performance criteria. One laboratory had 6 of 17 endpoints which did not meet the criteria, two had 5 that did not meet the criteria, and another laboratory had 4 which did not meet the CV criteria specified. The criterion for ventral prostate variability was violated in all laboratories and this may have contributed to the inability to detect the expected change in this endpoint for certain chemical/dose level combinations. It should be noted, however, that where the ventral prostate did detect an effect it was consistent in all labs that detected an effect and was consistent with expectations.

X. Published studies not directly part of the validation effort

Several male pubertal studies have been published in the literature. These studies were based on earlier published descriptions of the protocol (e.g., in the EDSTAC report (EDSTAC (1998)) or Stoker *et al.* (2000b)), but retain sufficient similarity to the final protocol that they lend some insight into the transferability, sensitivity, and consistency of the assay. A list of the references reviewed, and data extracted from these studies, are given in Appendix 3. Only certain highlights will be discussed here since it is not possible to discuss all of the similarities and differences in the way the studies were carried out. Discussions are arranged by putative mechanisms, even though the specific mechanisms of several of the test chemicals are not completely known.

A. Androgens

The only pubertal study in the literature that examined the applicability of the male pubertal assay to androgenic chemicals is Marty *et al.* (2001b). They found that 0.1 or 0.4 mg/kg/day (oral) of testosterone propionate (TP) did not alter the endpoints in the pubertal assay. However, the authors concluded that "oral administration limited assay sensitivity such that higher TP doses would be needed for detection" and that "[b]ecause androgens have been linked with alterations in several end points measured in the male pubertal assay, it seems likely that higher doses of TP would have been detected." For these reasons, this study is not regarded as indicative of the transferability of the assay.

B. Anti-androgens

Ashby and Lefevre's (2000) study showing delayed PPS and increased liver weights from p,p'-DDE at 100 mg/kg is consistent with the multi-chemical study's results at the same dose level[7] (Section VII.A.2.b). Ashby and Lefevre did not test at a lower

[7] Only the results of the 34-day exposure reported in the Ashby and Lefevre study were deemed relevant for this comparison to the 30-day exposure of the multichemical study. The 20-day exposures were too short for an appropriate comparison. Ashby and Lefevre's study tested several shorter exposure periods (14 and 20 days) as well as different initiation and termination days, and concluded that none of the curtailed dosing periods used in the study were able to detect all of the test agents.

dose level so there is no comparable value for the 50 mg/kg day level at which effects were seen in the multi-chemical study. No decrease was seen in epididymis weights as was seen in the multi-chemical study; and LABC weights, which were the only other weights affected in the multi-chemical study, were not reported.

Ashby *et al.* (2000) also tested dibutylphthalate at 500 mg/kg/day, which was the lower dose level used in the interlaboratory comparison study (Section IX.D.2). They reported a statistically significant 1.9-day delay in PPS, as well as decreases in epididymal and seminal vesicle weights (using unadjusted weights; analysis for covariance with weaning body weight was not performed). Final body weight was not different from controls, but liver and kidney weights were elevated. These results are generally consistent with the results in the interlaboratory comparison study. No hormone or histology reports were provided in this study.

Shin *et al.* (2002) reported a delay in PPS from flutamide at 5 and 25 mg/kg/day but not 1 mg/kg/day, in a modified pubertal male assay. The 25 mg/kg/day dose level was the same as used in the multi-dose-level study described in Section VII.B.2.b. Dosing in Shin et al.'s study did not begin until PND 33 and lasted only 20 days. Epididymis, ventral prostate, seminal vesicles, and LABC weights decreased while testis weight was not affected. Serum testosterone increased. Adrenal weight increased in the 25 mg/kg/day group. Thyroid weight was unchanged and serum T_4 levels were unaffected. TSH was not measured. The results in this study are consistent with what was seen in the multi-dose-level study, although they must be interpreted with caution since the initiation of dosing and the duration of exposure were significantly different from the initiation and duration in the multi-dose-level study.

These studies, published by independent laboratories, provide significant additional evidence that the male pubertal assay is transferable and reproducible for the anti-androgens tested.

C. Steroidogenesis inhibitors

Marty *et al.* (2001b) did not find a delay in PPS at a dose level of 24 mg/kg of ketoconazole. This dose level was lower than the lowest dose (50 mg/kg) tested in any of the studies done for this validation effort (the multi-chemical study), where such a

delay was observed. However, they did observe a significant decrease in absolute and relative epididymal weight at this dose, an effect which was seen in one of the two validation-related studies that tested this chemical, at 100 mg/kg. There was no significant difference in terminal body weight from controls even though body weight gain was significantly different from controls, suggesting that MTD may not have been reached. The results from this study are thus in agreement with the multi-chemical study results (Section VII.A.2.f) which showed an increasing number of endpoints affected as dose increased. It should be noted that terminal body weight was not affected even at the 50 and 100 mg/kg/day levels tested in the multi-chemical study, so it is not surprising that fewer endpoints were affected at the 24 mg/kg/day dose level in the Marty et al. study.

In addition, the same study provided evidence that an aromatase inhibitor (testolactone, 220 mg/kg/day) and a 5-alpha-reductase inhibitor (finasteride, 20 and 80 mg/kg/day), could be detected by the male pubertal. Testolactone delayed pubertal onset and altered the growth of androgen dependent tissues and suppressed serum testosterone concentrations, whereas finasteride resulted in decreases in androgen dependent tissue weights. However, exposure to fadrozole, a more specific aromatase inhibitor, did not affect any of the male pubertal endpoints in the same study. Body weights were significantly but not excessively decreased at both doses of fadrozole, suggesting that the dose levels were appropriate.

The ketoconazole, testolactone, and finasteride results provide supporting evidence to the information in the TherImmune 1 and multi-chemical (RTI) studies that compounds that affect steroidogenesis can be identified by the male pubertal protocol. They also provide evidence that the protocol is transferable and reproducible. However, the results with fadrozole suggest that compounds specific to aromatase inhibition will not be detected with this assay.

D. Estrogens

Ashby *et al.* (2000) found a 10.9-day delay in PPS using diethylstilbestrol at 40 ug/kg/day for a 34-day exposure period using the male pubertal assay. Seminal vesicle, epididymides, testes, and prostate weights declined significantly, even though

there were only 10 animals per group; final body weight was 85% of body weight in controls, and liver weight also decreased. Analysis was done on absolute weights only; covariance with weaning body weight was not reported.

Tan *et al.* (2003) studied nonylphenol and bisphenol A (BPA) in a male pubertal assay, each at 100 mg/kg/day. Their conclusion was that both nonylphenol and BPA delayed PPS, judging from the number of animals which did not show PPS by the end of the study at PND 55 (0, 8, and 4 for controls, nonylphenol, and BPA respectively). Analysis of PPS was done differently from the method specified by the protocol, so a direct comparison is not possible. It should be noted that the mean age of PPS in controls (PND 48.4 ± 2.5) was outside of the performance criteria described in Section VIII (43.12 ± 2.95). Nonylphenol decreased seminal vesicle weight significantly but did not affect testis or epididymis, the only other two androgen-related organs measured. Liver weight was increased compared to controls. BPA caused no increases in androgen-related organ weights, but caused increases in kidney, liver, and thyroid weights. BPA also led to moderate to severe hydronephrosis and testicular histopathological changes. Body weights were not statistically different from controls, and adrenal weights were unaffected.

The Ashby *et al.* (2000) study suggests that the male pubertal assay responds to strong estrogens, while the Tan *et al.* (2003) study provides support that weak estrogens may also cause effects. The conclusions from the latter study, however, must be interpreted carefully given the kidney effects seen from BPA exposure, suggesting that MTD may have been exceeded.

E. HPG axis

Trentacoste *et al.* (2001) tested atrazine at doses similar to Stoker *et al.* (2000a) in a male pubertal study that included most but not all of the reproductive organ weights and age at PPS, but which terminated at 25 days of exposure rather than 31 days. They attributed the differences in SV and VP weights between treated and controls to changes in growth rather than to the chemical, as evidenced by results of a separate food restriction study they also reported. This is consistent with the Stoker *et al.* (2000a) study, which included a paired-feeding component to examine feed restriction.

Effects on testis and epididymis weights were not significant in either study. The Trentacoste *et al.* (2001) study thus confirms the reproducibility of results across laboratories and across strains (SD. vs. Wistar) for the reproductive organ weights. However, it is unclear from the Trentacoste *et al.* (2001) report whether the delays in PPS which were seen at 100 and 200 mg/kg (approximately 3 and 4 days respectively) were also attributable to feed restriction: data on PPS from the feeding study component were not reported. In the Stoker et al. study, the delay in PPS at 200 mg/kg could not be attributed to reduced growth, although it remained a possibility at 100 mg/kg. The Trentacoste *et al.* (2001) study appears to confirm that the pubertal male assay is sensitive to atrazine, a chemical which affects the HPG axis, that strain differences are not significant for this chemical, and that the protocol is transferable.

F. Thyroid-related activity

Yamasaki *et al.* (2002) found that 1 mg/kg of PTU was sufficient to decrease T_3 and T_4, increase thyroid weight and increase thyroid follicular cell hypertrophy when exposed orally to SD rats in the pubertal protocol. This is completely consistent with the results of the RTI multi-chemical study at 2 mg/kg of PTU (Section VII.A.2.e, Table 11) and demonstrates that the male pubertal assay is both transferable and consistent across laboratories for (this type of) thyroid interaction at a dose that is not otherwise overtly toxic.

PTU at 240 mg/kg has also been shown to increase TSH, decrease T_4, cause thyroid follicular cell hypertrophy, and delay puberty in the male SD rat in a study by Marty *et al.* (2001a) using the male pubertal protocol. These results are identical to the results seen in the TherImmune 1 (transferability) study at the same high dose, thus supporting the transferability and reproducibility of effects for this mechanism of thyroid activity across laboratories. Marty *et al.* (2001a) also found a 7 day delay in PPS. This extreme delay is consistent with the 9-day delay seen in the TherImmune transferability study using the same 240 mg/kg dose. Both Marty *et al.* (2001a) and the TherImmune 1 study found a significant decrease in absolute testes and epididymal weights. Although little can be deduced about the anti-androgenicity of PTU due to the extreme (~50%)

body weight difference from controls in both studies at this dose, the consistency of effect on the relevant endpoints across laboratories is noteworthy.

G. Other mechanisms

The Detailed Review Paper (Appendix 2) discusses the relationship of dopamine, thyrotropin releasing hormone, and prolactin secretion, and the relationship between prolactin and male pubertal develoment. Marty *et al.* (2001a) tested a dopamine antagonist (haloperidol, 2 and 4 mg/kg) and a dopamine agonist (bromocryptine, 10 and 50 mg/kg). Statistically significant decreases in terminal body weight when compared to controls were seen at both doses for both chemicals but were not excessive. The high dose of bromocryptine resulted in delayed PPS, a decrease in absolute prostate and seminal vesicle weights, and a decrease in absolute thyroid weight. (Adjustment for covariance with weaning weight was not reported.) Haloperidol caused a decrease in absolute liver weight compared to controls, but not relative to body weight. Haloperidol also caused a decrease in absolute thyroid weight compared to controls, at both doses. T_4 was decreased but TSH was not different from controls, at both doses. No statistically significant effects were seen in serum testosterone levels, from either haloperidol or bromocryptine although the high CV may have prevented detection of what otherwise appears to be a 50% reduction in the mean from haloperidol exposure. Thyroid histopathological changes with haloperidol were minimal. This study was done in two blocks of 10 males per treatment group, which were then combined prior to analysis based on comparability of control values for weanling weight, body weight at PPS, age at PPS, and terminal body weight.

This study showed the male pubertal assay's ability to detect the dopamine agonist bromocryptine, which decreases prolactin levels. The assay did not detect the dopamine antagonist haloperidol, which increases prolactin levels, through effects on androgen-dependent organs; but it may be significant that prolactin levels themselves, though apparently elevated, were not statistically significantly elevated in this study. Haloperidol was detected as interacting with the thyroid system. The results with haloperidol are not consistent with the finding of delayed PPS and decreased testis, epididymis, seminal vesicle, and LABC weights from pimozide in the transferability

study (TherImmune 1, Section VI.B.5), although the transferability study's results may have been complicated by the significant loss in body weight (15 to 25% compared to controls).

XI. Data interpretation

The male pubertal assay is intended to be one of a suite of *in vitro* and *in vivo* assays for determining the potential of a substance to interact with the endocrine system. Therefore, it is important to emphasize that the data interpretation of a specific chemical will be a combination of the results from a number of these Tier-1 screening assays taken as a whole and not merely the sum of results of assays interpreted in isolation. That said, there are certain guidelines that can be given for interpreting data from a male pubertal assay.

First, the dose levels tested should be examined to see if a Maximum Tolerated Dose was used. Indications that MTD was approached but not exceeded include clinical observations and/or body weight loss compared to controls at termination that does not exceed approximately 10%. Histopathology of the kidney and/or other organs, and/or significant deviations from standard blood chemistry values may be indications that MTD was exceeded.

Negative results for interaction with the endocrine system in the pubertal assay will generally require demonstration that the highest dose level tested was at or near the MTD. Positive results in the assay generally require no such proof, but will generally require demonstration that interference due to body weight loss *per se* was not a factor in generating the results. Thus, studies that suggest thyroid activity only at a dose level causing more than approximately 6% body weight loss at termination compared to controls may need to be repeated at a lower dose level. Similarly, studies which suggest interaction with non-thyroid endocrine systems only at a dose level that causes more than approximately 10% body weight loss at termination compared to controls may need to be repeated.

The endpoint values for the control group should be compared to the performance criteria. Comparison should be made on the basis of the measured values, not adjusted values. Any endpoints which do not meet the performance criteria

in controls will generally be disregarded for the test chemicals if they are negative but may provide useful information if they are positive.

Information that is missing due to inability to meet a performance criterion is not the same as a negative result. The more endpoints that are missing, the less likely the study will be regarded as adequate. No firm rules can be given for the minimum number of endpoints that must be available for evaluation since some of the endpoints are somewhat redundant (e.g., androgen-dependent tissue weights) while others are not (thyroid-related endpoints). In general, however, missing one or two performance criteria will not be regarded as fatal to the study.

More emphasis will be placed on meeting performance criteria for the coefficients of variation than for the endpoint control means. Laboratories may submit historical data for their own colonies to substantiate claims that tissue weights or other endpoints in the study being evaluated are in line with historical values of controls in that laboratory.

Once the usable data set has been identified through application of the performance criteria, it is evaluated to see if there is evidence of interaction of the test chemical with the endocrine system.

Due to the covariance of certain organ weights with body weight, care should be taken in interpreting pituitary, liver, and kidney weight changes. Only if a change in the organ weight relative to body weight is significant for these particular organs (i.e., not all the organs) should the weights adjusted for covariance with body weight at weaning for these particular organs be interpreted as relevant.

Endpoints other than pituitary, liver, and kidney weights should not be evaluated based on their values relative to terminal body weight, nor should an analysis of covariance with terminal body weight be used for interpretation. Since endocrine-active agents themselves may have an effect on body weight, it is most appropriate to adjust for covariance with body weight at weaning, before chemical treatment began.

The male pubertal data described in this report provide general profiles of changes in the assay endpoints for various modes of action such as androgen agonism, androgen antagonism, alteration of steroidogenesis, thyroid toxicity, and interference with HPG function. These profiles can be used to establish a “weight of evidence” for

general mechanisms of interaction of a test chemical with the endocrine system. For example, an antiandrogen such as vinclozolin delays puberty, impairs reproductive tract development (e.g., decreased VP, SV, LABC, epididymis weight) and increases testosterone at higher doses, so a test chemical with similar responses would likely be suspected of having an antiandrogenic interaction. A similar profile would be expected if the compound inhibits testosterone synthesis. One way to discern a compound that inhibits steroidogenesis from one that is antiandrogenic is to evaluate serum testosterone (a required endpoint) as this endpoint will obviously be decreased.

The pubertal male assay includes redundant androgen-dependent endpoints and in general, all would be expected to respond similarly. However, it is possible that a chemical may not alter all the endpoints measured, or the effect may not be dose dependent, or it may occur only at the high dose. In these kinds of cases, data from the other proposed assays in the Tier 1 battery would provide added insight into a potential effect. For example, any effect of the compound on the androgen receptor should also be indicated by the data from the androgen receptor binding assay or the Hershberger assay. EDSTAC recommended a certain amount of redundancy across assays.

Another example of how a chemical would produce a particular profile would be the way in which the thyroid homeostasis is disrupted. For example, the polybrominated diphenyl ether mixture, DE-71, disrupts thyroid hormones by inducing the clearance of thyroxin by hepatic enzyme induction. The ensuing profile of effects includes a decrease in T_4, decrease in T_3 (although not required in the male pubertal) and a subsequent increase in TSH. In many cases, thyroid weight and thyroid histology appear less sensitive than the changes in the hormone concentration because the hormonal changes can occur more rapidly than any frank change in histology. For example, DE-71 exposure in the interlab study only one lab observed histological changes associated with the hypothyroidism, yet the endocrine effect was seen in all three laboratories.

Other chemicals may target the HPG axis and there are certain profiles that may indicate altered brain-pituitary function. For example, a chemical may delay puberty, lead reduced androgen dependent tissues, yet be negative in a androgen receptor binding assay, negative for alterations in steroidogenesis and the Hershberger. This is

the profile that was observed following atrazine exposure (Stoker *et al.* (2000a);Stoker *et al.* (2002)). Again, the interpretation of the results of the male pubertal assay will be enhanced using a weight of evidence from all the assays included in the Tier 1 battery.

The results of the interlaboratory study with 2-chloronitrobenzene are an excellent example of the patterns or profiles of effects that may be observed with an unknown chemical in the screening process. For example, there was a decrease in serum testosterone and a decrease in the growth of the androgen dependent tissues. Therefore, 2-CNB is either altering steroidogenesis or targeting the hypothalamic/pituitary secretion of pituitary hormones, but is not an antiandrogen through receptor binding activity. If that were the case, testosterone would have been increased.

Examples of endocrine profiles which can be identified in the male pubertal protocol are shown in Table 31. It should be noted that consistency with a known "profile" is not a requirement for determining that a test chemical interacts with the endocrine system, nor is consistency among supposedly redundant endpoints. As shown in the interlaboratory validation study, laboratories may vary in their abilities to detect certain endpoints but the overall conclusion of interaction with the endocrine system will not be affected.

In addition to the redundancy of the endpoints in the protocol to detect reproductive and thyroid effects, the requirement for a minimum of two doses provides an opportunity to examine the relationship between dose and response. If an endpoint is positive at the lower dose in just one endpoint and no effect is seen at the higher dose, then the effect and the overall conclusions would need to be questioned. Thus, the dose response informs the weight of evidence approach discussed above.

Table 31. Endocrine profiles which can be identified in the male pubertal protocol

Androgen antagonist	Steroidogenesis inhibitor or HPG suppression	Hypothyroidism
↑Age of Puberty	↑ Age of Puberty	↓ T_4
↓Ventral prostate, seminal vesicles, LABC, epididymis	↓Ventral prostate, seminal vesicles, LABC, epididymis	↑TSH
↑Testosterone	↓Testosterone or no effect	↑Thyroid wt.
		↑Follicular cell height ↓Colloid area
		↑Liver wt (for compounds which induce hepatic clearance of thyroxine) or no effect

XII. Additional issues concerning sensitivity, specificity, and reproducibility

During the development and validation of the male pubertal assay, concerns have been raised in various quarters that sensitivity may not have been optimized and that neither specificity nor reproducibility has been adequately demonstrated. This section discusses these concerns in the context of the validation of the male pubertal assay for use in a screening program.

A. Phytoestrogens in feed

One concern that has been raised is that the soy and alfalfa content of most major rat feeds may contain sufficient phytoestrogens to interfere with the sensitivity of endocrine assays. There are reports that some laboratory rodent dietary formulations contain levels of phytoestrogens that are sufficient to induce alterations in uterine weight and histology (Boettger-Tong *et al.* (1998)). However, there appear to be no studies providing data relevant to the effect of phytoestrogens on the male pubertal assay's endpoints. In adult (not pubertal) males, the weak phytoestrogen genistein was administered to by gavage at doses of 0, 50, 120, 400, or 1000 mg/kg/day for a period of 21 days (Okazaki *et al.* (2002)). While high-dose rats had statistically significant decreases in body weights (96% of control) and feed consumption, treatment-related effects on clinical signs, accessory sex gland unit weights, or testicular or epididymal

histology were not detected. Serum hormone concentrations were also not significantly altered. Although it was concluded that an MTD was not achieved in this study, the 1000 mg/kg/day certainly exceeds any reasonable estimate of the phytoestrogen concentration present in commercial rodent diets. Similarly, O'Connor *et al.* (2000) reported that coumestrol did not alter male organ weights and testicular histology at dosages up to 2.5 mg/kg/day, i.p. These results are in contrast to the changes that occurred in these parameters after administration of the estrogen receptor agonist 17b-estradiol (O'Connor *et al.* (1998)), and reflects the lower potency of coumestrol.

In the interlaboratory validation study described above for the male pubertal assay, two laboratories (Charles River/Argus and Huntingdon) used the same batch of rat feed (Harlan Teklad 2018CM, with a "genistein equivalent" content of 212 ppm) while one laboratory (WIL) used a different feed (Purina 5002, with a "genistein equivalent" content of 319 ppm). The 50% higher phytoestrogen content had no apparent effect on the overall results for any of the chemicals tested. While there were some differences in specific endpoints, these differences were not consistent and may reflect differences in laboratory techniques across laboratories rather than a systematic effect due to phytoestrogens. For example, at the high dose of DBP WIL did not detect a delay in PPS that was noted by both other labs, but it did detect a decrease in ventral prostate weights that was not detected by the other two labs. Similarly, WIL did not detect the increase in testosterone noted by the other labs at the high dose of vinclozolin, but did detect a decrease in ventral prostate weight that the other labs did not detect. Since other androgen endpoints responded as expected, and since the other labs were inconsistent between themselves on certain endpoints, and since the overall classification of this chemical was the same, it seems reasonable to conclude that phytoestrogen content of the feed, at least up to the limit of 300 - 350 ppm specified in the protocol as the acceptable maximum, is not a major determinant of the sensitivity of the pubertal assay. While the issue deserves additional, more-focused study, it does not appear to be a sufficient reason to disqualify the male pubertal assay from use in screening. The significant amount of research that it will take to resolve whether reducing the phytoestrogen content of feeds will improve the sensitivity of the assay does not appear to be within the scope of the immediate implementation of a screening

program for endocrine activity, although it may well be a part of determining whether the assay can be improved in the future.

Relevant to this discussion, the phytoestrogen contents in the diets used in the validation of the uterotrophic assay were analyzed and the impact of the diet on uterine weight examined by an international group of experts for the Organisation of Economic Cooperation and Development (Owens *et al.* (2003)). There was little indication that the phytoestrogen content of the diet influenced uterine weight in these studies, up to a concentration of about 300 – 350 mg/kg. As an increase in uterine weight is a very sensitive indicator of a estrogenic effect, the lack of influence on the uterotrophic assay indicates that reasonable concentrations of dietary phytoestrogens would not influence the outcome of the male pubertal study, which appears to be less sensitive to estrogenic chemicals than the uterotrophic assay.

Reluctance to require the use of low-phytoestrogen feeds in the absence of data demonstrating that this would improve the sensitivity of the male pubertal assay stems in part from uncertainty that phytoestrogens at the current typical levels in feed are responsible even for the effects seen in female endocrine assays. There are indications, for example, that the metabolizable energy content of the feed, not the phytoestrogen content, is responsible for the estrogenic effects reported (Odum *et al.* (2004)). Changing the diet changes many parameters, and showing correlations between certain parameters such as phytoestrogen content and certain effects without proof of a causal association indicates a clear need for further research but does not provide sufficient indication that the male pubertal assay is inadequate for use in a screening program for endocrine disruption.

B. Rat strain differences

Concerns have been raised both internally and from the EDMVS that the strain of rat used may affect the ability to detect a response in endocrine assays. In response to EDMVS' recommendation to write a white paper on the issue of rat-strain effects on pubertal assay endpoints, EPA prepared such a paper and presented both the White Paper (Appendix 12) and an expert reviewer's comments (Appendix 13) to the EDMVS in August 2003. EPA concluded that although it appears that some strains of rats are

differentially sensitive to endocrine effects, it is not possible at this time to determine which strain will be the most susceptible across all (or most) endpoints. Because the pubertal assay has multiple endpoints, it was not possible to choose an optimal strain (if indeed an optimal strain exists). While EDMVS did consider recommending that multiple strains be used, it decided that this would not be feasible for the multi-endpoint pubertal assays.

The transferability study demonstrated that the expected endocrine-mediated changes in male pubertal development could be detected in both Sprague-Dawley and Long Evans rats (Table 5). The increased variability and later onset of PPS in the Long-Evans rat as compared with the Sprague-Dawleys was shown to be attributable to a vendor-associated disparity, and the apparent insensitivity of Sprague-Dawleys to DBP in that study (while Long Evans were sensitive) was shown not to be strain-related since the interlaboratory validation study showed that DBP could be detected (at the same dose and at a lower dose as was used in the transferability study) in Sprague-Dawleys in three laboratories.

At this time, there is a preference of Crl:CD®(SD) rat for the male pubertal protocol. While the EPA recognizes there are reasons to believe that this strain might be particularly insensitive to endocrine disruptors (see Appendix 13), the data currently available appear to show that it is no worse (or better) than other strains for screening for endocrine activity using the endpoints in the male pubertal assay. Other considerations therefore form the basis for EPA's preference. This strain of rat is readily available and there is a wealth of data and endocrine data available from pubertal studies, thus making it possible to establish performance criteria. As shown in the figures illustrating the performance criteria for each endpoint for Sprague-Dawley and Wistar rats (Section IX.D.1), Wistar and SD rats are within the same range of means for the endpoints, emphasizing the comparability of these strains.

In summary, EPA is aware of the potential for differences between strains and therefore expresses a preference for standardization using the Sprague-Dawley rat. Wistar rats may also be used, and performance criteria have been developed for this strain as well as for Sprague-Dawleys. Given the data currently available and the amount of research it is likely to take to determine the best strategy for optimizing the

use of rat strains, the current uncertainty about the effect of strain on sensitivity does not disqualify this assay for use in a screening program.

C. Specificity

Another concern is that there are no chemicals which have been shown to be entirely negative in this assay. The lack of negative reference chemicals (that is, chemicals which have been tested for all the endocrine activities that can be identified by the pubertal male assay and which are known to be negative for all of these activities) made it difficult to test the specificity of this assay. As noted in the section describing the choice of chemicals for the interlaboratory validation study (Section IX.C), a good-faith effort was made to identify a chemical that was both toxic to other systems but without endocrine effects. Upon testing in this assay, however, the chemical gave positive results. Since at this time it is not known from other assays whether this chemical interacts with the endocrine system, it is not possible to determine whether the pubertal male assay is non-specific or the chemical is indeed interacting with the endocrine system.

It is clear, however, that the male pubertal assay's androgen-related endpoints do not respond to all stresses. There are several chemicals which are known thyroid toxicants that have been shown to be positive for the thyroid effects and negative for the endocrine and reproductive effects in the male pubertal assay. One example of this is perchlorate (Stoker *et al.* (2006)). This chemical altered thyroid hormones, TSH and thyroid histology and caused no effects on any of the reproductive tract weights or puberty onset.

Similarly, the assay's thyroid-related endpoints do not respond to all stresses. The studies on atrazine and ketoconazole are examples of androgen-related endpoints being affected while thyroid-related endpoints are not. Thus, there is reason to believe that the assay is specific to interaction with the endocrine system rather than to general stress.

D. Selection of dose levels

Concern has been raised that the assay provides no guidance on estimation of MTD for pubertal animals in a study of this length. Because accurate and replicable determination of the MTD is important for establishing a chemical as not interacting with the endocrine system, some have argued that MTD-determination should be part of the protocol.

The problem of MTD determination, however, is not unique to the pubertal assay. All assays, for any effect, must show that an adequate challenge has been presented to the system before a negative finding from that challenge can be accepted. MTD determination is not usually considered part of those protocols, nor is the accuracy of MTD determination considered a measure of the validity of those assays.

The EPA recognizes that some investigators may choose to perform special studies to estimate the proper dose levels to use in the pubertal assay. This route may be chosen in some cases because information which is often available for the MTD of adult rats for a 28-day or 90-day exposure may not be applicable to the 30-day exposure in juvenile/pubertal rats on which this assay is based. Such studies are not required, however. The only requirement is that a dose level at or near the MTD be tested before making a claim that the substance does not interact with the endocrine system.

Due to the importance of the MTD determination, EPA has clarified what it will consider evidence of exceeding the MTD for this assay. In addition to clinical observations that indicate stress, terminal body weight loss compared to controls that exceeds approximately 10% and is statistically significantly different from controls may be used as evidence that MTD has been exceeded. Terminal body weight loss that is not statistically different from controls may be an indication that MTD was not reached. In addition, abnormal blood chemistry values at termination (particularly creatinine and blood urea nitrogen (BUN)) may indicate that MTD was exceeded. Finally, histopathology of the kidney (or any other organ where gross observations indicate damage) may be used as evidence that MTD was exceeded. Blood chemistry and histopathology of the kidney are not required, however.

In some cases the second, lower, dose level required by the pubertal protocol can provide useful information even if the high dose level exceeds the MTD.

E. Adjustment for body weight at weaning

When examining organ weights, researchers often consider the difference in the organ-to-body-weight-ratio (or similarly, analysis of covariance with terminal body weight) as the appropriate indicator of whether treatment affected organ weight. Such analyses separate the effect due to treatment from any effect that may be due simply to size of the animal.

Those analyses, however, are not appropriate when treatment itself may affect body weight. (See Section VII.C.3 for a discussion of endocrine-active compounds and their relationship to body weight.) It would separate the effect on each organ only if that effect exceeded the treatment's effect on body weight.

Since it is not possible to measure what the terminal body weight would have been without treatment, body weight at the last point before treatment is used as a surrogate. In this way, effects of treatment on specific organs are separated from the effect of treatment on overall body weight while still taking into account, as best as possible, the effect of the animal's (untreated) size.

The adjustment of organ weights (and age and weight at PPS, but not hormone levels) for covariance with body weight at weaning is an additional measure to take into account untreated size. The procedure described in the protocol for distributing weanlings to test groups randomly based on weight so that all groups are similar in mean and standard deviation at the beginning of treatment also helps to ensure that the group means of treated animals vs. controls for organ weights are not due to differences in untreated size.

XIII. Summary of the male pubertal protocol

Validation of multi-endpoint, apical, *in vivo* assays such as the male pubertal assay is limited by the availability of appropriate reference chemicals to test, as well as by time and other resources. Nevertheless, within this context, the male pubertal protocol has proven to be transferable, sensitive, and reproducible over many different

modes of action. The following sections on strengths and weaknesses discuss the assay in terms of its potential place in a battery of assays for identifying interaction with the endocrine system.

A. Strengths of the male pubertal protocol

The strengths of the male pubertal protocol in the context of a screening program to identify the ability of a test chemical to interact with the estrogen, androgen, and/or thyroid system in humans are that it is an *in vivo* assay, it is performed in a mammalian model, it is an apical assay, it has redundant confirmatory endpoints, it involves the pubertal period of development, and it has a well-established base of knowledge of the relationship of the endpoints to endocrine activity. In many cases, the profile of responses across the various endpoints can suggest mechanisms of action that might be operative and this can help focus attention in later studies.

The fact that this is an *in vivo* assay allows greater confidence, when compared to an *in vitro* assay, that metabolism is accounted for. In some known cases such as vinclozolin and methoxychlor, it is the metabolites which are the most active agents and *in vitro* systems may not identify the parent compound as having the potential to interact with the endocrine system when taken in by a complete organism. Thus, the use of an *in vivo* system reduces the likelihood of false negative or false positive results. In addition, the integrated nature of the endocrine system in the developing organism and the relationship of the endocrine toxicity to other systemic effects cannot be simulated *in vitro*. Also, the use of a mammalian model usually gives rise to greater confidence that results are relevant to humans than if a phylogenetically more removed model is used.

The fact that this is an apical assay is both a strength and a weakness. The ability to detect in a single assay many different modes of action, as demonstrated during the validation process, is efficient. For example, the assay provides the opportunity to measure, in one assay, both reproductive and thyroid responses.

Having redundant confirmatory endpoints is helpful in an assay such as this where variability in proficiency in measuring specific endpoints can be a factor in the ability of the assay to detect a response. Performance criteria, particularly for weak

positive controls, can help ensure a minimum level of sensitivity, but redundant endpoints provide additional aid.

The longer duration of this assay when compared to the female pubertal assay seems to afford greater time for expression of thyroid effects and thus greater sensitivity.

The pubertal assay focuses on a period of development when the endpoints selected for the assay are particularly sensitive to endocrine modulation. Sensitivity is therefore greater than for other life stages which are feasible to include in a screen for endocrine activity.

One of the strengths particular to the male pubertal protocol as an assay for screening for interaction with the endocrine system lies in the fact that the measures included for the identification of endocrine effects are based on a solid knowledge of how the reproductive and thyroid axis mature in the rat. Thus, the extensive basic literature in this area, reviewed in the DRP, provides the background for the underlying assumptions in the assay endpoints and assists in the interpretation of results. The assay offers certain advantages over the proposed *in vitro* tests for determining alterations in AR/ER binding or steroidogenesis in that the *in vitro* assays cannot account for metabolic activation of xenobiotics which would result in “false negative” responses.

Although extensive, the male pubertal protocol consists of several relatively straightforward measures that can be performed in most professional laboratories. This is not to say that training and expertise is not needed to perform the technical aspects of tissue dissection, serum collection, hormone assay, tissue preparation, histology and histopathology. However, all of these procedures are routinely performed in contract laboratories and are an integral part of the current requirements for pesticide registration. Thus, the technical difficulties of this assay should not be a barrier to its implementation under Good Laboratory Practices. In the early stages of prevalidation, laboratories appeared to have difficulty in measuring some of the endpoints in the protocol. For example, the measurements of small tissue weights (i.e., adrenal, pituitary) did eventually require a more explicit description of the tissue dissection procedures in the protocol itself. Also, the variability in hormone data in some of the

early studies was a cause for concern. Many of these early issues did not appear in the later studies once the methods were described more precisely and the participating laboratories became aware of the areas of concern.

B. Weaknesses of the male pubertal protocol

There are limitations to this protocol. One potential weakness is the variability of the testosterone measurements. The hormone assays themselves are not difficult, as they are available commercially and include their own quality control samples. Collecting the sample in a non-stressed animal will reduce the variability. The rest of the variability is due to the changing levels of testosterone over time in the juvenile male rat, which has been shown by Monosson *et al.* (1999). Even with the variability, effects were significant in the interlaboratory comparison study with a decrease following exposure to DBP and an increase following exposure to 2-CNB and vinclozolin, thus suggesting that testosterone measurement can still be useful. The availability of redundant androgen-dependent endpoints in the assay also reduces reliance on the testosterone measurement.

A significant limitation of the EDSP's implementation of the protocol, but not the protocol itself, is the absence of concurrent weak positive controls. As explained in Section VIII, requirement of a sufficient number of positive controls to cover all the endpoints and all the potential modes of action for an apical assay appears to be infeasible. This weakness has been mitigated by the inclusion of performance criteria for the controls, but is not eliminated.

The male pubertal assay is one of the longer assays being considered for the Tier 1 battery. (The duration is not necessary to identify androgenicity but is appropriate for identification of anti-androgens and thyroid-active compounds.) To the extent that ideal screens are short, the duration of this assay may be regarded by some as a disadvantage. Again, this is a limitation only in the context of a particular use, not a shortcoming of the protocol *per se*.

Although identification of mechanism of action is not necessary for identification of interaction with the endocrine system, some observers may feel that the inability of this assay to isolate mechanisms of action is a limitation. For example, although the

protocol did detect the adverse effects of atrazine, it would have been difficult to determine from this assay alone whether this compound blocked steroidogenesis or disrupted the central control of puberty (HPG). This inability to define a mechanism of action is not a weakness in the context of the purpose of Tier 1 screening, which is solely to identify the ability to interact with the endocrine system, but may be a weakness relative to the ability of other, non-apical, assays.

A potential concern is the limited ability to detect highly-specific aromatase inhibitors. Although the male pubertal was able to detect the moderately specific aromatase inhibitors testolactone, ketoconazole and finasteride (a 5α-reductase inhibitor) it was unable to detect fadrozole, a highly specific aromatase inhibitor. It is not known if there are any highly-specific aromatase inhibitors that are environmentally relevant.

The ability to detect estrogenic compounds may also be limited. While the assay makes no claim to be able to detect estrogenic compounds, it was successful in detecting DES and possibly nonylphenol, according to reports from modified pubertal assays in the literature. Methoxychlor caused a decrease in seminal vesicle weights at the high dose tested in the multi-chemical study but the lack of response in any other endpoint of the assay might have led this to be discounted had this been an unknown test compound. Also, the lack of effect on body weight may indicate that the dose level was not high enough.

C. Conclusion

In summary, EPA believes that the male pubertal protocol has proven to be transferable, sensitive to the kinds of interactions with the endocrine system it claims to detect, and reproducible in independent contract laboratories. It also finds reason to believe that the assay is specific even though this is not testable at this time. While there may be ways to improve the assay in the future, the assay appears to be appropriate for use in a screening program to identify interaction with the endocrine system.

XIV. References

1. Aguilar, E., Pinilla, L., Guisado, R., Gonzalez, D., and Lopez, F. (1984). Relation between body weight, growth rate, chronological age and puberty in male and female rats. Rev.Esp.Fisiol. **40**, 83-86.

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

3. Beattie, C. W. and Schwartz, N. B. (1973). Blockade of the proestrous LH surge in cyclic rats by barbiturate administration on diestrus. Proc.Soc.Exp.Biol.Med. **142**, 933-935.

4. Blackburn, D. M., Gray, A. J., Lloyd, S. C., Sheard, C. M., and Foster, P. M. (1988). A comparison of the effects of the three isomers of dinitrobenzene on the testis in the rat. Toxicol.Appl.Pharmacol. **92**, 54-64.

5. Boettger-Tong, H., Murthy, L., Chiappetta, C., Kirkland, J. L., Goodwin, B., Adlercreutz, H., Stancel, G. M., and Makela, S. (1998). A case of a laboratory animal feed with high estrogenic activity and its impact on in vivo responses to exogenously administered estrogens. Environ.Health Perspect. **106**, 369-373.

6. Bronson, F. H. and Heideman, P. D. (1990). Short-term hormonal responses to food intake in peripubertal female rats. Am.J.Physiol **259**, R25-R31.

7. Bucher, J. (1993). NTP technical report on the toxicity studies of 2-Chloronitrobenzene (CAS No. 88-73-3) and 4-Chloronitrobenzene (CAS No. 100-00-5) Administered by Inhalation to F344/N Rats and B6C3F1 Mice. Toxic.Rep.Ser. **33**, 1-F25.

8. Capen, C. C. (1997). Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicol.Pathol. **25**, 39-48.

9. Chapin, R. E., Gulati, D. K., Barnes, L. H., and Teague, J. L. (1993). The effects of feed restriction on reproductive function in Sprague-Dawley rats. Fundam.Appl.Toxicol. **20**, 23-29.

10. Cook, J. C., Mullin, L. S., Frame, S. R., and Biegel, L. B. (1993). Investigation of a mechanism for Leydig cell tumorigenesis by linuron in rats. Toxicol.Appl.Pharmacol. **119**, 195-204.

11. Cooper, R. L., Goldman, J. M., and Stoker, T. E. (1999). Neuroendocrine and reproductive effects of contemporary-use pesticides. Toxicol.Ind.Health **15**, 26-36.

12. Cooper, R. L., Stoker, T. E., Tyrey, L., Goldman, J. M., and McElroy, W. K. (2000). Atrazine disrupts the hypothalamic control of pituitary-ovarian function. Toxicol.Sci. **53**, 297-307.

13. Craft, E. S., DeVito, M. J., and Crofton, K. M. (2002). Comparative responsiveness of hypothyroxinemia and hepatic enzyme induction in Long-Evans rats versus C57BL/6J mice exposed to TCDD-like and phenobarbital-like polychlorinated biphenyl congeners. Toxicol.Sci. **68**, 372-380.

14. De Sandro, V., Chevrier, M., Boddaert, A., Melcion, C., Cordier, A., and Richert, L. (1991). Comparison of the effects of propylthiouracil, amiodarone, diphenylhydantoin, phenobarbital, and 3-methylcholanthrene on hepatic and renal T4 metabolism and thyroid gland function in rats. Toxicol.Appl.Pharmacol. **111**, 263-278.

15. Delaere, K. P. and Van Thillo, E. L. (1991). Flutamide monotherapy as primary treatment in advanced prostatic carcinoma. Semin.Oncol. **18**, 13-18.

16. EDSTAC. 1998. Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) final report. U.S. Environmental Protection Agency.

17. Engelbregt, M. J., van Weissenbruch, M. M., Popp-Snijders, C., Lips, P., and Delemarre-van de Waal HA (2001). Body mass index, body composition, and leptin at onset of puberty in male and female rats after intrauterine growth retardation and after early postnatal food restriction. Pediatr.Res. **50**, 474-478.

18. Feron, V. J., de Groot, A. P., Spanjers, M. T., and Til, H. P. (1973). An evaluation of the criterion "organ weight" under conditions of growth retardation. Food Cosmet.Toxicol. **11**, 85-94.

19. Foster, P. M. (2006). Disruption of reproductive development in male rat offspring following in utero exposure to phthalate esters. Int.J.Androl **29**, 140-147.

20. Friedmann, A. S. (2002). Atrazine inhibition of testosterone production in rat males following peripubertal exposure. Reprod.Toxicol. **16**, 275-279.

21. Fukuda, H., Greer, M. A., Roberts, L., Allen, C. F., v Critchlow, and Wilson, M. (1975). Nyctohemeral and sex-related variations in plasma thyrotropin, thyroxine and triiodothyronine. Endocrinology **97**, 1424-1431.

22. Gaido, K. W., Maness, S. C., McDonnell, D. P., Dehal, S. S., Kupfer, D., and Safe, S. (2000). Interaction of methoxychlor and related compounds with estrogen receptor alpha and beta, and androgen receptor: structure-activity studies. Mol.Pharmacol. **58**, 852-858.

23. Gray, L. E., Jr., Wolf, C., Lambright, C., Mann, P., Price, M., Cooper, R. L., and Ostby, J. (1999). Administration of potentially antiandrogenic pesticides (procymidone, linuron, iprodione, chlozolinate, p,p'-DDE, and ketoconazole) and

toxic substances (dibutyl- and diethylhexyl phthalate, PCB 169, and ethane dimethane sulphonate) during sexual differentiation produces diverse profiles of reproductive malformations in the male rat. Toxicol.Ind.Health **15**, 94-118.

24. Gupta, C., Shapiro, B. H., and Yaffe, S. J. (1980). Reproductive dysfunction in male rats following prenatal exposure to phenobarbital. Pediatr.Pharmacol.(New York.) **1**, 55-62.

25. Hamilton, G. D. and Bronson, F. H. (1986). Food restriction and reproductive development: male and female mice and male rats. Am.J.Physiol **250**, R370-R376.

26. Hodgson,E. (1987). Measurement of toxicity. In A textbook of modern toxicology (E.Hodgson and P.E.Levi, Eds.), pp. 252-253. Elsevier Science Publishing Co., Inc., Amsterdam, The Netherlands.

27. Hood, A., Liu, J., and Klaassen, C. D. (1999). Effects of phenobarbital, pregnenolone-16alpha-carbonitrile, and propylthiouracil on thyroid follicular cell proliferation. Toxicol.Sci. **50**, 45-53.

28. Imperato-McGinley, J., Sanchez, R. S., Spencer, J. R., Yee, B., and Vaughan, E. D. (1992). 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.

29. Kassim, N. M., McDonald, S. W., Reid, O., Bennett, N. K., Gilmore, D. P., and Payne, A. P. (1997). The effects of pre- and postnatal exposure to the nonsteroidal antiandrogen flutamide on testis descent and morphology in the Albino Swiss rat. J.Anat. **190 (Pt 4)**, 577-588.

30. Kelce, W. R., Lambright, C. R., Gray, L. E., Jr., and Roberts, K. P. (1997). 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.

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

32. KENNEDY, G. C. and MITRA, J. (1963). Body weight and food intake as initiating factors for puberty in the rat. J.Physiol **166**, 408-418.

33. Kieffer, J. D., Mover, H., Federico, P., and Maloof, F. (1976). Pituitary-thyroid axis in neonatal and adult rats: comparison of the sexes. Endocrinology **98**, 295-304.

34. Lambright, C., Ostby, J., Bobseine, K., Wilson, V., Hotchkiss, A. K., Mann, P. C., and Gray, L. E., Jr. (2000). Cellular and molecular mechanisms of action of

linuron: an antiandrogenic herbicide that produces reproductive malformations in male rats. Toxicol.Sci. **56**, 389-399.

35. Laws, S. C., Ferrell, J. M., Stoker, T. E., Schmid, J., and Cooper, R. L. (2000). The effects of atrazine on female wistar rats: an evaluation of the protocol for assessing pubertal development and thyroid function. Toxicol.Sci. **58**, 366-376.

36. Maric, D., Simonovic, I., Kovacevic, R., Krsmanovic, L., Stojilkovic, S., and Andjus, R. K. (1982). Effects of short-term and long-term hyperprolactinemia on the developmental pattern of androgen and LH levels in the immature male rat. J.Endocrinol.Invest **5**, 235-241.

37. Marty, M. S., Crissman, J. W., and Carney, E. W. (2001a). Evaluation of the male pubertal assay's ability to detect thyroid inhibitors and dopaminergic agents. Toxicol.Sci. **60**, 63-76.

38. Marty, M. S., Crissman, J. W., and Carney, E. W. (2001b). Evaluation of the male pubertal onset assay to detect testosterone and steroid biosynthesis inhibitors in CD rats. Toxicol.Sci. **60**, 285-295.

39. McClain, R. M., Levin, A. A., Posch, R., and Downing, J. C. (1989). The effect of phenobarbital on the metabolism and excretion of thyroxine in rats. Toxicol.Appl.Pharmacol. **99**, 216-228.

40. McMullin, T. S., Brzezicki, J. M., Cranmer, B. K., Tessari, J. D., and Andersen, M. E. (2003). Pharmacokinetic modeling of disposition and time-course studies with [14C]atrazine. J.Toxicol.Environ.Health A **66**, 941-964.

41. Merry, B. J. and Holehan, A. M. (1979). Onset of puberty and duration of fertility in rats fed a restricted diet. J.Reprod.Fertil. **57**, 253-259.

42. Merry, B. J. and Holehan, A. M. (1985). The endocrine response to dietary restriction in the rat. Basic Life Sci. **35**, 117-141.

43. Monosson, E., Kelce, W. R., Lambright, C., Ostby, J., and Gray, L. E., Jr. (1999). 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.

44. Murphy, W. M., Soloway, M. S., and Barrows, G. H. (1991). Pathologic changes associated with androgen deprivation therapy for prostate cancer. Cancer **68**, 821-828.

45. Mylchreest, E., Sar, M., Cattley, R. C., and Foster, P. M. (1999). Disruption of androgen-regulated male reproductive development by di(n-butyl) phthalate during late gestation in rats is different from flutamide. Toxicol.Appl.Pharmacol. **156**, 81-95.

46. NIEHS. 3-1-1997. Validation and regulatory acceptance of toxicological test methods: a report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (NIH 97-3981). DHHS/NTP/NIEHS/NTP.

47. O'Connor, J. C., Davis, L. G., Frame, S. R., and Cook, J. C. (2000). Evaluation of a Tier I screening battery for detecting endocrine-active compounds (EACs) using the positive controls testosterone, coumestrol, progesterone, and RU486. Toxicol.Sci. **54**, 338-354.

48. O'Connor, J. C., Frame, S. R., Biegel, L. B., Cook, J. C., and Davis, L. G. (1998). Sensitivity of a Tier I screening battery compared to an in utero exposure for detecting the estrogen receptor agonist 17 beta-estradiol. Toxicol.Sci. **44**, 169-184.

49. O'Connor, J. C., Frame, S. R., Davis, L. G., and Cook, J. C. (1999). Detection of thyroid toxicants in a tier I screening battery and alterations in thyroid endpoints over 28 days of exposure. Toxicol.Sci. **51**, 54-70.

50. O'Connor, J. C., Frame, S. R., and Ladics, G. S. (2002). Evaluation of a 15-day screening assay using intact male rats for identifying antiandrogens. Toxicol.Sci. **69**, 92-108.

51. Odum, J., Tinwell, H., Tobin, G., and Ashby, J. (2004). Cumulative dietary energy intake determines the onset of puberty in female rats. Environ.Health Perspect. **112**, 1472-1480.

52. OECD. 1996. Final report of the OECD workshop on harmonization of validation and acceptance criteria for alternative toxicological test methods. Organisation for Economic Cooperation and Development.

53. Okazaki, K., Okazaki, S., Nakamura, H., Kitamura, Y., Hatayama, K., Wakabayashi, S., Tsuda, T., Katsumata, T., Nishikawa, A., and Hirose, M. (2002). A repeated 28-day oral dose toxicity study of genistein in rats, based on the 'Enhanced OECD Test Guideline 407' for screening endocrine-disrupting chemicals. Arch.Toxicol. **76**, 553-559.

54. Owens, W., Ashby, J., Odum, J., and Onyon, L. (2003). The OECD program to validate the rat uterotrophic bioassay. Phase 2: dietary phytoestrogen analyses. Environ.Health Perspect. **111**, 1559-1567.

55. Owens, W., Gray, L. E., Zeiger, E., Walker, M., Yamasaki, K., Ashby, J., and Jacob, E. (2007). The OECD program to validate the rat Hershberger bioassay to screen compounds for in vivo androgen and antiandrogen responses: phase 2 dose-response studies. Environ.Health Perspect. **115**, 671-678.

56. Pazos-Moura, C. C., Moura, E. G., Dorris, M. L., Rehnmark, S., Melendez, L., Silva, J. E., and Taurog, A. (1991). Effect of iodine deficiency and cold exposure

on thyroxine 5'-deiodinase activity in various rat tissues. Am.J.Physiol **260**, E175-E182.

57. Perheentupa, A., Bergendahl, M., and Huhtaniemi, I. (1995). Modulation of gonadotropin secretion at the pituitary level by testosterone in gonadotropin-releasing hormone-treated male rats during food deprivation. Biol.Reprod. **52**, 808-813.

58. Ronnekleiv, O. K., Ojeda, S. R., and McCann, S. M. (1978). Undernutrition, puberty and the development of estrogen positive feedback in the female rat. Biol.Reprod. **19**, 414-424.

59. Shin, J. H., Kim, H. S., Moon, H. J., Kang, H., Kim, T. S., Seok, J. H., Kim, I. Y., Park, K. L., Han, S. Y., and Nam, S. Y. (2002). Effects of flutamide on puberty in male rats: an evaluation of the protocol for the assessment of pubertal development and thyroid function. J.Toxicol.Environ.Health A **65**, 433-445.

60. Shin, J. H., Moon, H. J., Kim, T. S., Kang, I. H., Ki, H. Y., Choi, K. S., and Han, S. Y. (2006). Repeated 28-day oral toxicity study of vinclozolin in rats based on the draft protocol for the "Enhanced OECD Test Guideline No. 407" to detect endocrine effects. Arch.Toxicol. **80**, 547-554.

61. Shiroozu, A., Taurog, A., Engler, H., and Dorris, M. L. (1983). Mechanism of action of thioureylene antithyroid drugs in the rat: possible inactivation of thyroid peroxidase by propylthiouracil. Endocrinology **113**, 362-370.

62. Siglin, J. C., Mattie, D. R., Dodd, D. E., Hildebrandt, P. K., and Baker, W. H. (2000). A 90-day drinking water toxicity study in rats of the environmental contaminant ammonium perchlorate. Toxicol.Sci. **57**, 61-74.

63. Stoker, T. E., Cooper, R. L., Lambright, C. S., Wilson, V. S., Furr, J., and Gray, L. E. (2005). In vivo and in vitro anti-androgenic effects of DE-71, a commercial polybrominated diphenyl ether (PBDE) mixture. Toxicol.Appl.Pharmacol. **207**, 78-88.

64. Stoker, T. E., Ferrell, J. M., Laws, S. C., Cooper, R. L., and Buckalew, A. (2006). Evaluation of ammonium perchlorate in the endocrine disruptor screening and testing program's male pubertal protocol: ability to detect effects on thyroid endpoints. Toxicology **228**, 58-65.

65. Stoker, T. E., Guidici, D. L., Laws, S. C., and Cooper, R. L. (2002). The effects of atrazine metabolites on puberty and thyroid function in the male Wistar rat. Toxicol.Sci. **67**, 198-206.

66. Stoker, T. E., Laws, S. C., Crofton, K. M., Hedge, J. M., Ferrell, J. M., and Cooper, R. L. (2004). Assessment of DE-71, a commercial polybrominated diphenyl ether (PBDE) mixture, in the EDSP male and female pubertal protocols. Toxicol.Sci. **78**, 144-155.

67. Stoker, T. E., Laws, S. C., Guidici, D. L., and Cooper, R. L. (2000a). The effect of atrazine on puberty in male wistar rats: an evaluation in the protocol for the assessment of pubertal development and thyroid function. Toxicol.Sci. **58**, 50-59.

68. Stoker, T. E., Parks, L. G., Gray, L. E., and Cooper, R. L. (2000b). Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC recommendations. Endocrine Disrupter Screening and Testing Advisory Committee. Crit Rev.Toxicol. **30**, 197-252.

69. Tan, B. L., Kassim, N. M., and Mohd, M. A. (2003). Assessment of pubertal development in juvenile male rats after sub-acute exposure to bisphenol A and nonylphenol. Toxicol.Lett. **143**, 261-270.

70. Trentacoste, S. V., Friedmann, A. S., Youker, R. T., Breckenridge, C. B., and Zirkin, B. R. (2001). Atrazine effects on testosterone levels and androgen-dependent reproductive organs in peripubertal male rats. J.Androl **22**, 142-148.

71. USEPA. 12-28-1998. Endocrine Disruptor Screening Program; Proposed statement of policy. Federal Register 63(248), 71542-71568. U.S. Environmental Protection Agency.

72. USEPA. 1999. Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. USEPA/SAB.

73. USEPA. 2007. Validation of screening and testing assays proposed for the EDSP. Version 5.4. USEPA/OPPTS/OSCP/EDSP.

74. Wani, J. H., Agrawal, A. K., and Shapiro, B. H. (1996). Neonatal phenobarbital-induced persistent alterations in plasma testosterone profiles and testicular function. Toxicol.Appl.Pharmacol. **137**, 295-300.

75. Widdowson, E. M. and McCance, R. A. (1960). Some effects of accelerating growth. I. General somatic development. Proc.R.Soc.Lond B Biol.Sci. **152**, 188-206.

76. Yamada, T., Kunimatsu, T., Miyata, K., Yabushita, S., Sukata, T., Kawamura, S., Seki, T., Okuno, Y., and Mikami, N. (2004). Enhanced rat Hershberger assay appears reliable for detection of not only (anti-)androgenic chemicals but also thyroid hormone modulators. Toxicol.Sci. **79**, 64-74.

77. Yamasaki, K., Tago, Y., Nagai, K., Sawaki, M., Noda, S., and Takatsuki, M. (2002). Comparison of toxicity studies based on the draft protocol for the 'Enhanced OECD Test Guideline no. 407' and the research protocol of 'Pubertal Development and Thyroid Function in Immature Male Rats' with 6-n-propyl-2-thiouracil. Arch.Toxicol. **76**, 495-501.

Appendix 1. Male pubertal protocol

Appendix 2. Detailed Review Paper

Appendix 3. Literature studies since the Detailed Review Paper

Appendix 4. Transferability study (TherImmune 1) summary report

Appendix 5. Transferability study (TherImmune 1) detailed table of results

Appendix 6. Multi-chemical study (RTI) summary report

Note: This summary report was written using, in part, organ weights adjusted for covariance with terminal body weight, as requested by EPA. However, EPA later decided that adjustment for covariance with weaning body weight is more appropriate. Thus, some of the descriptions and conclusions in this report may not coincide with EPA's final analysis and interpretation. The analysis of covariance with weaning body weight on which EPA's interpretation is based is provided as Appendix 8.

Appendix 7. Multi-chemical study (RTI) detailed table of results

Appendix 8. Multi-chemical study (RTI) ANCOVA with body weight at weaning

Appendix 9. Multi-dose study (TherImmune 2) summary report

Note: This summary report was written using, in part, organ weights adjusted for covariance with terminal body weight, as requested by EPA. However, EPA later decided that adjustment for covariance with weaning body weight is more appropriate. Thus, some of the descriptions and conclusions in this report may not coincide with EPA's final analysis and interpretation. The analysis of covariance with weaning body weight on which EPA's interpretation is based is provided as Appendix 11.

Appendix 10. Multi-dose study (TherImmune 2) detailed table of results

Appendix 11. Multi-dose study (TherImmune 2) ANCOVA with body weight at weaning

Appendix 12. White Paper on rat strain differences

Appendix 13. Reviewer's comments on White Paper on rat strain differences

Appendix 14. Interlaboratory validation study summary report (Charles River/Argus)

Appendix 15. Interlaboratory validation study summary report (Huntingdon)

Appendix 16. Interlaboratory validation study summary report (WIL)

Appendix 17. Interlaboratory validation study analysis report (Battelle)

Appendix 18. Interlaboratory validation study detailed table of results

Appendix 19. Interlaboratory validation study, comparison of results table