Pure Appl. Chem., Vol. 75, Nos. 11–12, pp. 2081–2098, 2003.

Topic 3.9

Experience with new testing guidelines with endocrine-sensitive endpoints*

Rochelle W. Tyl‡

Center for Life Sciences and Toxicology, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709-2194, USA

Abstract: Concerns about the effects of endocrine disruptors on humans and wildlife have resulted in revised governmental testing guidelines (e.g., U.S. Environmental Protection Agency, Organization for Economic Cooperation and Development, U.S. Food and Drug Administration), adding endpoints to enhance their capability to detect endocrine active compounds. Based on experience with these testing guidelines, I present my opinions and data on study design, performance, results, endpoints, interpretation, and recommendations for improvement. New regulatory endpoints must be reproducible, robust, sensitive, relevant, and consistent. These new endpoints are appropriate: anogenital distance, examination of culled pups on postnatal day 4, examination for retained nipples/areolae in preweanling males, weanling necropsy, acquisition of puberty in offspring, prebreed estrous cyclicity, reproductive organ weights and histopathology, and andrology. Endpoints considered not appropriate are ovarian primordial follicle counts, stage of estrus for parental females at demise, and single blood sample at necropsy to measure circulating levels of relevant hormones. In conclusion: (1) regulatory guidelines represent only the minimum requirements and should/will change as science improves and scientific and societal concerns arise; (2) scientists must interact to identify and refine new methods, endpoints, and mechanisms; (3) the objective of these studies is to provide good science and useful information for risk assessment. Regulatory agencies should convene workshops with stakeholders to discuss issues such as new fundamental/applied information, mechanisms, and current and new endpoints. Interactions will lead to enhanced science, periodic revisions of testing guidelines, and improved risk assessment.

BACKGROUND

During the late 1980s, circumstantial evidence was accumulating that humans and wildlife were exhibiting effects consistent with alterations in endocrine function. These included decreasing sperm counts worldwide, increasing incidences of prostate and breast cancer, accelerated puberty, and increased incidences of male reproductive malformations in humans. For wildlife, these effects included reductions in reproductive success, increases in malformed offspring, alterations in sexual and reproductive behaviors in fish-eating birds of the Great Lakes, alterations in external reproductive organs in alligators in Lake Apopka (Florida) and in Florida panthers, accelerated reductions (or total losses) of localized populations of amphibians, imposexes in harbor snails, inappropriate secondary and tertiary

*Report from a SCOPE/IUPAC project: Implication of Endocrine Active Substances for Human and Wildlife (J. Miyamoto and J. Burger, editors). Other reports are published in this issue, *Pure Appl. Chem.* **75**, 1617–2615 (2003).

‡Tel.: 919-541-5972; Fax: 919-541-5956; E-mail: rwt@rti.org

sex characteristics (i.e., male structures in females, female structures in males) in various fish populations, and other effects. Dr. Theo Colborn of the World Wildlife Fund recognized these events and coined the term "endocrine disruption" to describe effects from endogenous (e.g., phytoestrogens) and anthropogenic (i.e., manmade; examples include phthalate plasticizers, Kraft pulp paper mill effluents, synthetic hormones in waste water from use of contraceptive pills, nonylphenol, octylphenol, and bisphenol A-based plastics) materials in the environment. Almost single-handedly, she organized a series of conferences/workshops (Wingspread conferences) on various aspects of the effects and mechanisms of endocrine classes (e.g., steroids, thyroid hormone) on humans and wildlife, with participants from various areas of expertise, including academic endocrinologists, researchers using animal models, wildlife biologists, epidemiologists, neurologists, and andrologists. Dr. Colborn also testified before the U.S. Congress.

The U.S. Environmental Protection Agency (USEPA), in response to the growing evidence for effects of endocrine active compounds, revised its original testing guideline for Reproductive Toxicity Effects (Fig. 1; 40 Code of Federal Regulations [CFR] 798.4700 [1]), with a draft of a new guideline with limited circulation in 1994 and a public draft in 1996 for comments, and established the Endocrine Disruptors Screening and Testing Advisory Committee (EDSTAC) in 1996. The USEPA Toxic Substance Control Act (TSCA) regulators finalized the new testing guidelines in 1997 (799.9380; USEPA TSCA [2]), and the USEPA Office of Prevention, Pesticides and Toxic Substances (OPPTS) regulators, including both TSCA and Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), finalized the new testing guidelines in 1998 (870.3800; USEPA OPPTS; Fig. 2 [3]). The OECD promulgated its new reproductive toxicity testing guideline in 2001 (OECD, No. 416; Fig. 3 [4]), and the U.S. Food and Drug Administration (FDA) presented its new reproductive toxicity testing guideline in the Redbook (Fig. 4) [5].

All of the new guidelines retained the original study design in terms of exposures and matings (Fig. 1), but added endocrine-sensitive endpoints to enhance the efficacy of the testing guideline to detect endocrine-mediated effects (Fig. 2). My laboratory has completed ten studies under the new OPPTS

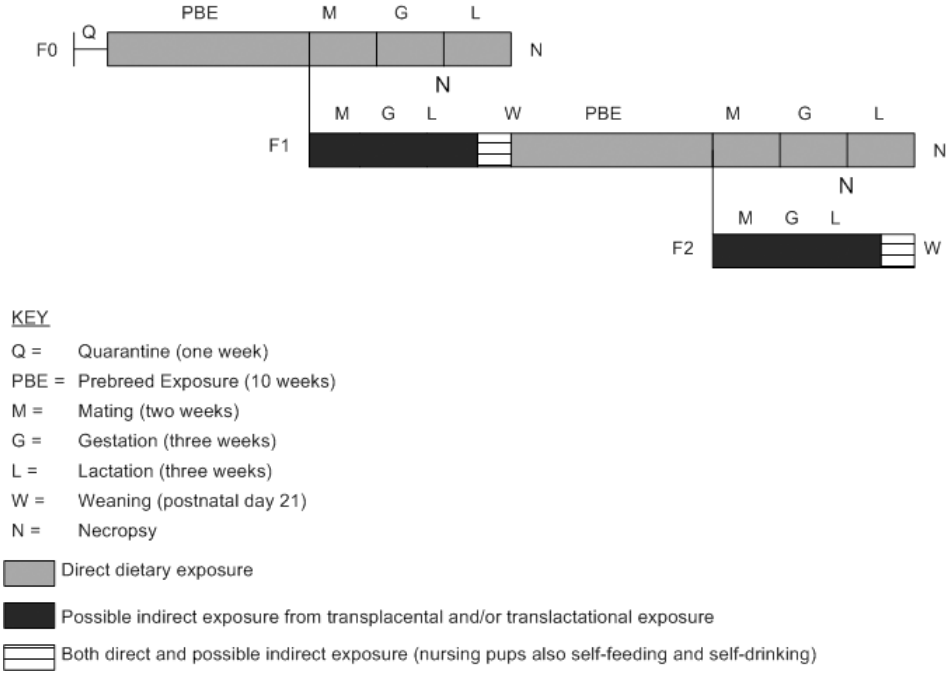
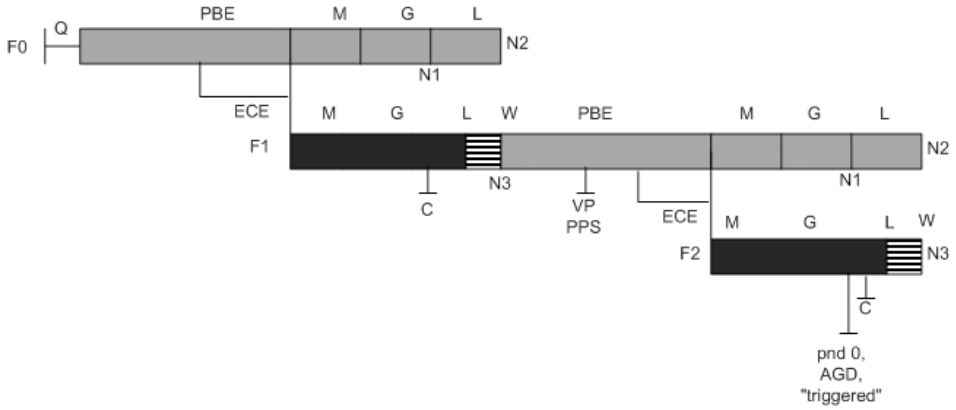


Fig. 1 Previous USEPA (TSCA and FIFRA) Testing Guidelines since 1978.



KEY

Q = Quarantine (one week)

PBE = Prebreed Exposure (10 weeks)

M = Mating (two weeks)

G = Gestation (three weeks)

L = Lactation (three weeks)

VP = Vaginal patency (evaluated in F1 females on postnatal day 22 to acquisition)

PPS = Preputial separation (evaluated in F1 males on postnatal day 35 to acquisition)

W = Weaning (postnatal day 21)

N1 = Necropsy of all paternal animals (organ weights, histology, andrological assessments)

N2 = Necropsy of all maternal animals (organ weights, histology, ovarian follicle assessments)

N3 = Necropsy of selected weanlings, three/sex/litter, if possible (organ weights)

ECE = Estrous Cyclicity Evaluations (three weeks)

C = Cull litters to 10 pups (with equal sex ratio) on postnatal day 4

AGD = Anogenital distance measured in F2 pups on pnd 0 if triggered by effects on F1 reproductive development

N = Necropsy

Direct dietary exposure

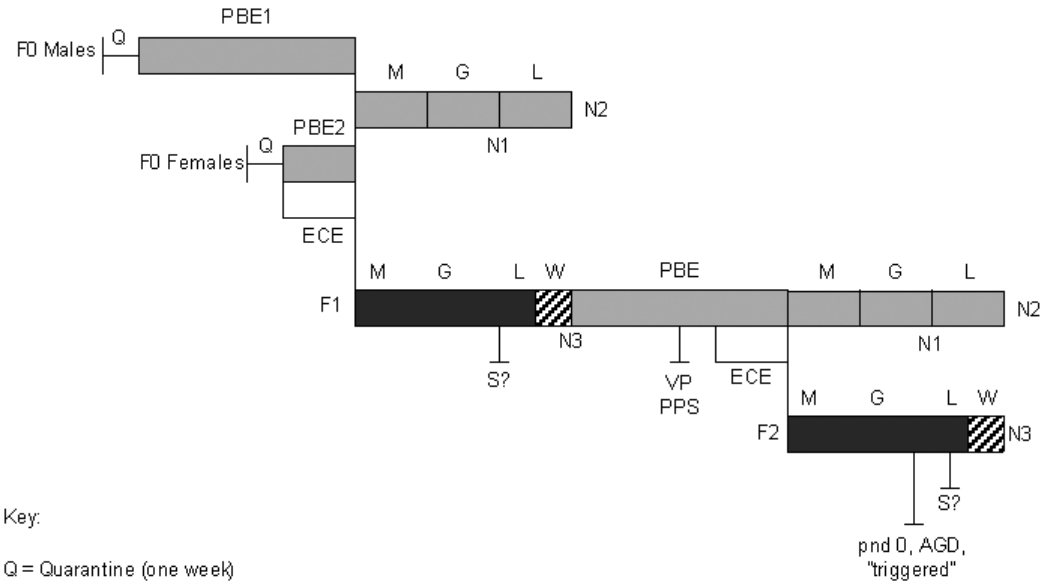
Possible indirect exposure from transplacental and/or translactational exposure

Both direct and possible indirect exposure (nursing pups also self-feeding and self-drinking)

Fig. 2 Most recent USEPA TSCA Final Test Guidelines (USEPA 799.9380, 1997) and OPPTS (FIFRA/TSCA) Final Testing Guidelines (USEPA 870.3800, 1998), "Reproductive and Fertility Effects".

testing guidelines, as well as FDA Segment I and III studies, OECD 415 and 416 studies, and OECD-modified 422 and 421 studies, with inclusion of the new endocrine-sensitive endpoints.

Based on our experience, I offer you my thoughts on the study design, performance, endpoints, interpretation, and reporting of study results, including my recommendations to correct perceived deficiencies.



Key:

Q = Quarantine (one week)

PBE1 = Prebreed exposure in F0 males; ten weeks for rats, eight weeks for mice

PBE2 = Prebreed exposure in F0 females ("exposure during growth and for several complete estrous cycles"; 14 days)

PBE = Prebreed exposure in F1 males and females (10 weeks for rats, eight weeks for mice)

M = Mating (two weeks)

G = Gestation (approximately three weeks; 18-19 days)

L = Lactation (three weeks)

VP = Vaginal patency (evaluated in F1 females on postnatal day 22 to acquisition)

PPS = Preputial separation (evaluated in F1 males on postnatal day 25 to acquisition)

W = Weaning (postnatal day 21)

N1 = Necropsy of all paternal animals (organ weights, histology, andrological assessments)

N2 = Necropsy of all maternal animals (organ weights, histology, ovarian follicle assessments in F1)

N3 = Necropsy of selected weanlings, one/sex/litter, with tissues retained (brain, spleen, thymus)

ECE = Estrous cyclicity evaluation (F0 for two weeks, F1 for three weeks)

S? = Standardization of litters to ten pups (with equal sex ratio) on postnatal day 4 is "optional"

AGD = Anogenital distance measured in F2 pups on postnatal day 0 if triggered by effects on F1 reproductive development ("alterations in F1 sex ratio or timing of sexual maturation")

N = Necropsy

Direct dietary exposure

Possible indirect exposure from transplacental and/or translactational exposure

Both direct and possible indirect exposure (nursing pups also self-feeding and self-drinking)

Fig. 3 Most recent OECD Guideline for the Testing of Chemicals (No. 416), Two-Generation Reproductive Toxicity Study (adopted 22 January 2001).

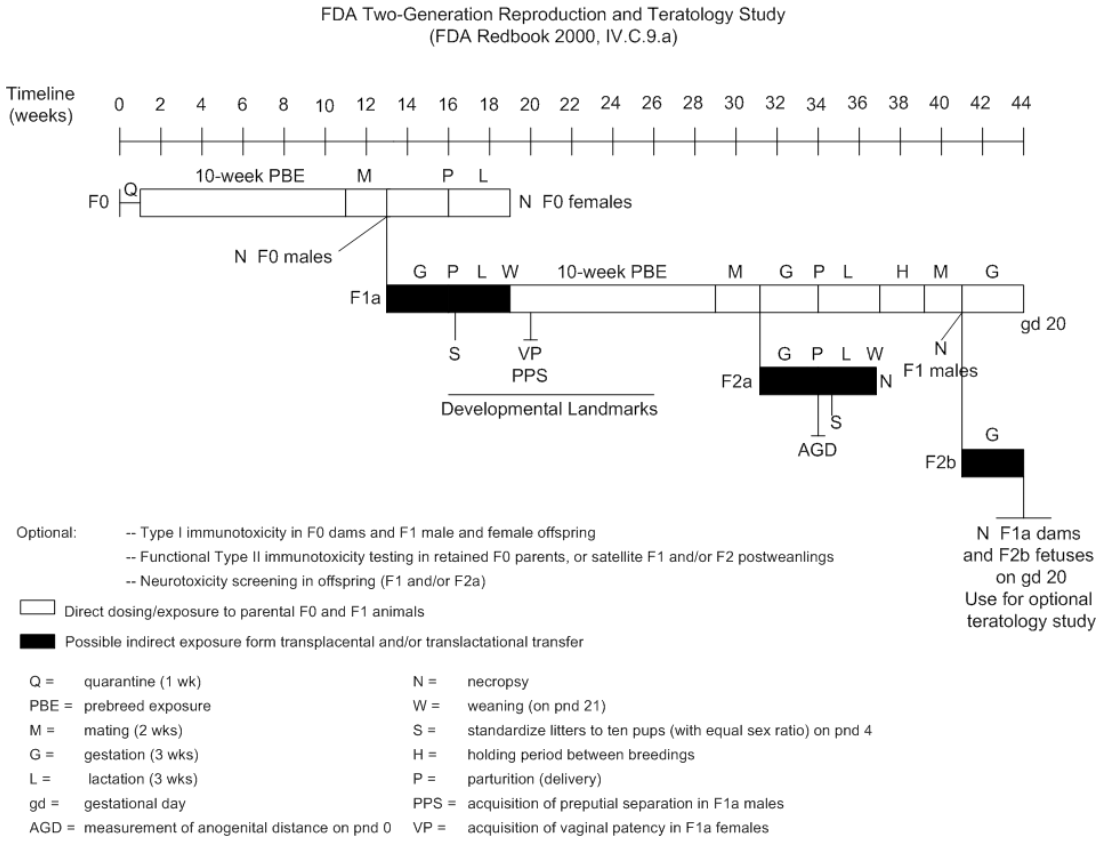


Fig. 4 Most recent FDA Two-Generation Reproductive and Teratology Study (FDA Redbook 2000, IV.C.9.a).

STUDY DESIGN

The multigeneration study design is "apical". Intact animals are exposed beginning as postpubescent animals through adulthood and reproduction (F0 generation), beginning as gametes through adulthood and reproduction (F1 generation—the most important generation), and beginning as gametes to weaning (F2 generation). Therefore, a number of critical, sensitive life stages are exposed, including prenatal, perinatal, lactational, prepubescent, peripubescent, adult, and reproductive. Route of administration mimics the route known, anticipated, or modeled for the target species. Doses (or dietary or drinking-water concentrations) are usually selected to include a clear adult effect level (usually exceeding environmental exposure by order[s] of magnitude) to very low environmentally relevant doses. There are many endpoints to detect endocrine-mediated outcomes, from centrally mediated effects (e.g., brain, hypothalamus) to locally mediated effects (e.g., on gonads, sex accessory organs, other glands and organs). As an apical test (and not a "specific" one), it will not necessarily identify the specific mechanism(s), but its design evaluates multiple possible mechanisms. The endpoints chosen become very important (endpoints are discussed in more detail below).

PERFORMANCE

Staff must be well trained and competence reaffirmed, with intra- and inter-technician variability ascertained and shown to be low. Historical control data from the performing laboratory are also very important to provide context and interpretation for observed effects (e.g., the concurrent control value is very low, but the dose group value is well within the historical control range).

The choice of methodologies used to evaluate endpoints is also critical. For example, anogenital distance is measured in some laboratories with a ruler and a hand-held animal, resulting in large variability. In my laboratory, we use an ocular diopter with a stage micrometer attached to a dissecting microscope for newborn anogenital distances. The selection of which instruments will determine the accuracy, with some models accurate to 0.2 mm (when newborn control rat values are typically 2.0 mm for males and 1.0 mm for females) and some accurate to 0.01 mm. This means that differences of 0.2 to 0.3 mm may be statistically significantly different but not necessarily biologically significant. For weanling and adult anogenital distances, my laboratory uses digital vernier calipers (these work best on animals after euthanasia prior to necropsy). Other simple things that may affect the study performance include checking balance calibrations before and after weighing animals and organs, keeping procedures consistent across generations and studies, and documenting clinical observations carefully, completely, and consistently. These studies, typically used for risk assessment by cognizant governmental agencies, are ideally performed in compliance with stringent good laboratory practice (GLP) principles, regulations, and standards. This adherence to GLPs will ensure appropriate procedures and training for the study performance.

ENDPOINTS

It is my thesis that endpoints in this type of study (for risk assessment purposes) must first be shown to be robust, reproducible, appropriately sensitive, biologically plausible, and relevant to the adverse outcomes of concern. Definitions of the attributes of such endpoints are as follows:

1. *Reproducible*: These endpoints must be reliable: the same findings occur under the same conditions within the initial reporting laboratory (intra-laboratory) and among other laboratories (inter-laboratory). If the results from endpoints are not reproducible, they cannot form the basis for future research and are most likely not useful for risk assessment.
2. *Robust*: These endpoints must be present after comparable routes of exposure (e.g., dosed feed or dosed water). The use of oral gavage, a bolus dose once per day, may result in exacerbation of the effect on an endpoint if the parent material is the proximate toxicant and is metabolized to a non-toxic metabolite, and, if bolus dosing overwhelms the metabolic capacity of the organism or preparation, it may result in diminution or loss of the effect on an endpoint if the parent compound must be metabolized to the active form. Different effects may be observed by non-oral routes, such as inhalation, topical application, and injection, since these routes bypass "first-pass" metabolism by the liver. The findings from routes unrelated to human or environmental exposures may not be useful for risk assessment. These findings must also be present at the same routes and doses over time.
3. *Sensitive*: These endpoints should not be dependent on unique conditions (e.g., intrauterine position [IUP]), especially those that are not relevant to the species at risk. These endpoints should not exhibit high variability (insensitive) or be greatly affected by confounders (too sensitive).
4. *Relevant*: These endpoints must be biologically plausible and related to "adverse" effects of interest/concern. If there are no "adverse" effects at the dose/duration/route evaluated, these endpoints should be predictive of other "adverse" effects at higher doses, after longer exposure duration, and/or by different routes, etc.
5. *Consistent*: These endpoints must occur in the presence of effects in other related, relevant endpoints, if possible, at the same dose, timing, duration, routes of exposure, etc.

Examples of such appropriate and inappropriate endpoints for risk assessment (in my opinion) are as follows (data from my laboratory and elsewhere are also presented):

1. *Anogenital distance (both sexes at birth, at weaning, at adulthood; Table 1)*: This is considered a very appropriate endpoint. It is dihydrotestosterone- (DHT-) mediated, and endocrine-mediated effects persist into adulthood. However, it is confounded by body weight. Therefore, the current practice is to present the data as mm, mm/gram (mm/g) body weight, mm/cube root of the body weight, and/or to analyze the data by analysis of covariance (ANCOVA), with the body weight at measurement (e.g., birth, weaning) as the covariate, to account for differences in body weight (especially in groups where there is systemic toxicity, such as reduced parental and offspring body weights). Very small changes in this parameter (e.g., ≤0.05 mm on postnatal day [pnd] 0) may indicate only body weight-related delays in development, vs. large changes (e.g., ≥0.15 mm on pnd 0) more likely indicative of effects from endocrine disruption. Male pups with demasculinized (feminized) anogenital distance are more likely to exhibit reproductive system malformations, but the correlation is not perfect (i.e., some males with shortened anogenital distance exhibit no malformations, and some males with normal anogenital distance do exhibit malformations). It is also a reasonable predictor at lower doses of increased incidence of male reproductive malformations from perinatal exposure at higher doses. Current OECD and USEPA guidelines trigger anogenital distance in newborn F2 offspring only if effects on reproductive development are detected in F1 offspring (e.g., delays in acquisition of puberty, alterations in sex ratio). Since F2 offspring are terminated at weaning, effects on anogenital distance in F2 newborn pups cannot be related to any postwean effects, such as puberty, estrous cycling, adult reproductive system structures, or func-

Table 1 Anogenital distance in rats.

			Female anogenital distance (mm)				**Male anogenital distance (mm)**			
Study	PND	Gen.	Control		High dose		Control		High dose	
Code			BW (g)	AGD	BW (g)	AGD	BW (g)	AGD	BW (g)	AGD
A	0	F2	6.08	0.73	6.02*	0.63**	6.47	2.25	5.64*	2.11*
B	0	F2	5.93	0.76	5.85	0.79*	6.27	2.01	6.22	2.00
C	0	F2	5.97	0.95	5.99	0.96	6.28	1.98	6.38	2.00
	0	F3	5.98	0.92	6.14	0.94	6.36	1.97	6.49	1.96
D	0	F2	6.09	0.96	5.81	0.96	6.43	2.11	6.20	2.00*
E	0	F2	6.00	0.97	6.12	0.98	6.34	2.05	6.46	2.06
F[a]	0	F1	6.35	0.96	5.91*	0.92	6.76	2.06	6.15**	1.71***
	0	F2	6.21	0.98	5.93	0.99	6.63	2.05	6.29	1.77***
G (E2)	0	F1	6.17	0.95	6.56*	0.95	6.47	2.00	6.74	2.03
H (F)[b]	1	F1	NOT DONE					2.9		1.4*
I (L)[c]	1	F1	NOT DONE					2.9		2.7*
J (TP)[d]	2	F1		1.72		1.64	[3.07–3.33][e]	3.80		2.92*
K (TB)[f]	2	F1		1.5		2.2**		UNAFFECTED		

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (study codes A through G are from the author's laboratory)

[a]An antiandrogen (Tyl, unpublished observations).

[b]Flutamide [6].

[c]Linuron [7].

[d]Testosterone propionate [5].

[e]At the top dose, external sex distinction could not be made for most offspring; this is the range of AGD values for these pups [5].

[f]17ß-trembolone [8].

tions. We recommend measuring anogenital distance in newborn F1 animals as well, a generation that is thoroughly evaluated through sexual maturity and reproduction. This will enable interpretation of the consequences or predictability, if any, of altered anogenital distance.

2. *Necropsy of culled pups on pnd 4*: Neither the OECD nor OPPTS guidelines specify what to do with the culled pups (standardizing litters is optional in OECD 416). My strong recommendation is to necropsy the culled pups. Testes, epididymides, ovaries, and uteri can be weighed to detect changes in size better than by visual examination alone. Ectopic/undescended testes can be detected, as well as major malformations (e.g., missing parts of epididymides, missing testes and/or epididymides, necrosis, swelling, flaccidity of testes) in these organs.
3. *Retained/reduced nipples and areolae in preweanlings (Table 2, A and C)*: This is not included in the current guidelines. Involution of fetal nipples in males is DHT-mediated, and retained nipples persist into adulthood. In the author's laboratory, retained nipples have never been observed in control preweanling CD® (Sprague–Dawley [SD]) male rats, although areolae are present in our laboratory in 0 to 2.7 % of control males on pnd 11–13 (based on examination of over 3000 males in toto). Reduced nipples and/or areolae in female preweanlings are observed when they are exposed in utero to a strong androgen (Table 2, B and D). This is a sensitive indicator of altered testosterone (T) and/or DHT levels (e.g., effects on synthesis, degradation, receptor binding, transcriptional activation). Male pups with retained nipples are more likely to exhibit reproductive system malformations, but the correlation is not perfect (i.e., some males with nipples exhibit no malformations, and some males with no nipples do exhibit malformations). It is also a reasonable predictor at lower doses of male reproductive malformations from perinatal exposures at higher doses. Reduction in areolae/nipples in preweanling females was associated with reproductive system abnormalities [6]. In my laboratory, we examine preweanling rat offspring on pnd 11–13 and preweanling mouse offspring on pnd 9–11.

Table 2 Nipple and areolae retention/reduction in preweanling rats (pnd 11–13).

A. Nipple retention in male rats

		Control		High dose	
Study code	Gen.	% Males	No./male	% Males	No./male
A[a]	F1	0.00	0.00	19.23***	0.72***
	F2	0.00	0.00	16.46	0.51*
B[b]	F1	0.00	0.00	0.00	0.00
	F2	0.00	0.00	0.00	0.00
	F3	0.00	0.00	0.00	0.00
C (L)[c]	F1	–	0.8	–	3.3*[c]
D (F)[d]	F1	16	0.5[c]	100.0	12.0*

B. Nipple reduction in female rats

		Control		High dose	
Study code	Gen.	% Females	No./female	% Females	No./female
A (TP)[e]	F1	100.0	12.0	6.67****	0.27****
B (TB)[f]					

(*continues on next page*)

Table 2 (*Continued*).

C. Areolae retention in male rats

		Control		High dose	
Study code	Gen.	% Males	No./male	% Males	No./male
A[a]	F1	2.63	0.07	32.5***	1.29**
	F2	2.13	0.05	72.15***	3.14***
B[b]	F1	0.00	0.00	0.83	0.01
	F2	2.26	0.05	2.26	0.05
	F3	0.00	0.00	0.00	0.00

D. Areolae reduction in female rats

		Control		High dose	
Study code	Gen.	% Females	No./female	% Females	No./female
A (TP)[e]	F1	100.0	12.03	30.0***	0.53****
B (TB)[f]	F1 (total areolae)		12.0		1.0**
	F1 (normal areolae)		12.0		0.0**

Values represent incidence of nipples and/or areolae on pnd 13: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

[a]An antiandrogen (Tyl, unpublished observations).

[b]Bisphenol A [9].

[c]L = linuron [7].

[d]Flutamide [6]

[e]Testosterone proprionate [5].

[f]17ß-Trenbolone [8].

4. *Weanling necropsy*: As currently specified in both OECD and OPPTS guidelines, limited organ weights (brain, spleen, thymus) are collected from only 1 pup/sex/litter. In the OPPTS guideline, up to 3/sex/litter are necropsied, but organ weights are taken for only 1 pup/sex/litter. In my laboratory, we weigh organs from all 3/sex/litter; it does not increase the power for statistical analysis since the number of litters is the same, but it better characterizes the parameters per litter. We also recommend weighing ovaries, uterus with cervical vagina, testes, epididymides, seminal vesicles, and liver (and possibly kidneys and adrenals), especially if these are possible target organs.
5. *Acquisition of puberty (in both sexes; Table 3)*: In females, it is indicated by vaginal opening or patency (VP); in males, it is indicated by preputial separation (PPS; balanopreputial separation). Within one or a few days post VP, the female exhibits her first estrus, so age at first estrus (absolute age and/or interval from VP to first estrus) is also useful. VP is dependent on 17ß-estradiol, and PPS is dependent on testosterone (T). Less useful in males is testes descent (into the scrotal sacs from the abdominal cavity through the inguinal canal and ring), which occurs during lactation (pnd 15–20) and may be mediated by T and/or DHT. In control CD® (SD) rats in the author's laboratory, the grand mean age at VP is 31.1 days and 41.9 days at PPS (based on 20 studies from 1996 to 2002). Acquisition of puberty in both sexes is affected by body weight, so the current approach is to covary the age at acquisition by the body weight at acquisition (so that pups are at equivalent physiological states, regardless of age), or by an arbitrary calendar date or age, preferably an age that is during the time of acquisition of puberty (so that pups are at equivalent ages, regardless of physiological state). Other suggested body weights as covariates include those at weaning, birth (the last is least useful in the author's opinion; it also requires that the pups are individually identified at birth), or by some measure of weight gain during the postlactational,

Table 3 Acquisition of puberty in rats.

Study code	Gen.	Acquisition of vaginal patency (VP)				Acquisition of preputial separation (PPS)			
		Control		High dose		Control		High dose	
		BW (g)	VP age (days)	BW (g)	VP age (days)	BW (g)	PPS age (days)	BW (g)	PPS age (days)
A	F1	–	32.1	–	31.2[NS,a]	–	42.2	–	41.7[NS]
B	F1	–	31.6	–	31.1[NS]	–	42.4	–	44.0[NS]
C	F1	103.12	30.0	96.87	30.05[NS]	223.65	41.9	219.59	42.70**[,b]
D	F1	99.52	32.5	91.97[††,c]	36.0**	210.72	43.6	184.83[††]	47.8**
	F2	104.24	31.7	88.43[††]	33.8**	201.98	41.2	191.67	44.9**
E	F1	97.83	30.5	91.91	31.8**	220.07	43.1	207.01	44.7**
	F2	97.03	30.6	92.51	31.3**	202.75	42.2	197.74	43.6**
Fd	F1	102.52	30.5	92.32[††]	33.0***	215.70	41.9	194.02[†††]	45.8***
	F2	105.04	31.0	102.50	34.5***	219.74	42.1	200.13[††]	47.9***
	F3	105.59	31.3	99.04	33.8***	209.33	42.1	186.76[†††]	45.2***
G	F1	110.03	29.9	112.02	30.5[NS]	235.5	41.5	236.8	41.8[NS]
H	F1	116.19	32.3	102.26	33.5**	214.49	41.1	209.51	44.8**
I	F1	104.60	30.6	93.30[††]	32.03***	213.56	41.6	198.76[††]	43.7***
J	F1	–	31.15	–	35.04***	–	41.31	–	44.61***
K	F1	110.88	31.3	96.91[††]	32.0***	206.86	40.9	205.35	43.6***
L (AA)	F1	108.25	31.4	106.30	34.1***	208.17	40.9	207.63	45.2***
M (E2)	F1	104.16	31.2	53.94***	24.8***	212.04	41.7	241.04***	48.7***
N (F)[e]	F1	NOT DONE				NOT DONE[f]			
O (TP)[g]	F1		34.74		34.05		43.62		44.28

[a]NS = Not statistically significant (by ANOVA and pairwise tests); Study codes A through M from the author's laboratory (AA = antiandrogen, E2 = 17ß-estradiol).

[b]** = $p < 0.01$; *** = $p < 0.001$; by ANOVA or ANCOVA (body weight at acquisition is covariate).

[c]†† = $p < 0.01$; ††† = $p < 0.001$; by Dunnett's test.

[d]Bisphenol A [9].

[e]Flutamide [6].

[f]100 % of the F1 male offspring at the high dose exhibited hypospadias, so PPS could not be determined [6].

[g] Testosterone propionate [5].

prepubertal period (the selection of the end date for weight gain is problematic). Small delays in acquisition (≤3 days) may only indicate body weight-related delays in development from systemic toxicity. Large changes in delays (≥5 days) and any acceleration more likely indicate effects from endocrine disruption.

6. *Reproductive organ weights (in both sexes at adulthood; Table 4):* These should be presented as absolute and relative to terminal body weight and should include: (a) ovaries with oviducts and uterus with cervix and vagina for females; (b) testes, epididymides, prostate (whole, and dorsolateral and ventral lobes separately; dissection should be postfixation, if possible), seminal vesicles, coagulating glands, preputial glands, bulbourethral (Cowper's) glands, and levator ani/bulbocavernosus (LABC) complex for males; and (c) systemic organs in both sexes (e.g., liver, thyroid, adrenal glands, pituitary, brain [regions]). Relative weights will correct for effects on body weights (i.e., systemic toxicity). Reproductive organ weights (testes and epididymides and ovaries and uterus) can also be collected in the weanling animals selected for necropsy.

Table 4 Reproductive organ weights in male rats.

Study code	Gen.	Organ		Control	High dose
MALES					
A (AA)	F0	Paired testes:	A	3.447 ± 0.058	3.570 ± 0.064
			R	0.578 ± 0.010	0.611 ± 0.013
		Paired epididymides:	A	1.424 ± 0.019	1.424 ± 0.008
			R	0.238 ± 0.004	0.244 ± 0.005
		Prostate:	A	0.846 ± 0.034	0.821 ± 0.039
			R	0.142 ± 0.006	0.042 ± 0.005
		SV/CG:	A	2.362 ± 0.059	2.303 ± 0.063
			R	0.396 ± 0.010	0.396 ± 0.003
	F1	Paired testes:	A	3.598 ± 0.050	2.858 ± 0.179***
			R	0.600 ± 0.010	0.521 ± 0.030
		Paired epididymides:	A	1.351 ± 0.028	1.208 ± 0.054*
			R	0.226 ± 0.006	0.217 ± 0.009
		Prostate:	A	0.756 ± 0.037	0.563 ± 0.029***
			R	0.126 ± 0.006	0.103 ± 0.005**
		SV/CG:	A	2.145 ± 0.052	1.752 ± 0.094***
			R	0.358 ± 0.009	0.316 ± 0.016
B (TP)	F1	LABC (mg):	A	162.9 ± 9.1	512.8 ± 67.2*
		V. prostate (mg)	A	16.6 ± 3.0	206.7 ± 21.3*
		SV/CG (mg)	A	46.5 ± 3.0	776.5 ± 116.2*
C(TB)	F1	LABC (mg):	A	173.7 ± 9.3	458.7 ± 13.9*
		V. prostate (mg)	A	19.3 ± 0.6	38.7 ± 4.2*
		SV/CG (mg)	A	39.3 ± 4.9	90.9 ± 16.4*
D	F0	Paired testes:	A	3.48 ± 0.06	3.43 ± 0.04
			R	0.63 ± 0.01	0.80 ± 0.01***
		Paired epididymides:	A	1.46 ± 0.03	1.36 ± 0.02
			R	0.27 ± 0.01	0.32 ± 0.01***
		Prostate:	A	1.05 ± 0.06	0.73 ± 0.04***
			R	0.19 ± 0.01	0.17 ± 0.01
		SV/CG:	A	2.24 ± 0.08	1.79 ± 0.07***
			R	0.41 ± 0.01	0.41 ± 0.02
E	F1	Paired testes:	A	3.72 ± 0.05	3.24 ± 0.005***
			R	0.66 ± 0.01	0.78 ± 0.02***
		Paired epididymides:	A	1.45 ± 0.02	1.30 ± 0.02***
			R	0.26 ± 0.01	0.31 ± 0.01***
		Prostate:	A	0.73 ± 0.04	0.58 ± 0.02*
			R	0.13 ± 0.01	0.14 ± 0.01
		SV/CG:	A	2.19 ± 0.07	1.82 ± 0.07***
			R	0.39 ± 0.01	0.44 ± 0.02*
F	F2	Paired testes:	A	3.71 ± 0.05	3.27 ± 0.05***
			R	0.63 ± 0.01	0.78 ± 0.01***
		Paired epididymides:	A	1.45 ± 0.03	1.31 ± 0.02***
			R	0.25 ± 0.01	0.31 ± 0.00***
		Prostate:	A	0.65 ± 0.03	0.49 ± 0.02***
			R	0.11 ± 0.01	0.12 ± 0.01
		SV/CG:	A	2.29 ± 0.07	1.76 ± 0.07***
			R	0.39 ± 0.01	0.42 ± 0.02

(*continues on next page*)

Table 4 (*Continued*).

Study code	Gen.	Organ		Control	High dose
MALES					
G	F3	Paired testes:	A	3.65 ± 0.04	3.19 ± 0.07***
			R	0.23 ± 0.01	0.87 ± 0.02***
		Paired epididymides:	A	1.34 ± 0.02	1.20 ± 0.02***
			R	0.27 ± 0.01	0.33 ± 0.01***
		Prostate:	A	0.56 ± 0.02	0.42 ± 0.02***
			R	0.11 ± 0.00	0.11 ± 0.00
		SV/CG:	A	1.79 ± 0.01	1.43 ± 0.05***
			R	0.36 ± 0.01	0.39 ± 0.01
H (CA)	Adult exposure	Testes		3.2 ± 0.1	3.0 ± 0.1
		Epididymides		1.16 ± 0.02	0.87 ± 0.02*
		ASG		0.587 ± 0.022	0.227 ± 0.017*
		SV		0.412 ± 0.017	0.139 ± 0.009**
		Prostate		0.172 ± 0.010	0.087 ± 0.009
I (DBP)	Adult exposure	Testes		3.3 ± 0.1	3.2 ± 0.1
		Epididymides		1.02 ± 0.03	1.09 ± 0.02
		ASG		0.532 ± 0.016	0.558 ± 0.018
		SV		0.392 ± 0.011	0.417 ± 0.016
		Prostate		0.143 ± 0.006	0.144 ± 0.005
J (F)	Adult exposure	Testes		3.3 ± 0.1	3.3 ± 0.1
		Epididymides		1.26 ± 0.03	0.79 ± 0.03*
		ASG		0.588 ± 0.021	0.200 ± 0.01*
		SV		0.390 ± 0.025	0.119 ± 0.009*
		Prostate		0.190 ± 0.013	0.082 ± 0.007*
K (L)	Adult exposure	Testes		3.3 ± 0.1	3.3 ± 0.1
		Epididymides		1.16 ± 0.03	1.05 ± 0.03*
		ASG		0.597 ± 0.016	0.532 ± 0.026*
		SV		0.443 ± 0.014	0.407 ± 0.020
		Prostate		0.153 ± 0.008	0.124 ± 0.007*
L (V)	Adult exposure	Testes		3.1 ± 0.1	3.3 ± 0.0
		Epididymides		1.12 ± 0.03	1.02 ± 0.02**
		ASG		0.562 ± 0.015	0.474 ± 0.024*
		SV		0.387 ± 0.013	0.334 ± 0.019
		Prostate		0.172 ± 0.007	0.138 ± 0.008*
M (F)	Adult exposure	Testes		3.3 ± 0.1	3.4 ± 0.1
		Epididymides		1.16 ± 0.02	1.17 ± 0.02
		ASG		0.596 ± 0.016	0.526 ± 0.018*
		SV		0.437 ± 0.012	0.372 ± 0.014*
		Prostate		0.157 ± 0.008	0.152 ± 0.013

(*continues on next page*)

Table 4 (*Continued*).

Study code	Gen.	Organ	Control	High dose
MALES				
N (K)	Adult exposure	Testes	3.4 ± 0.1	3.3 ± 0.1
		Epididymides	1.22 ± 0.02	1.11 ± 0.05*
		ASG	0.578 ± 0.020	0.475 ± 0.026*
		SV	0.410 ± 0.015	0.357 ± 0.020*
		Prostate	0.166 ± 0.009	0.119 ± 0.010*

AA = antiandrogen (Tyl, unpublished observations).
TP = testosterone propionate [5].
TB = 17ß-trenbolone [8].
BPA = bisphenol A [9].
CA = cyproterone acetate.
DBP = di-*n*-butyl phthalate [10].
Fl = flutamide [10].
L = linuron [10].
V = vinclozolin [10].
Fa = fadrazole; aromatase inhibitor [11].
K = ketaconazole; testosterone biosynthesis inhibitor [11].
*, **, *** = $p < 0.05$, 0.01, 0.001 vs. control group value.
SV/CG = seminal vesicles/coagulating glands.
ASG = accessory sex glands.
LABC = levator ani bulbocavernosus complex.
A, D–G = multigeneration studies with exposure in the feed; exposures to F0 animals began at puberty, exposure to F1, F2, and F3 animals began as gametes (F3 animals not bred).
B and C = immature castrated male rats were dosed for 8–13 days by subcutaneous injection.
H–N = 15-day exposure by oral gavage to adult intact male rats.
A = absolute organ weight in grams (g).
R = relative organ weight as percentage of terminal body weight.

7. *Gross Necropsy Observations (Fig. 5)*: Careful dissections under a dissecting microscope by experienced technical staff can identify alterations (e.g., size, shape, components) in male and female reproductive systems on gestation day (gd) 20 and pnd 4 (ovaries, oviducts, uterus, cervix and vagina, cranial suspensory ligament in females; testes, epididymides [caput, corpus, cauda], afferent and efferent ducts [vas deferens], seminal vesicles, gubernaculum in males), on pnd 21 (same plus male coagulating glands, prostate, bulbourethral glands), and as adults (same plus male preputial glands, LABC complex, penis, prepuce [foreskin]), plus functional assessments (see item 9).
8. *Histopathology*: Organs with gross lesions, changes in absolute and relative weight, or identified as target organs from the list in item 7 above. Systemic organs should not be ignored, as effects on reproduction or development may be mediated by systemic toxicity.
9. *Andrology*: Includes epididymal (cauda) sperm number, motility (and progressive motility), morphology, and testicular homogenization-resistant spermatid head counts (SHC) to calculate daily sperm production (DSP) and efficiency of DSP. Epididymal sperm motility must be done immediately after demise (within 2 min in the author's laboratory); the other endpoints can be evaluated at a later date. The age of the male at demise is critical. With an average duration of spermatogenesis in rats of approximately 70 days and of epididymal transit time in rats of 10 to 14 days, there are essentially no sperm in the epididymides in a 60-day old rat, very little in the cauda of a 70-day old rat, and close to adult values in rats at and above 85 days of age.

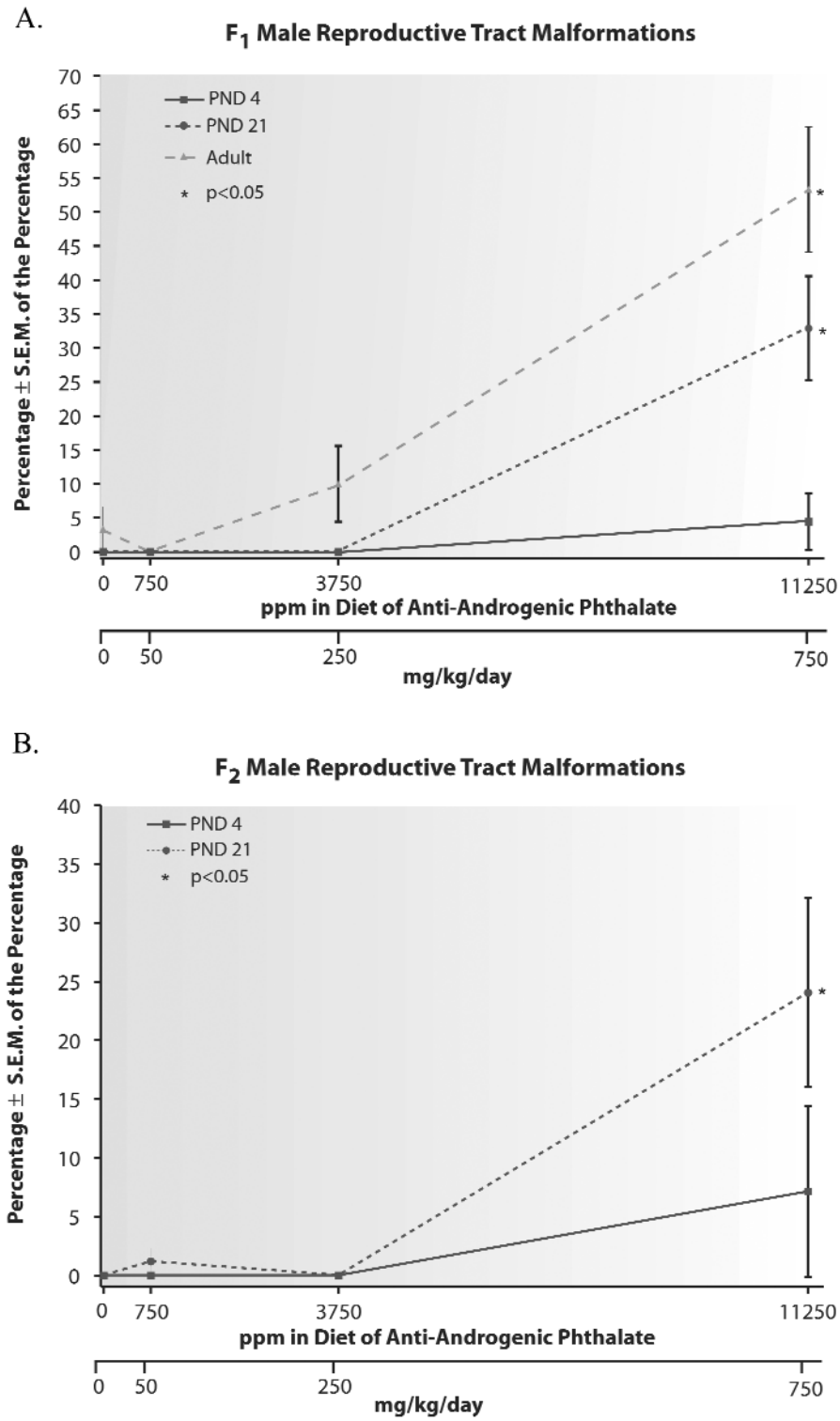


Fig. 5 Male Reproductive Tract Malformations by Gross Necropsy From an Antiandrogen (Tyl, unpublished). A. F1 males at pnd 4 and 21, and as adults; B. F2 males at pnd 4 and 21.

10. *Other*: Such as estrous cyclicity (evaluated daily for at least 2 weeks; 3 weeks is better), precoital interval, gestational length, and circulating hormone concentrations (basal and after challenge); the last is best done by longitudinal evaluations (in satellite groups of cannulated rats), etc.
11. Current regulatory guidelines specify selection of one F1 pup/sex/litter at weaning to be retained to adulthood for breeding to generate F2 offspring. Will we miss subtle lesions or effects at low incidence? Should more pups per litter be retained postwean? This will not improve statistical power since that is based on the number of litters/group, but it will improve the characterization and sensitivity of parameters for each litter. This concern is being evaluated under the USEPA EDSP contract.

Examples of inappropriate endpoints (the author's current list) are:

1. *Ovarian primordial follicle counts (Table 5)*: As currently done, this endpoint is too "noisy" (insensitive), with large variance terms. Interpretation is also a problem. A better approach would be to quantify all stages of oocyte development: primordial, small antral, large antral, and mature (Graäfian) follicles. This is very labor intensive (and, therefore, expensive), but it provides a complete assessment of oogenesis. Therefore, it should be done (in my opinion) only if other endpoints indicate possible ovarian toxicity (e.g., changes in absolute and relative ovarian weight, total implants per litter).

Table 5 Ovarian primordial follicle counts.

Study code	Gen.	Control	High dose
A	F0	140.7 ± 11.6[a]	146.1 ± 10.8
	F1	134.0 ± 8.8	135.3 ± 9.4
B	F0	132.3	114.4
	F1	150.0	148.75
C (BPA)	F0	315.9 ± 41.6	453.2 ± 26.3*
	F1	353.0 ± 35.4	409.7 ± 46.8
	F2	409.2 ± 32.7	378.0 ± 25.5
	F3	384.6 ± 55.7	355.4 ± 38.3
D	F0	242.1 ± 11.6	216.2 ± 14.1
	F1	182.2 ± 14.9	164.5 ± 9.1
E	F0	351.6 ± 38.6	405.3 ± 45.7
	F1	391.2 ± 41.4	385.5 ± 57.1
F	F0	281.3 ± 24.9	321.1 ± 51.2
	F1	357.9 ± 38.7	368.8 ± 30.0
G (AA)	F0	281.1 ± 34.1	272.9 ± 35.1
	F1	368.4 ± 26.3	414.9 ± 56.5

[a]Data presented as mean ± SEM.
* = $p < 0.05$ (in the presence of reduced absolute and relative ovarian weights).
C (BPA) = bisphenol A [9].
G (AA) = antiandrogen (Tyl, unpublished observations).
A – G is from the author's laboratory.

2. *Stage of estrus at scheduled necropsy (Table 6)*: This is a compromise endpoint; earlier draft versions of the current USEPA OPPTS reproductive toxicity testing guidelines specified that the parental females were all to be necropsied at the same stage of estrus. This endpoint, as currently specified in the finalized guidelines (1998), is considered insensitive. This is also very difficult to analyze statistically. In my laboratory, we analyze the percentage of females in each estrous stage for each group. Some laboratories only identify three stages (proestrus [P], estrus [E], and diestrous [D]); other laboratories also identify metestrus (M) for a total of four stages. Females iden-

tified as in transition (e.g., P/E, E/M, M/D) in my laboratory are included in the earlier stage (e.g., P, E, M); otherwise, the statistical analyses are close to impossible. The specified 3-week premating evaluation of estrous cyclicity is much more informative.

Table 6 Stage of estrus at scheduled necropsy and in-life estrus cycle length[a,b].

	Bisphenol A (ppm in the feed)						
	0	0.015	0.3	4.5	75	750	7500
A. F0 Females							
No. females evaluated	30	29	30	30	29	29	29
No. in proestrus	11$^{\pounds\pounds}$	4$^{\Phi}$	10	3$^{\Phi}$	5	3$^{\Phi}$	1$^{\Phi\Phi}$
% in proestrus	40.74$^{\Psi\Psi\Psi}$	13.79	35.71	10.34	18.52	10.34	3.70
No. in estrus	4	4	5	9	5	6	0
% in estrus	14.81	13.79	17.86	31.03	18.52	20.69	0.00
No. in metestrus	3	4	3	1	3	0	1
% in metestrus	11.11	13.79	10.71	3.45	11.11	0.00	3.70
No. in diestrus	9$^{\pounds\pounds}$	17	10	16	14	20$^{\Phi}$	25$^{\Phi\Phi\Phi}$
% in diestrus	33.33$^{\Psi\Psi\Psi}$	58.62	35.71	55.17	51.85	68.97	92.59
Estrous cycle length (days)	4.58 ± 0.25	4.41 ± 0.09	4.48 ± 0.20	4.50 ± 0.11	4.57 ± 0.14	4.45 ± 0.18	4.26 ± 0.09
B. F1 Females							
No. females evaluated	27	30	29	30	28	30	28
No. in proestrus	5	8	3	3	5	4	4
% in proestrus	20.00	26.67	10.34	10.34	18.52	14.29	15.38
No. in estrus	0	5	4	4	2	2	0
% in estrus	0.00	16.67	13.79	13.79	7.41	7.14	0.00
No. in metestrus	2	0	2	0	0	0	0
% in metestrus	8.00	0.00	6.90	0.00	0.00	0.00	0.00
No. in diestrus	18	17	20	22	20	22	22
% in diestrus	72.00$^{\Psi}$	56.67	68.97	75.86	74.07	78.57	84.62
Estrous cycle length (days)	4.41 ± 0.10	4.47 ± 0.13	4.19 ± 0.09	4.70 ± 0.23	4.94 ± 0.21	4.40 ± 0.18	4.54 ± 0.11
C. F2 Females							
No. females evaluated	30	30	29	29	30	30	28
No. in proestrus	2	6	1	3	4	3	0
% in proestrus	7.14	20.00	3.45	10.34	13.33	10.71	0.00
No. in estrus	8$^{\pounds\pounds}$	6	9	5	13	11	0$^{\Phi}$
% in estrus	28.57	20.00	31.03	17.24	43.33	39.29	0.00
No. in metestrus	3	1	4	3	2	0	1
% in metestrus	10.71	3.33	13.79	10.34	6.67	0.00	3.70
No. in diestrus	15$^{\pounds\pounds}$	17	15	18	11	14	26$^{\Phi\Phi}$
% in diestrus	53.57	56.67	51.72	62.07	36.67	50.00	96.30
Estrous cycle length (days)	4.54 ± 0.21	4.61 ± 0.24	4.39 ± 0.14	4.47 ± 0.23	4.17 ± 0.07	4.56 ± 0.24	4.56 ± 0.11
D. F3 Females[c]							
No. females evaluated	30	30	30	29	30	30	30
No. in proestrus	3	2	4	3	5	9	5
% in proestrus	10.34$^{\Psi}$	6.67	13.33	10.34	16.67	31.03	16.67
No. in estrus	9	12	12	13	15	4	10
% in estrus	31.03	40.00	40.00	44.83	50.00	13.79	33.33
No. in metestrus	2	1	4	3	2	3	3
% in metestrus	6.90	3.33	13.33	10.34	6.67	10.34	10.00
No. in diestrus	15	15	10	10	8	13	12
% in diestrus	51.72	50.00	33.33	34.48	26.67	44.83	40.00
Estrous cycle length (days)	4.32 ± 0.18	4.34 ± 0.12	4.32 ± 0.09	4.39 ± 0.18	4.66 ± 0.21	4.59 ± 0.21	4.31 ± 0.09

[a]Stage of estrus (Tyl, unpublished observations); in-life estrus cycle length [9].

(*continues on next page*)

Table 6 (*Continued*).

[b]For presentation and statistical analysis purposes those females in two stages were pooled in the following manner: proestrus/estrus was considered proestrus; estrus/metestrus and estrus/diestrus were considered estrus; metestrus/diestrus was considered metestrus; and diestrus/proestrus was considered diestrus. The females for which the stage could not be determined or no cells were present were not included in the statistical analysis.
[c]F3 females were not mated prior to scheduled necropsy as adults.
$^{\pounds\pounds}p < 0.01$; Chi-square test.
$^{\Phi}p < 0.05$; Fisher Exact test.
$^{\Phi\Phi}p < 0.01$; Fisher Exact test.
$^{\Phi\Phi\Phi}p < 0.001$ Fisher Exact test.
$^{\Psi}p < 0.05$; Cochran–Armitage test.
$^{\Psi\Psi\Psi}p < 0.01$; Cochran–Armitage test.

3. *Single sample of circulating hormone concentrations*: Due to the pulsatile nature of circulating hormone levels and complex positive and negative feedback loops, a single blood sample from adult animals at necropsy (i.e., a "single snapshot in time") is considered insensitive. It would only detect profound changes in the animals that would also be detectable by results in other parameters (see above for suggestions for longitudinal evaluation).

CONCLUSIONS

1. Regulatory guidelines represent the minimum requirements. There is no reason not to enhance them if and as appropriate.
2. Regulatory guidelines will change over time as the science improves and new societal, regulatory, and scientific concerns arise. The sooner we begin discussions on how to improve them, the sooner the regulatory agencies will consider amending them.
3. Basic scientists will continue to provide new endpoints, mechanisms, and approaches for consideration of inclusion into the guidelines. Applied scientists will continue to identify new needs and to refine parameters from the basic researcher. This interaction between basic and applied science (between specific and apical investigations) must be encouraged, supported, and valued.
4. The objectives in performing these studies are to perform sound science and to provide useful information to the regulatory agencies for risk assessment. The better the studies, the better the risk assessment. My last recommendation is for the regulatory agencies to periodically convene workshops with basic (academic, governmental) scientists, industrial scientists, contract performing laboratories, and regulators to discuss the rapid increase in information (e.g., target tissues, critical life phases of exposure, endpoints, interpretation), and the assessment of current and proposed endpoints. My hope is that this will lead to periodic revision and improvement of the regulatory testing guidelines.

ACKNOWLEDGMENTS

The author wishes to thank Ms. Christina Myers (supervisor for reproductive toxicity studies and data analyst), Ms. Melissa Marr (laboratory supervisor), and her reproductive and developmental toxicology staff at RTI International who did the work and made this presentation possible. Thanks to Ms. Cathee Winkie (Dr. Tyl's administrative coordinator) for her patient typing (and retyping…) of the presentation, poster, and manuscript. Thanks to Ms. Christina Myers and Ms. Cathee Winkie for their assistance in generating the tables and figures.

REFERENCES

1. U.S. Environmental Protection Agency (EPA). Toxic Substances Control Act Test Guidelines; Final Rule, Section 798.4700, reproduction and fertility effects, *Federal Register* **50** (188), 43834–43838 (Friday, 27 September 1985).
2. U.S. Environmental Protection Agency (EPA). 40CFR Part 799, Toxic Substances Control Act Test Guidelines; Final Rule, Section 799.9380, TSCA reproduction and fertility effects, *Federal Register* **62** (158), 43834–43838 (Friday, 15 August 1997).
3. U.S. Environmental Protection Agency (EPA). Office of Prevention, Pesticides and Toxic Substances (OPPTS), Health Effects Testing Guidelines, OPPTS 870.3800, reproduction and fertility effects (Final Guideline, August 1998).
4. Organization for Economic Cooperation and Development, *OECD Guideline for the Testing of Chemicals; Proposal for Updating Guideline 414: Prenatal Developmental Toxicity Study*, pp. 1–11 (adopted 22 January 2001).
5. U.S. Food and Drug Administration (FDA). "Redbook 2000", Toxicological Principles for the Safety of Food Ingredients, Center for Food Safety and Applied Nutrition, Office of Premarket Approval (CFSAN, OPA) 20 July 2000.
6. C. J. Wolf, A. Hotchkiss, J. S. Ostby, G. A. LeBlanc, L. E. Gray, Jr. *Toxicol. Sci.* **65** (1), 71–86 (2002).
7. B. S. McIntyre, N. J. Barlow, P. M. D. Foster. *Toxicol. Sci.* **62**, 236–249 (2001).
8. B. S. McIntyre, N. J. Barlow, P. M. D. Foster. *Toxicol. Sci.* **65** (1), 62–70 (2002).
9. V. S. Wilson, C. Lambright, J. Ostby, L. E. Gray, Jr. *Toxicol. Sci.* **70** (2), 202–211 (2002).
10. R. W. Tyl, C. B. Myers, M. C. Marr, B. F. Thomas, A. R. Keimowitz, D. R. Brine, M. M. Veselica, P. A. Fail, T. Y. Chang, J. C. Seely, R. L. Joiner, J. H. Butala, S. S. Dimond, S. Z. Cagen, R. N. Shiotsuka, G. D. Stropp, J. M. Waechter. *Toxicol. Sci.* **68** (1), 121–146 (2002).
11. J. C. O'Connor, S. R. Frame, G. S. Ladics. *Toxicol. Sci.* **69** (1), 92–108 (2002a).
12. J. C. O'Connor, S. R. Frame, G. S. Ladics. *Toxicol. Sci.* **69** (1), 79–91 (2002b).