2009

Inter-Laboratory Control Data for Reproductive Endpoints Required in the OPPTS 870.3800/ OECD 416 Reproduction and Fertility Test

M. S. Marty
Dow Chemical Company, mmarty@dow.com

B. Allen
Bruce Allen Consulting

R. E. Chapin
Pfizer Global Research and Development

R. Cooper
U.S. Environmental Protection Agency

G. P. Daston
Procter & Gamble

See next page for additional authors

Authors
M. S. Marty, B. Allen, R. E. Chapin, R. Cooper, G. P. Daston, J. A. Flaws, P. M. D. Foster, S. L. Makris, E. Mylchreest, D. Sandler, and R. W. Tyl

Original Article

Inter-Laboratory Control Data for Reproductive Endpoints Required in the OPPTS 870.3800/OECD 416 Reproduction and Fertility Test

M. S. Marty,[1]* B. Allen,[2] R. E. Chapin,[3] R. Cooper,[4] G. P. Daston,[5] J. A. Flaws,[6] P. M. D. Foster,[7] S. L. Makris,[8] E. Mylchreest,[9] D. Sandler,[10] and R. W. Tyl[11]

[1]The Dow Chemical Company, Midland, Michigan
[2]Bruce Allen Consulting, Chapel Hill, North Carolina
[3]Pfizer Global Research and Development, Groton, Connecticut
[4]U.S. EPA NHEERL, Research Triangle Park, North Carolina
[5]Procter & Gamble Company, Cincinnati, Ohio
[6]University of Illinois, Urbana, Illinois
[7]National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina
[8]U.S. EPA, Washington, DC
[9]DuPont Haskell Laboratories, Newark, Delaware
[10]ILSI Health and Environmental Sciences Institute, Washington, DC
[11]RTI International, Research Triangle Park, North Carolina

BACKGROUND: The U.S. EPA revised the Reproduction and Fertility Effects Test Guideline (OPPTS 870.3800/OECD 416) in 1998, adding numerous endpoints in an effort to incorporate new methodologies, improve the sensitivity for detecting reproductive toxicants, and more efficiently utilize study animals. Many of these new endpoints have not been used in regulatory reproductive toxicology studies prior to their inclusion in the test guidelines; thus, the Health and Environmental Sciences Institute (HESI) of the International Life Sciences Institute (ILSI) initiated the Reproductive Endpoints Project to examine the utility of these new endpoints. **METHODS:** This report provides a retrospective analysis of 43 multi-generation studies (16 in Wistar rats, 27 in Sprague-Dawley rats) conducted according to the latest version of the test guidelines. It focuses on vehicle (negative) control values (means and ranges) for the various endpoints to examine inter-laboratory variability. **RESULTS:** Based on the compiled data, the most variable endpoints across laboratories and their associated coefficients of variation (CV) for each generation were: percent abnormal sperm (166–205%), testicular spermatid concentration (126–147%), postimplantation loss (97–104%), primordial follicle counts (69%, only measured in P2 females), and epididymal sperm concentration (52–57%). Absolute and relative prostate and thymus weights, weanling uterine weights, and anogenital distance had CVs of 25–50%. Sources of variability included procedural differences between laboratories, inherent biological variability, and/or small sample sizes for some endpoints. **CONCLUSIONS:** These inter-laboratory control data provide a means for laboratories to review their performance on reproductive toxicity measures, and provide perspective for interpreting their own control data and data from treated animals. *Birth Defects Res (Part B)* 86:470–489, 2009. © 2009 Wiley-Liss, Inc.

Key words: ***reproductive toxicity; multi-generation; regulatory; test guideline; reproduction; fertility; endocrine***

INTRODUCTION

In 1998, the United States Environmental Protection Agency (U.S. EPA) revised the Reproductive and Fertility Effects Test Guideline (OPPTS 870.3800; multi-generation studies), adding new endpoints in an effort to improve the sensitivity of the study to detect reproductive toxicants. In addition, regulators felt that additional information could be gained along with more efficient use of study animals. New endpoints that were added to the study are listed in Table 1. The difficulty with the addition of these new endpoints is that many of these endpoints had not been used in a regulatory setting prior

*Correspondence to: M. S. Marty, The Dow Chemical Company, 1803 Building, Midland, MI 48674. E-mail: mmarty@dow.com
Received 18 August 2009; Revised 20 August 2009; Accepted 21 August 2009

E. Mylchreest's present address is Wyeth Research, 641 Ridge Road, Chazy, New York, 12921.

Disclaimer: The views expressed in this document are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

D. Sandler's present address is U.S. Department of Agriculture, Food Safety and Inspection Service, 1400 Independence Ave., S.W., Washington, DC 20005.

Published online in Wiley InterScience (www.interscience.wiley.com)
DOI: 10.1002/bdrb.20208

Table 1
New Endpoints Added to OPPTS 870.3800 in 1998 (or OECD 416 in 2001)

Adult female:
Evaluation of the estrous cycle for 3 weeks prior to mating in P1 and P2 females
Organ weights in P1 and P2 females[a]
Primordial follicle counts in P2 females
Postimplantation loss in P1 and P2 females
Expanded histopathology of parental reproductive organs
Adult male:
Organ weights in P1 and P2 males[a]
Epididymal sperm counts in P1 and P2 males
Testicular spermatid counts in P1 and P2 males
Sperm motility in P1 and P2 males
Sperm morphology in P1 and P2 males
Expanded histopathology of parental reproductive organs
Offspring:
Puberty onset in F1/P2 males and females
Anogenital distance in F2 offspring if triggered by a change in sex ratio or age at puberty onset
Weanling organ weights (brain, spleen, and thymus) in the F1 and F2 offspring (1 pup/sex/litter)[b]
Weanling necropsy with the examination of 3 pups/sex/litter compared with 1 pup/sex/litter required previously
Histopathological examination of treatment-related macroscopic abnormalities in weanlings

[a]Reproductive organs, brain, pituitary, liver, kidneys, adrenals, spleen, and target organs.
[b]Weanling uterine weight was added as an endpoint by the Japanese Ministry of Forestry and Fisheries (JMAFF) in 2002.

to their inclusion in the test guidelines, and their utility remains unproven.

In 1997, the International Life Sciences Institute (ILSI) Health Sciences Institute (HESI) sponsored a workshop to discuss the new endpoints and their interpretation. At that time, workshop participants agreed that it would be advantageous to review data in ~5 years after adopting the new guidelines to determine whether improvement in the sensitivity of the Reproductive and Fertility Effects Test Guideline to detect reproductive toxicants had been achieved.

Thus, a follow-up study was sponsored by the ILSI HESI Developmental and Reproductive Toxicology Committee to evaluate the utility of the new reproductive endpoints added to the Reproduction and Fertility Effects test guideline in 1998 (OPPTS 870.3800 or the OECD 416 guideline adopted in 2001). As part of this evaluation, the steering committee agreed to compile a data set of inter-laboratory vehicle control values that could be used by laboratories and regulators to establish typical mean values and ascertain the normal variability around the means for the new endpoints added to the test guideline. The second goal of the HESI project, evaluating the performance of the new endpoints relative to more traditional endpoints, will be discussed in a separate report.

In order to gather multi-generation study data for review, industrial companies, trade groups, and contract research laboratories were contacted and asked to submit study data collected in accordance with the OPPTS 870.3800 (OECD 416) test guideline. Forty-four reports from 11 companies were received. The data were entered into a database to facilitate database-level analyses. Results of these analyses were evaluated by a steering committee comprised of experts from industry, government, and academia. This report presents data on inter-laboratory control values for the new endpoints, as well as some more traditional endpoints. Issues related to variability are discussed.

MATERIALS AND METHODS

Data Submission

Numerous laboratories known to conduct toxicity studies according to the OPPTS 870 guidelines were contacted to ascertain whether these laboratories could submit data for review. Participating laboratories submitted data collected in accordance with the OPPTS 870.3800 and/or OECD 416 guidelines. Forty-four reports from 11 companies were received. Companies were asked to submit finalized study reports with summary tables (i.e., summary data containing means and standard deviations [SD] for continuous variables, incidence data for counted endpoints, and sample sizes). Due to the volume of data contained in each multi-generation study, individual animal data and appendices were not included in the submissions. To insure confidentiality, the text of the reports was modified to remove any chemical-specific information (e.g., chemical identity, structure, chemical class, etc.). The data were entered into a database to facilitate database-level analyses. In addition, companies were asked to answer one supplemental question as to whether the test material affected the reproductive system (i.e., organ weights or histopathology) in a shorter term study. Results of these analyses were evaluated by a steering committee comprised of experts from industry, government, and academia.

Definitions

Generations are defined as P1: first-parental generation randomized and placed on study as young adults; F1: first-generation offspring produced by the P1 parents; P2: second-parental generations selected from a subset of the F1 offspring; and F2: second-generation offspring produced by the P2 parents.

Data Analysis

Control values (means and ranges) for various reproductive parameters were compiled from the

database; thus, the mean values reported in Tables 3–16 are the mean of the study means given in each report. The standard deviations represent the variance of mean values across studies. Data were analyzed separately by strain as either Sprague-Dawley-derived (Crl:CD®(SD)IGS BR, Crl:CD®(SD)IGS BR–VAF/Plus®, Crl:CD (SD), Crl:CD® (SD) BR, Crl:CD® BR, Crl:CD® BR–VAF/Plus®, CD® [Sprague-Dawley-derived]) or Wistar-derived (CrlGlxBrlHan:WI, Chbb = THOM [SPF]). When presenting inter-laboratory control values, minimum, maximum, and percentile values were given so that the reader might understand the distribution of the mean values across studies. Data are presented as mean, minimum value, 25th percentile, median (50th percentile), 75th percentile, and maximum value. Thus, the minimum value is the lowest mean reported across studies, the maximum value is the highest mean reported across studies, and so on. Percentile values are not interpolated, but rather represent actual values at the defined percentiles.

RESULTS AND DISCUSSION

Inter-Laboratory Control Data: General Principles

Data from Reproductive and Fertility Guideline studies (OPPTS 870.3800/OECD 416) were entered into a database and summary control data (means and variances) were calculated by strain (Sprague-Dawley- or Wistar-derived). There were 44 studies submitted to the data set: 27 for Sprague-Dawley-derived rats, 16 studies for Wistar-derived rats, and 1 study for Alpk:AP$_f$SD rats. For each set of endpoints, data for Sprague-Dawley-derived rats appear in Table 1 followed by data for

Table 2
ANOVA Analysis for Intra- and Inter-Laboratory Variability for Highly Variable Endpoints

	CD rats			
	Intra-laboratory variability		Inter-laboratory variability	
Endpoint	1st Gen.	2nd Gen.	1st Gen.	2nd Gen.
% Abnormal sperm			X	X
Testicular spermatid concentration	X	X		
Epididymal sperm concentration			X	X
Post implantation loss			X	X
Primordial follicle counts	NA	X		
Prostate Wts (abs)	≈	X	≈	
Prostate Wts (rel)		X	X	
Thymus Wts (abs), female adults		X	X	
Thymus Wts (abs), male adults	≈	≈	≈	≈
Thymus Wts (rel), female adults		X	X	
Thymus Wts (rel), male adults	X			X

Based on a subset of data for Sprague-Dawley rats from laboratories that reported results for multiple studies. X identifies whether intra- or inter-laboratory variability was greatest for each endpoint for each generation. ≈ identifies values in which intra- and inter-laboratory variability were similar (i.e., ≤12% difference). Shaded boxes identify consistent results across generations. NA = not applicable; primordial follicle counts required in P2 females only.

Table 3
Control Values: Male Reproductive Parameters in Sprague-Dawley-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Fertility index (%)	P1	27	89.1	6.6	73.3	86.4	90.0	93.3	100.0
	P2	27	86.4	8.6	65.5	80.0	86.7	93.3	100.0
Mating index (%)	P1	27	95.7	5.4	80.0	93.3	96.7	100.0	100.0
	P2	27	94.7	4.7	82.8	93.3	96.7	96.7	100.0
Age at puberty onset (d)	P1	–	–	–	–	–	–	–	–
	P2	27	45.3	2.1	41.2	43.6	45.2	47.0	49.0
Body weight at PPS (g)	P1	–	–	–	–	–	–	–	–
	P2	15	210.9	39.7	118.9	207.6	227.0	233.3	240.7
Conc. epid. sperm (no. $\times 10^6$/g)	P1	19	609.9	318.2	6.4	444.1	476.2	906.7	1,144.0
	P2	18	595.7	337.6	8.2	420.8	506.6	791.0	1,189.4
Conc. test. sperm. (no. $\times 10^6$/g)	P1	17	142.2	178.6	72.9	80.5	91.6	124.6	826.5
	P2	17	159.5	235.1	76.2	85.7	93.2	133.5	1,063.9
Sperm motility (%)	P1	23	82.2	11.3	50.9	82.8	85.5	89.0	93.1
	P2	23	82.7	9.3	55.0	79.9	85.9	87.7	96.2
Abnormal sperm (%)	P1	14	2.6	4.2	0.0	0.1	1.1	3.1	15.9
	P2	14	3.8	7.8	0.0	0.2	0.9	3.6	30.0

Table 4
Control Values: Male Reproductive Parameters in Wistar-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Fertility index (%)	P1	16	95.3	5.4	84.0	92.0	96.3	100.0	100.0
	P2	16	94.4	4.5	87.0	91.5	95.8	97.0	100.0
Mating index (%)	P1	16	98.7	2.8	90.0	99.2	100.0	100.0	100.0
	P2	16	98.7	2.0	95.8	96.0	100.0	100.0	100.0
Age at puberty onset (d)	P1	–	–	–	–	–	–	–	–
	P2	17	43.5	1.4	40.9	42.8	43.6	44.0	47.1
Body weight at PPS (g)	P1	–	–	–	–	–	–	–	–
	P2	7	162.7	31.3	92.2	170.0	174.2	176.1	180.6
Conc. epid. sperm (no. $\times 10^6$/g)	P1	14	569.8	81.2	451.0	513.3	575.5	638.0	697.0
	P2	14	608.5	88.8	444.0	548.5	641.0	674.8	705.0
Conc. test. sperm. (no. $\times 10^6$/g)	P1	14	102.6	18.8	55.0	95.5	102.5	112.8	131.0
	P2	14	108.3	18.5	82.0	96.0	107.5	119.5	144.0
Sperm motility (%)	P1	17	87.3	3.3	81.0	85.9	89.0	90.0	91.0
	P2	17	86.9	4.1	78.1	84.0	87.0	90.0	93.0
Abnormal sperm (%)	P1	15	2.8	1.1	1.7	2.0	2.1	3.3	5.2
	P2	16	2.2	0.8	1.0	1.6	2.4	2.7	3.5

Table 5
Control Values: Female Reproductive Parameters in Sprague-Dawley-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Fertility index (%)	P1	27	89.8	5.9	73.3	88.6	91.7	93.3	96.7
	P2	27	87.5	7.7	72.4	80.0	89.7	93.3	100.0
Gestation index (%)	P1	18	99.2	2.6	89.5	100.0	100.0	100.0	100.0
	P2	18	99.0	1.9	95.2	100.0	100.0	100.0	100.0
Mating index (%)	P1	27	96.2	4.4	86.7	93.3	96.7	100.0	100.0
	P2	27	94.8	4.7	82.8	93.3	96.7	96.7	100.0
Postimplantation loss	P1	7	4.9	5.0	0.5	1.0	1.2	8.6	12.2
	P2	7	3.9	3.7	0.7	0.9	1.2	7.3	8.4
Age at puberty onset (d)	P1	–	–	–	–	–	–	–	–
	P2	27	33.5	1.9	29.0	32.3	33.7	34.3	38.8
Body weight at VO (g)	P1	–	–	–	–	–	–	–	–
	P2	15	124.8	40.4	90.5	105.7	109.1	124.7	223.0
Estrous cycle length (d)	P1	18	4.4	0.3	4.1	4.2	4.3	4.5	5.3
	P2	18	4.4	0.3	4.0	4.3	4.3	4.5	5.5
Time to mating (d)	P1	18	2.9	0.6	1.8	2.5	2.8	3.1	4.4
	P2	18	3.1	0.5	2.4	2.7	3.0	3.4	4.0
Gestational length (d)	P1	25	22.1	0.4	21.6	21.8	22.0	22.3	23.2
	P2	26	22.1	0.4	21.5	21.8	22.0	22.4	22.9
Primordial follicle count (total)	P1	–	–	–	–	–	–	–	–
	P2	19	116.1	80.2	10.7	89.0	114.6	146.5	384.0

Wistar-derived rats (e.g., male reproductive parameters for Sprague-Dawley and Wistar rats are reported in Tables 3 and 4, respectively). Due to the limited number of studies submitted, data for Alpk:AP$_f$SD rats are not presented. As noted by the sample sizes in the various data tables, means and variances were not reported for all endpoints in all of the submitted studies.

To compile the database, the mean value for each continuous endpoint was collated from each submitted study. In order to clearly represent the mean values in the data set, tables include a presentation of the mean (i.e., mean of the compiled means), median of the compiled mean values, minimum study mean, and maximum study mean values for each endpoint, as well as other percentile values in the distribution of the study means. The data distribution can be used to glean information on the frequency with which values appear and, hence, how typical a value is likely to be. In cases where the mean and median are similar, this suggests that the data are more evenly distributed. However, in some cases, the mean and median differ. For example, the mean for abnormal sperm is 2.56 and 3.80% for the first (P1) and second (P2) parental generations, respectively, whereas the median values for these generations are 1.05 and 0.90%, respectively (Table 3, discussed in greater detail below). In this case, the mean value is similar to the 75th percentile value, both of which differ markedly from the maximum values (15.9–30%). This disparity between the 75th percentile and the maximum values suggests that such extreme maximum values were not common across laboratories.

Table 6
Control Values: Female Reproductive Parameters in Wistar-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Fertility index (%)	P1	16	95.2	7.0	75.0	92.0	98.3	100.0	100.0
	P2	16	95.4	4.6	87.0	91.9	95.9	100.0	100.0
Gestation index (%)	P1	16	99.0	1.8	96.0	99.0	100.0	100.0	100.0
	P2	16	99.7	1.3	95.0	100.0	100.0	100.0	100.0
Mating index (%)	P1	16	98.7	2.8	90.0	99.2	100.0	100.0	100.0
	P2	16	98.7	2.0	95.8	96.0	100.0	100.0	100.0
Postimplantation loss	P1	8	1.4	1.4	0.3	0.4	0.8	1.9	4.2
	P2	8	1.2	0.7	0.4	0.8	1.0	1.4	2.5
Age at puberty onset (d)	P1	–	–	–	–	–	–	–	–
	P2	17	32.6	1.8	29.9	31.5	32.0	34.1	35.3
Body weight at VO (g)	P1	–	–	–	–	–	–	–	–
	P2	7	103.5	28.6	90.7	91.5	92.4	95.2	168.1
Estrous cycle length (d)	P1	16	4.2	0.4	3.9	4.0	4.1	4.3	5.3
	P2	16	4.4	0.5	3.6	4.1	4.2	4.6	5.5
Time to mating (d)	P1	15	2.7	0.2	2.3	2.6	2.7	2.9	3.1
	P2	15	2.5	0.5	1.9	2.2	2.3	2.6	3.7
Gestational length (d)	P1	16	22.0	0.2	21.8	21.8	21.9	22.0	22.3
	P2	16	22.0	0.1	21.8	21.9	22.0	22.1	22.3
Primordial follicle count (total)	P1	–	–	–	–	–	–	–	–
	P2	15	165.1	57.4	51.3	150.5	169.0	196.5	250.0

Table 7
Control Values: P1 and P2 Reproductive Organ Weights for Adult Sprague-Dawley-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Rel. epididymides wt (mean unilateral) (%)	P1	18	0.131	0.009	0.117	0.125	0.131	0.138	0.154
	P2	18	0.127	0.009	0.112	0.122	0.125	0.131	0.150
Epididymides weight (mean unilateral) (g)	P1	26	0.738	0.038	0.683	0.716	0.737	0.740	0.838
	P2	26	0.734	0.042	0.630	0.716	0.739	0.748	0.879
Rel. prostate wt (%)	P1	18	0.186	0.063	0.105	0.133	0.184	0.227	0.312
	P2	18	0.177	0.051	0.089	0.128	0.182	0.211	0.272
Prostate weight (g)	P1	18	1.063	0.400	0.592	0.758	1.030	1.219	2.155
	P2	18	1.039	0.321	0.527	0.810	1.130	1.239	1.652
Rel. sem. ves. wt (%)	P1	18	0.413	0.075	0.252	0.363	0.423	0.470	0.522
	P2	18	0.387	0.073	0.272	0.322	0.403	0.436	0.511
Sem. ves. weight (g)	P1	18	2.331	0.406	1.627	2.012	2.326	2.639	2.908
	P2	18	2.234	0.354	1.668	1.958	2.320	2.449	2.855
Rel. testes weight (mean unilateral) (%)	P1	18	0.315	0.022	0.268	0.306	0.312	0.333	0.348
	P2	18	0.313	0.015	0.297	0.301	0.308	0.323	0.349
Testes weight (mean unilateral) (g)	P1	18	1.782	0.076	1.667	1.735	1.753	1.829	1.990
	P2	18	1.822	0.101	1.629	1.761	1.843	1.872	2.069
Females									
Rel. ovarian wt (mean unilateral) (%)	P1	21	0.019	0.003	0.015	0.017	0.018	0.022	0.025
	P2	22	0.019	0.003	0.016	0.017	0.018	0.020	0.027
Ovarian weight (mean unilateral) (g)	P1	21	0.061	0.008	0.052	0.055	0.060	0.066	0.081
	P2	23	0.062	0.008	0.050	0.057	0.060	0.066	0.081
Rel. uterine wt (%)	P1	18	0.213	0.039	0.164	0.183	0.214	0.229	0.309
	P2	19	0.204	0.037	0.140	0.173	0.209	0.231	0.266
Uterine wt (g)	P1	18	0.671	0.116	0.490	0.580	0.660	0.717	0.970
	P2	20	0.650	0.103	0.493	0.565	0.635	0.721	0.850

For some endpoints, variance (as measured by SD) exceeded 20% of the mean value within the same generation, resulting in relatively large coefficients of variation (CV). These ''high variability'' endpoints in Sprague-Dawley-derived rats included: postimplantation loss (CV 97–104%), epididymal sperm concentration (CV 52–57%), testicular spermatid concentration (CV 126–147%), percent abnormal sperm (CV 166–205%), primordial follicle count (total) (CV 69%), anogenital distance (CV 28% in males, 42% in females), prostate weights (CV 29–38%), thymus weights (CV 29–35% in males, 14–32% in females), and weanling uterine weights (CV 16–52%).

Table 8
Control Values: P1 and P2 Reproductive Organ Weights for Adult Wistar-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Rel. epididymides wt (mean unilateral) (%)	P1	17	0.145	0.012	0.120	0.139	0.145	0.153	0.168
	P2	17	0.144	0.012	0.123	0.136	0.141	0.154	0.165
Epididymides weight (mean unilateral) (g)	P1	17	0.637	0.061	0.556	0.580	0.648	0.690	0.723
	P2	17	0.634	0.051	0.542	0.586	0.656	0.672	0.704
Rel. prostate wt (%)	P1	17	0.260	0.033	0.179	0.254	0.263	0.277	0.313
	P2	17	0.241	0.031	0.168	0.234	0.248	0.260	0.290
Prostate weight (g)	P1	17	1.141	0.157	0.959	1.031	1.073	1.264	1.432
	P2	17	1.063	0.143	0.800	0.966	1.058	1.156	1.318
Rel. sem. ves. wt (%)	P1	17	0.295	0.024	0.245	0.282	0.296	0.309	0.330
	P2	17	0.278	0.024	0.244	0.258	0.275	0.298	0.319
Sem. ves. weight (g)	P1	17	1.313	0.211	1.001	1.145	1.312	1.423	1.860
	P2	17	1.231	0.164	1.040	1.136	1.207	1.283	1.740
Rel. testes weight (mean unilateral) (%)	P1	17	0.417	0.045	0.330	0.409	0.420	0.453	0.485
	P2	17	0.432	0.052	0.330	0.398	0.429	0.476	0.536
Testes weight (mean unilateral) (g)	P1	17	1.829	0.075	1.696	1.767	1.831	1.884	1.948
	P2	17	1.890	0.051	1.816	1.843	1.888	1.911	2.001
Females									
Rel. ovarian wt (mean unilateral) (%)	P1	17	0.023	0.003	0.018	0.021	0.024	0.025	0.027
	P2	17	0.023	0.002	0.018	0.022	0.024	0.025	0.027
Ovarian weight (mean unilateral) (g)	P1	17	0.059	0.007	0.048	0.053	0.057	0.062	0.073
	P2	17	0.060	0.008	0.047	0.054	0.059	0.064	0.076
Rel. uterine wt (%)	P1	17	0.266	0.040	0.189	0.239	0.267	0.297	0.327
	P2	17	0.264	0.040	0.186	0.244	0.272	0.292	0.324
Uterine wt (g)	P1	17	0.688	0.102	0.533	0.624	0.652	0.720	0.864
	P2	17	0.686	0.131	0.483	0.626	0.672	0.725	0.980

There are numerous factors that contribute to data variability. First, there is inherent biological variability in an endpoint, resulting in diverse values within and across laboratories. Second, different laboratories may use different procedures to evaluate some of the endpoints, thereby contributing to variability in the values across laboratories. In addition, there are differences across laboratories in the ages of the animals, source of animals and animal diets used, among other factors. A combination of these factors is likely. Additional discussion of some factors that contribute to specific endpoint variability is included below.

To examine potential sources of variability, the contributions of inter-laboratory and intra-laboratory variability were examined for highly variable endpoints using analysis of variance (ANOVA). The intra-laboratory variability was equal to the mean square error (MSE), whereas the inter-laboratory variability was equal to [(model mean square−MSE)/number of observations per laboratory]. Due to the larger number of studies available, data for Sprague-Dawley rats from laboratories that reported multiple studies were used for this comparison. The intra-laboratory variability represents variability contributed by both inherent biological variability and procedural differences in sampling across studies conducted in the same laboratory. Presumably, procedural differences within the same laboratory are minimal if laboratories have documented standard operating procedures (SOPs). The inter-laboratory variability for each endpoint represents procedural differences across laboratories as inherent biological variability is accounted for in the intra-laboratory measurement. If an endpoint is highly variable across laboratories, this may suggest that some performance criteria are needed to insure adequate data collection (e.g., standardize the number of animals and/or sections examined for ovarian follicle counts). As shown in Table 2, inter-laboratory procedural differences appear to account for the majority of variability in postimplantation loss, percent abnormal sperm, and epididymal sperm concentration. In contrast, inherent biological variability appears to account for greater variance in testicular spermatid concentration and primordial follicle counts. Intra-laboratory variability was shown to impact absolute prostate weights in P2 adult males, although intra- and inter-laboratory differences both contributed to differences in prostate weights in P1 males. For some endpoints (relative prostate weights; absolute and relative thymus weights), it was difficult to resolve whether intra- or inter-laboratory differences contributed more to variability as the results differed across generations. There were insufficient laboratories reporting anogenital distance and weanling uterine weights to allow a comparison for these endpoints.

Observations on Specific Endpoints

Mating and Fertility Indices. Male reproductive parameters in Sprague-Dawley- and Wistar-derived rats are presented in Tables 3 and 4, respectively, and female reproductive parameters are presented in Table 5 (Sprague-Dawley) and 6 (Wistar). For this assessment, it was assumed that reproductive indices were calculated using equations similar to those outlined in Foster (1999). In both Sprague-Dawleys and Wistars, mating, fertility, and gestation indices were generally high (mean > 86%).

Table 9
Control Values: Additional P1 and P2 Organ Weights for (A) Adult Male and (B) Adult Female, Sprague-Dawley-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
(A) *Males*									
Relative adrenal wt (%)	P1	18	0.011	0.002	0.008	0.010	0.011	0.012	0.014
	P2	18	0.011	0.001	0.009	0.010	0.010	0.011	0.014
Adrenal weight (g)	P1	18	0.061	0.005	0.048	0.059	0.061	0.064	0.068
	P2	18	0.062	0.005	0.054	0.058	0.062	0.065	0.071
Relative brain wt (%)	P1	18	0.383	0.035	0.320	0.362	0.388	0.400	0.450
	P2	18	0.410	0.092	0.334	0.362	0.373	0.404	0.637
Brain weight (g)	P1	18	2.163	0.077	2.020	2.113	2.150	2.214	2.310
	P2	18	2.150	0.069	2.040	2.114	2.135	2.169	2.300
Relative pituitary wt (%)	P1	14	0.0025	0.0005	0.0020	0.0020	0.0024	0.0030	0.0030
	P2	14	0.0025	0.0005	0.0020	0.0020	0.0026	0.0030	0.0030
Pituitary weight (g)	P1	14	0.0143	0.0018	0.0110	0.0130	0.0144	0.0155	0.0176
	P2	14	0.0146	0.0021	0.0090	0.0141	0.0152	0.0155	0.0170
Relative spleen wt (%)	P1	18	0.148	0.006	0.137	0.145	0.148	0.150	0.163
	P2	18	0.150	0.006	0.139	0.146	0.151	0.155	0.161
Spleen weight (g)	P1	18	0.847	0.070	0.760	0.795	0.832	0.865	1.010
	P2	18	0.878	0.066	0.760	0.850	0.867	0.915	1.000
Relative thymus wt (%)	P1	11	0.049	0.014	0.032	0.041	0.043	0.056	0.077
	P2	11	0.051	0.015	0.027	0.042	0.048	0.056	0.083
Thymus weight (g)	P1	11	0.283	0.096	0.172	0.219	0.246	0.321	0.473
	P2	11	0.297	0.104	0.155	0.232	0.281	0.342	0.520
Relative thyroid wt (%)	P1	7	0.0047	0.0005	0.0040	0.0044	0.0048	0.0050	0.0055
	P2	7	0.0046	0.0005	0.0039	0.0043	0.0046	0.0050	0.0050
Thyroid weight (g)	P1	7	0.0264	0.0029	0.0234	0.0243	0.0250	0.0285	0.0308
	P2	7	0.0271	0.0039	0.0224	0.0245	0.0272	0.0289	0.0336
(B) *Females*									
Relative adrenal wt (%)	P1	21	0.025	0.003	0.021	0.024	0.025	0.027	0.031
	P2	22	0.025	0.003	0.022	0.023	0.025	0.027	0.034
Adrenal weight (g)	P1	21	0.082	0.008	0.069	0.078	0.080	0.086	0.103
	P2	23	0.083	0.011	0.070	0.074	0.080	0.090	0.112
Relative brain wt (%)	P1	21	0.633	0.053	0.520	0.604	0.633	0.665	0.743
	P2	22	0.589	0.104	0.329	0.574	0.613	0.637	0.766
Brain weight (g)	P1	21	2.020	0.101	1.890	1.960	1.980	2.052	2.300
	P2	23	2.018	0.104	1.880	1.940	1.977	2.091	2.220
Relative pituitary wt (%)	P1	17	0.0055	0.0008	0.0040	0.0050	0.0060	0.0060	0.0070
	P2	18	0.0053	0.0008	0.0040	0.0050	0.0051	0.0060	0.0070
Pituitary weight (g)	P1	17	0.0175	0.0025	0.0120	0.0160	0.0180	0.0188	0.0219
	P2	18	0.0171	0.0022	0.0120	0.0166	0.0171	0.0186	0.0199
Relative spleen wt (%)	P1	21	0.196	0.013	0.173	0.184	0.195	0.206	0.224
	P2	22	0.196	0.011	0.178	0.190	0.195	0.201	0.221
Spleen wt (g)	P1	21	0.628	0.053	0.567	0.590	0.610	0.660	0.750
	P2	23	0.638	0.051	0.577	0.606	0.630	0.645	0.760
Relative thymus wt (%)	P1	13	0.078	0.013	0.056	0.073	0.078	0.081	0.107
	P2	14	0.080	0.025	0.035	0.063	0.086	0.089	0.146
Thymus weight (g)	P1	13	0.254	0.036	0.210	0.223	0.252	0.282	0.330
	P2	15	0.260	0.083	0.121	0.205	0.253	0.296	0.484
Relative thyroid wt (%)	P1	8	0.0066	0.0009	0.0056	0.0060	0.0062	0.0071	0.0080
	P2	8	0.0063	0.0006	0.0057	0.0060	0.0061	0.0066	0.0073
Thyroid weight (g)	P1	8	0.0200	0.0032	0.0164	0.0174	0.0190	0.0233	0.0240
	P2	8	0.0203	0.0025	0.0166	0.0192	0.0200	0.0222	0.0238

However, reproductive capacity exhibited some variability in control animals as illustrated by the minimum values for the fertility indices.

There were some reproductive endpoints that were not consistently reported. Gestation index (Number of females with live born/Number of females pregnant $\times 100$) was the reproductive index least likely to be reported (34 of 44 laboratories reported this measure), although it is required in the Test Reporting section of OPPTS 870.3800. This reporting omission may not pose an issue as gestation index is unlikely to be a sensitive indicator of reproductive toxicity (i.e., a single live birth in a litter is sufficient to include the dam in the numerator for this calculation). Similarly, postimplantation loss (Tables 5 and 6), which serves as an indicator of in utero embryo/fetal loss, met the criterion for a highly variable

Table 10
Control Values: Additional P1 and P2 Organ Weights for (A) Adult Male and (B) Adult Female Wistar-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
(A) *Males*									
Relative adrenal wt (%)	P1	17	0.015	0.002	0.012	0.014	0.015	0.016	0.018
	P2	17	0.015	0.001	0.013	0.015	0.016	0.016	0.018
Adrenal weight (g)	P1	17	0.066	0.010	0.055	0.058	0.063	0.074	0.083
	P2	17	0.069	0.008	0.058	0.062	0.064	0.077	0.084
Relative brain wt (%)	P1	17	0.474	0.051	0.380	0.446	0.477	0.512	0.563
	P2	17	0.477	0.053	0.370	0.438	0.486	0.521	0.568
Brain weight (g)	P1	17	2.069	0.054	1.942	2.044	2.093	2.110	2.137
	P2	17	2.086	0.047	1.975	2.074	2.101	2.123	2.136
Relative pituitary wt (%)	P1	16	0.0027	0.0004	0.0020	0.0023	0.0030	0.0030	0.0030
	P2	16	0.0025	0.0005	0.0020	0.0020	0.0022	0.0030	0.0030
Pituitary weight (g)	P1	16	0.0112	0.0017	0.0090	0.0100	0.0104	0.0124	0.0150
	P2	16	0.0108	0.0016	0.0090	0.0099	0.0102	0.0114	0.0150
Relative spleen wt (%)	P1	17	0.172	0.018	0.149	0.161	0.165	0.179	0.216
	P2	17	0.172	0.014	0.153	0.160	0.165	0.183	0.203
Spleen weight (g)	P1	17	0.769	0.147	0.580	0.660	0.714	0.883	1.123
	P2	17	0.772	0.156	0.609	0.650	0.699	0.889	1.165
Relative thymus wt (%)	P1	8	0.073	0.030	0.044	0.050	0.058	0.101	0.116
	P2	8	0.081	0.034	0.052	0.056	0.060	0.119	0.126
Thymus weight (g)	P1	8	0.335	0.132	0.205	0.249	0.260	0.440	0.550
	P2	8	0.367	0.128	0.246	0.276	0.295	0.505	0.546
Relative thyroid wt (%)	P1	6	0.0057	0.0005	0.0050	0.0053	0.0060	0.0060	0.0060
	P2	6	0.0057	0.0008	0.0050	0.0050	0.0055	0.0060	0.0070
Thyroid weight (g)	P1	5	0.0225	0.0012	0.0206	0.0226	0.0226	0.0232	0.0237
	P2	6	0.0230	0.0032	0.0193	0.0206	0.0229	0.0245	0.0280
(B) *Females*									
Relative adrenal wt (%)	P1	17	0.034	0.003	0.028	0.032	0.034	0.037	0.039
	P2	17	0.035	0.003	0.029	0.033	0.035	0.036	0.039
Adrenal weight (g)	P1	17	0.090	0.015	0.072	0.076	0.089	0.103	0.111
	P2	17	0.091	0.014	0.072	0.079	0.092	0.103	0.114
Relative brain wt (%)	P1	17	0.747	0.080	0.600	0.695	0.723	0.806	0.900
	P2	17	0.755	0.079	0.610	0.691	0.744	0.821	0.879
Brain weight (g)	P1	17	1.925	0.029	1.884	1.903	1.921	1.939	1.978
	P2	17	1.936	0.041	1.818	1.920	1.930	1.965	2.001
Relative pituitary wt (%)	P1	16	0.0054	0.0007	0.0040	0.0050	0.0055	0.0060	0.0060
	P2	16	0.0050	0.0005	0.0040	0.0050	0.0050	0.0050	0.0060
Pituitary weight (g)	P1	16	0.0138	0.0017	0.0110	0.0129	0.0136	0.0146	0.0176
	P2	15	0.0130	0.0013	0.0108	0.0120	0.0129	0.0143	0.0149
Relative spleen wt (%)	P1	17	0.217	0.016	0.191	0.206	0.213	0.226	0.248
	P2	17	0.222	0.013	0.209	0.211	0.219	0.230	0.259
Spleen wt (g)	P1	17	0.571	0.093	0.446	0.496	0.580	0.608	0.798
	P2	17	0.580	0.092	0.460	0.524	0.550	0.630	0.823
Relative thymus wt (%)	P1	8	0.093	0.008	0.082	0.088	0.095	0.097	0.104
	P2	8	0.098	0.010	0.083	0.094	0.099	0.103	0.111
Thymus weight (g)	P1	8	0.257	0.018	0.230	0.248	0.257	0.266	0.286
	P2	8	0.267	0.020	0.235	0.252	0.271	0.284	0.288
Relative thyroid wt (%)	P1	6	0.0070	0.0006	0.0060	0.0070	0.0070	0.0070	0.0080
	P2	6	0.0080	0.0011	0.0060	0.0080	0.0080	0.0088	0.0090
Thyroid weight (g)	P1	6	0.0169	0.0015	0.0141	0.0167	0.0173	0.0178	0.0183
	P2	6	0.0187	0.0012	0.0169	0.0181	0.0186	0.0198	0.0200

endpoint. This may be related to the relatively small n value ($n = 7$ or 8) for the numbers of laboratories reporting postimplantation loss.

Time to mating, precoital interval (Tables 5 and 6), was included in nearly all reports with Wistar rats (14 of 16), whereas it was reported less frequently in Sprague-Dawley studies (18 of 27). It appears that the time to mating was slightly less in Wistar than Sprague-Dawley rats (mean times = 2.4–2.7 days in Wistar rats vs. 2.9–3.1 days in Sprague-Dawley rats). A related variable "number of estrous periods until mating" was seldom reported although it is required per the test guideline. This variable can be estimated from mean estrous cycle length and days to mating.

Age at Puberty Onset. Age at puberty onset (as well as mating and fertility indices and sperm parameters) is presented in Tables 3–6. Age at puberty onset in the P2 generation was reported in all studies and varied across a series of days in control animals from both strains. Puberty onset is an apical endpoint that

Table 11
Control Values: F1 and F2 Litter Sizes for Sprague-Dawley-Derived Rats

Endpoint	Gen.		*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Litter size, PND 0	P1	All	23	13.74	0.94	12.30	13.00	13.60	14.35	15.90
		Culled	18	13.61	1.01	12.30	12.93	13.45	14.35	15.90
		Not culled	5	14.22	0.43	13.60	14.20	14.20	14.30	14.80
	P2	All	23	13.77	0.96	12.50	13.00	13.60	14.20	16.40
		Culled	18	13.74	1.06	12.50	13.00	13.50	14.05	16.40
		Not culled	5	13.86	0.59	13.00	13.60	13.90	14.30	14.50
Litter size, PND 4	P1	All	12	13.27	1.22	11.20	12.60	13.05	13.98	15.70
		Culled	7	12.80	1.41	11.20	12.15	12.70	12.85	15.70
		Not culled	5	13.92	0.49	13.20	13.80	13.90	14.20	14.50
	P2	All	12	13.37	1.03	11.90	12.80	13.20	13.85	15.90
		Culled	7	13.29	1.28	11.90	12.80	12.90	13.35	15.90
		Not culled	5	13.48	0.65	12.40	13.50	13.50	14.00	14.00
Litter size, PND 7	P1	Culled	7	7.70	0.21	7.40	7.55	7.80	7.80	8.00
		Not culled	5	13.80	0.41	13.10	13.80	13.90	14.10	14.10
	P2	Culled	7	7.74	0.14	7.50	7.70	7.80	7.80	7.90
		Not culled	5	13.40	0.62	12.40	13.30	13.50	13.80	14.00
Litter size, PND 14	P1	Culled	7	7.64	0.29	7.20	7.45	7.80	7.80	8.00
		Not culled	5	13.78	0.40	13.10	13.80	13.90	14.00	14.10
	P2	Culled	7	7.73	0.16	7.50	7.65	7.80	7.80	7.90
		Not culled	5	13.36	0.60	12.40	13.30	13.40	13.70	14.00
Litter size, PND 21	P1	Culled	7	7.64	0.29	7.20	7.45	7.80	7.80	8.00
		Not culled	5	13.76	0.38	13.10	13.80	13.90	13.90	14.10
	P2	Culled	7	7.73	0.16	7.50	7.65	7.80	7.80	7.90
		Not culled	5	13.24	0.68	12.10	13.30	13.30	13.60	13.90

Table 12
Control Values: F1 and F2 Litter Sizes for Wistar-Derived Rats

Endpoint	Gen.	Studies	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Litter size, PND 0	P1	All	17	11.84	1.88	9.10	10.60	10.90	13.50	15.20
		Culled	15	11.67	1.91	9.10	10.45	10.80	13.30	15.20
	P2	All	17	11.48	1.67	9.90	10.20	10.40	13.00	14.70
		Culled	15	11.45	1.69	9.90	10.25	10.40	12.90	14.70
Litter size, PND 4	P1	All	14	11.36	2.03	7.80	10.10	10.65	12.95	14.50
		Culled	12	11.30	2.15	7.80	10.03	10.60	13.08	14.50
	P2	All	14	11.18	1.63	9.00	9.93	10.55	12.70	14.30
		Culled	12	11.15	1.66	9.00	9.98	10.55	12.50	14.30
Litter size, PND 7	P1	Culled	12	7.74	0.20	7.30	7.68	7.80	7.90	8.00
	P2	Culled	12	7.66	0.25	7.20	7.48	7.65	7.83	8.00
Litter size, PND 14	P1	Culled	12	7.70	0.17	7.30	7.68	7.75	7.80	7.90
	P2	Culled	12	7.62	0.30	7.10	7.40	7.65	7.83	8.00
Litter size, PND 21	P1	Culled	12	7.68	0.16	7.30	7.60	7.70	7.80	7.90
	P2	Culled	12	7.60	0.30	7.10	7.38	7.65	7.83	8.00

exhibits some inherent variability and can be influenced by a number of factors (for review, see Goldman et al., 2000; Stoker et al., 2000). Across submitted studies, the mean ages at preputial separation (PPS; Tables 3 and 4) ranged from 41.2–49.0 days in Sprague-Dawley rats and 40.9–45.0 days in Wistar rats, but overall CVs were low (3–5%). Mean age at preputial separation was ~2 days earlier in Wistar rats than Sprague-Dawley rats. Additional variability may occur across laboratories, depending on how the age at preputial separation is classified (i.e., as the age at which preputial separation has clearly started or the age at which preputial separation is completed). Similarly, mean ages at vaginal opening/vaginal patency (VO/VP; Tables 5 and 6) varied from 29.0–38.8 in Sprague-Dawley rats and 29.9–35.1 in Wistar rats. The CV was 6% in both strains. The mean age at vaginal opening was one day earlier in Wistar rats than Sprague-Dawley rats. One factor that may have contributed to variability in determining the age of puberty onset is the appearance of vaginal or preputial threads. Based on the available data, it is difficult to determine the incidence of these threads and whether they contributed to later ages for puberty onset. Data from animals with vaginal or preputial threads are likely to be handled differently in different laboratories. In the current inter-laboratory dataset, body weight measurements were often not reported on the day at which preputial separation or vaginal opening was achieved.

Table 13
Control Values: F1 and F2 Pup Body Weights (g) for Sprague-Dawley-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Body weight, PND 1	F1	17	7.1	0.28	6.6	6.9	7.1	7.2	7.8
	F2	17	7.0	0.24	6.6	6.9	7.0	7.2	7.5
Body weight, PND 4	F1	20	10.2	0.69	9.1	9.7	10.2	10.5	11.6
	F2	20	10.2	0.70	8.9	9.8	10.2	10.4	11.7
Body weight, PND 21[a]	F1	13	54.3	4.8	45.4	50.6	54.6	58.5	60.5
	F2	13	55.1	4.4	50.0	51.3	54.2	56.7	63.8
Anogenital distance[b]	F1	1					4.34		
	F2	9	3.83	1.07	2.56	2.80	4.32	4.41	5.30
Females									
Body weight, PND 1	F1	17	6.7	0.28	6.2	6.5	6.7	6.8	7.3
	F2	17	6.6	0.24	6.2	6.5	6.6	6.8	7.0
Body weight, PND 4	F1	20	9.7	0.70	8.5	9.3	9.6	9.9	11.1
	F2	20	9.7	0.69	8.4	9.2	9.7	9.9	11.2
Body weight, PND 21[a]	F1	13	52.0	4.1	44.1	49.2	52.1	55.6	57.6
	F2	13	52.4	4.1	47.1	49.2	51.6	53.5	60.4
Anogenital distance (mm)[b]	F1	1					2.41		
	F2	9	2.08	0.88	1.04	1.20	2.28	2.57	3.40

[a]Excludes 12 studies where litters were not culled (5) or for which the route of exposure was inhalation (7).
[b]Anogenital distance was recorded on PND 1; limited data for F2 PND 0 ($n = 1$) and PND 4 ($n = 1$) are not presented.

Table 14
Control Values: F1 and F2 Pup Body Weights (g) for Wistar-Derived Rats

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Body weight, PND 1	F1	17	6.3	0.26	5.8	6.1	6.4	6.5	6.8
	F2	17	6.4	0.21	5.9	6.2	6.4	6.5	6.7
Body weight, PND 4	F1	17	9.4	0.44	8.7	9.1	9.4	9.7	10.2
	F2	17	9.6	0.40	9.0	9.4	9.5	9.7	10.7
Body weight, PND 21[a]	F1	15	48.1	3.20	42.0	46.0	47.5	50.9	52.8
	F2	15	49.3	2.95	44.4	47.2	49.7	52.0	54.6
Females									
Body weight, PND 1	F1	17	6.0	0.26	5.4	5.8	6.1	6.2	6.4
	F2	17	6.0	0.18	5.7	5.9	6.1	6.1	6.3
Body weight, PND 4	F1	17	9.0	0.42	8.3	8.8	8.9	9.2	9.8
	F2	17	9.2	0.40	8.6	9.0	9.2	9.3	10.3
Body weight, PND 21[a]	F1	15	46.7	3.04	40.9	44.2	46.6	48.6	51.4
	F2	15	47.4	2.70	42.6	45.4	47.7	49.8	51.6

[a]Excludes 2 studies that did not cull to 4 pups/sex/litter on PND 4.

Thus, it is difficult to determine how variable body weights were across these datasets and how body weight influenced the age ranges for preputial separation and vaginal patency. Furthermore, there is evidence that the amount of calories consumed in the diet will impact puberty onset in female rats (Odum et al., 2004). There are likely to be variations in the rodent diets (their energy content as well as phytoestrogen levels) and even bedding across laboratories, which may affect puberty onset.

Sperm Parameters. In the 1998 guideline revision to OPPTS 870.3800, sperm parameters were added as endpoints to the reproduction and fertility effects study. The addition of multiple sperm parameters was based on the premise that toxicants with different modes of action may affect sperm endpoints differentially, making different endpoints more sensitive to different types of toxicants. Inter-laboratory control values for testicular spermatid and epididymal sperm concentrations, sperm motility, and percent abnormal sperm are reported in Tables 3 (Sprague-Dawley) and 4 (Wistar). Interestingly, epididymal and testicular sperm/spermatid concentrations and percent abnormal sperm were identified as ''high variability'' endpoints in Sprague-Dawley rats. This variability could arise from multiple sources including different procedures across laboratories and inter-individual variability. Within the control population, there is a background incidence of oligospermic males, which contributes to this variability.

Sperm Motility: Variability in sperm motility may result from numerous procedural differences between laboratories. Sperm from the cauda epididymis or the vas deferens are typically used and the method used to release sperm may affect motility measurements (Chapin and Conner, 1999). Differences in media used for sperm motility assessments may affect these values across

Table 15
Control Values: F1 and F2 Pup Organ Weights for Sprague-Dawley-Derived Rats[a]

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Relative brain wt (%)	F1	14	2.707	0.317	1.860	2.557	2.754	2.937	3.084
	F2	13	2.661	0.290	1.830	2.626	2.683	2.803	3.088
Brain weight (g)	F1	14	1.481	0.049	1.364	1.469	1.500	1.503	1.553
	F2	13	1.500	0.042	1.445	1.474	1.490	1.520	1.593
Relative spleen wt (%)	F1	14	0.437	0.031	0.364	0.419	0.434	0.463	0.478
	F2	13	0.428	0.031	0.363	0.407	0.431	0.440	0.471
Spleen weight (g)	F1	14	0.245	0.032	0.197	0.222	0.246	0.269	0.305
	F2	13	0.246	0.028	0.216	0.223	0.239	0.256	0.317
Relative thymus wt (%)	F1	14	0.425	0.039	0.371	0.402	0.410	0.439	0.508
	F2	13	0.432	0.037	0.384	0.407	0.421	0.448	0.500
Thymus weight (g)	F1	14	0.239	0.038	0.183	0.212	0.234	0.257	0.312
	F2	13	0.248	0.037	0.206	0.233	0.245	0.251	0.351
Females									
Relative brain wt (%)	F1	13	2.756	0.299	1.970	2.643	2.809	2.911	3.153
	F2	13	2.700	0.298	1.890	2.660	2.741	2.846	3.173
Brain weight (g)	F1	14	1.435	0.041	1.339	1.422	1.437	1.460	1.491
	F2	13	1.447	0.037	1.389	1.420	1.438	1.465	1.527
Relative spleen wt (%)	F1	13	0.448	0.046	0.361	0.421	0.443	0.486	0.531
	F2	13	0.446	0.035	0.358	0.432	0.448	0.475	0.492
Spleen wt (g)	F1	14	0.239	0.029	0.190	0.225	0.242	0.256	0.294
	F2	13	0.244	0.024	0.216	0.226	0.240	0.261	0.292
Relative thymus wt (%)	F1	13	0.458	0.034	0.410	0.435	0.444	0.479	0.522
	F2	13	0.460	0.040	0.414	0.425	0.456	0.464	0.548
Thymus weight (g)	F1	14	0.244	0.031	0.188	0.228	0.245	0.252	0.311
	F2	13	0.251	0.034	0.208	0.234	0.247	0.257	0.337
Relative uterine wt (%)	F1	3	0.098	0.016	0.087		0.091		0.116
	F2	3	0.072	0.016	0.062		0.065		0.091
Uterine weight (g)	F1	3	0.061	0.024	0.047		0.048		0.088
	F2	3	0.046	0.024	0.031		0.035		0.074

[a]Excludes 12 studies where litters were not culled (5) or for which the route of exposure was inhalation (7).

laboratories as some media may better support sperm motility. Most laboratories use a computer-assisted sperm analyzer (CASA, e.g., Hamilton-Thorne) to assess motility, but parameter settings used to detect motile, non-motile, and progressively motile sperm may vary from laboratory to laboratory. The impact of different variables on motility measurements (e.g., CASA parameters, temperature, pH, chamber size, sample density, interval between necropsy, and motility assessment, etc.) have been well described by Slott and Perreault (1993) and Seed et al. (1996). Slott and Perreault (1993) recommend that 200 sperm/sample be monitored during motility assessments; however, the number of sperm to evaluate is not specified in the test guidelines. As with all methods, it is critical for laboratories to conduct sperm motility assessments in a consistent manner within and between studies.

With so many variables, comparison of motility data across laboratories is difficult. Still, a minimum motility value of 70% has been suggested as an acceptable level in control rats (Seed et al., 1996; Chapin et al., 1992). While this value has not been rigorously tested, only 3 of the 79 mean control values for sperm motility (P1 and P2 values for both Sprague-Dawley and Wistar rats) were less than 70%. According to Chapin and Conner (1999), control values of 85–96% are commonly reported in rat studies, which is consistent with the median to maximum values in Sprague-Dawley rats and the 25% to maximum values in Wistar rats. Overall, sperm motility data had less variability among laboratories than sperm morphology and sperm/spermatid counts.

Interestingly, the guidelines require that the ''number and percent of progressively motile sperm'' be reported in each multi-generation study; however, few laboratories reported progressively motile sperm in their multi-generation study data. The rationale for not reporting progressively motile sperm may be related to the lack of a scientifically agreed-upon definition as to what constitutes ''progressively motile.'' When using CASA, progressive motility depends on user-defined parameters for average path velocity and straightness of linear index, values that likely differ from laboratory to laboratory in those that choose to collect and report progressive motility.

Epididymal sperm concentrations: There was remarkable consistency in the practice of presenting sperm count data as ''counts per g tissue'' in the multi-generation study reports. Still, epididymal sperm concentration was considered a ''highly variable'' endpoint with CVs greater than 50% in Sprague-Dawley rats. In this strain, mean epididymal sperm concentrations across studies ranged from 6.4 million to 1,144 million sperm/g cauda epididymis (P1) and from 8.2 million to 1,189.4 million sperm/g cauda epididymis (P2). In Wistar rats, epididymal sperm concentration variability was much lower, ranging from 451 million to 697 million and

Table 16
Control Values: F1 and F2 Pup Organ Weights for Wistar-Derived Rats[a]

Endpoint	Gen.	*n*	Mean	SD	Min	25%	Median (50%)	75%	Max
Males									
Relative brain wt (%)	F1	15	3.071	0.223	2.697	2.904	3.139	3.205	3.359
	F2	15	3.051	0.199	2.710	2.887	3.049	3.200	3.419
Brain weight (g)	F1	15	1.481	0.029	1.412	1.460	1.491	1.499	1.528
	F2	15	1.494	0.020	1.458	1.486	1.495	1.500	1.540
Relative spleen wt (%)	F1	15	0.453	0.026	0.416	0.434	0.443	0.470	0.503
	F2	15	0.466	0.023	0.423	0.450	0.472	0.484	0.498
Spleen weight (g)	F1	15	0.223	0.022	0.183	0.210	0.226	0.236	0.259
	F2	15	0.232	0.013	0.214	0.222	0.235	0.240	0.260
Relative thymus wt (%)	F1	15	0.415	0.053	0.319	0.364	0.434	0.443	0.485
	F2	15	0.421	0.057	0.332	0.354	0.452	0.465	0.477
Thymus weight (g)	F1	15	0.201	0.018	0.170	0.193	0.206	0.211	0.240
	F2	15	0.208	0.021	0.174	0.191	0.216	0.227	0.229
Females									
Relative brain wt (%)	F1	15	3.093	0.197	2.777	2.921	3.152	3.229	3.340
	F2	14	3.075	0.184	2.838	2.915	3.088	3.208	3.411
Brain weight (g)	F1	15	1.435	0.026	1.377	1.415	1.443	1.454	1.465
	F2	15	1.445	0.018	1.415	1.429	1.448	1.463	1.467
Relative spleen wt (%)	F1	15	0.463	0.020	0.437	0.447	0.461	0.474	0.503
	F2	14	0.474	0.028	0.429	0.457	0.471	0.497	0.520
Spleen wt (g)	F1	15	0.218	0.016	0.193	0.207	0.220	0.232	0.241
	F2	15	0.226	0.011	0.205	0.218	0.228	0.233	0.243
Relative thymus wt (%)	F1	15	0.440	0.056	0.342	0.386	0.461	0.476	0.520
	F2	14	0.447	0.069	0.349	0.368	0.483	0.492	0.522
Thymus weight (g)	F1	15	0.205	0.017	0.174	0.197	0.208	0.218	0.228
	F2	15	0.213	0.026	0.173	0.188	0.222	0.235	0.240
Relative uterine wt (%)	F1	1					0.142		
	F2	1					0.112		
Uterine weight (g)	F1	1					0.068		
	F2	1					0.053		

[a]Excludes 2 studies that did not cull to 4 pups/sex/litter on PND 4.

from 444 million to 705 million sperm/g cauda epididymis in the P1 and P2 generations, respectively. This may be due to fewer laboratories reporting data in Wistar rats, which supports the ANOVA analysis results (i.e., that inter-laboratory variability makes a greater contribution to differences in sperm concentrations than inherent variability). Aside from methodological differences, other factors also may influence sperm concentrations, such as the age of the rats at the time the sample was collected, and the timing of the necropsy and sperm sampling relative to the last (successful) mating day (because epididymal sperm may be depleted due to recent ejaculatory activity). In these multi-generation studies, necropsy of the males may occur at different time points after the required 12-week dosing period, particularly if a second breeding is required. Furthermore, strain differences in sperm counts have been reported (Chapin and Conner, 1999) and some variability in sperm counts has been documented (Morrissey et al., 1988).

Despite this variability, sperm concentration may offer value in assessing male reproductive toxicity. Sperm count has been definitively linked to fertility in all species studied. Chapin et al. (1997) showed that a 10% reduction in sperm counts was reflected as a change in mouse fertility across studies in the National Toxicology Program Continuous Breeding database. Power calculations estimate a 90% chance of detecting a 15% change in epididymal sperm count when $n = 20$ per group (Chapin and Conner, 1999) and a 31–32% change when $n = 15$ in rats (Blazak et al., 1985). Histological changes in the testis may correlate with altered epididymal sperm counts, but evidence exists to indicate that these parameters may be altered independent of one another (Chapin and Conner, 1999).

Testicular spermatid concentrations: As with the epididymal sperm concentration, the testicular homogenization-resistant spermatid head concentration was considered a ''highly variable'' endpoint with CVs greater than 125% in Sprague-Dawley rats. Spermatid concentrations (spermatids/g testis) ranged from 72.9 to 826.5 in the P1 generation and from 76.2 to 1,063.9 in the P2 generation in this strain. Similarly, Wistar rats had less variability in spermatid concentrations than Sprague-Dawley rats. Both inherent biological variability and methodological differences appear to contribute to the high degree of variability in this endpoint; however, inherent biological and/or intra-laboratory procedural factors contribute most to variability (Table 2).

Mean spermatid counts/g testicular tissue (mean = $142.2-159.5 \times 10^6$ for Sprague-Dawley-derived rats; $106.2-109.2 \times 10^6$ for Wistar-derived rats) were similar to previously reported spermatid concentrations in both strains (e.g., Tyl et al., 2004; Willoughby et al., 2000; Schneider et al., 2005; Suter et al., 1998). Despite this variability, spermatid counts generally are considered to

Table 17
Control Values From the Scientific Literature: Individual Studies Reporting Reproductive Parameters in Sprague-Dawley and Wistar Rats

		Tyl et al. (2004)			Willoughby et al. (2000)			Willoughby et al. (2000)			Schneider et al. (2005)		
Reference		Sprague-Dawley			Sprague-Dawley			Sprague-Dawley			Wistar		
Strain													
Endpoint	Gen.	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Fertility index (%) (males)	P1	30	90		28	100		27	92.6		25	96	
	P2	30	100		28	96.4		28	100		25	96	
Fertility index (%) (females)	P1	30	90		28	100		28	89.3		25	100	
	P2	30	100		28	96.4		28	100		25	100	
Mating index (%) (males)	P1	30	100		28	100		27	96.3		24	100	
	P2	30	96.7		28	96.4		28	100		24	100	
Mating index (%) (females)	P1	30	100		28	100		28	96.4		25	31.3	1.6
	P2	30	96.7		28	96.4		28	100			32.9	2.7
Gestation index (%)	P1	27	96.3		28	100					25	90.8	10.6
	P2	29	100		27	96.3						93.5	10.9
Age at puberty onset (females)	P1										25	43.6	1.2
	P2	30	31.4	0.3	28	35.1	1.7	28	34.7	2.0		42.8	1.5
Body wt at puberty onset (females)	P1										25	172.5	10.3
	P2	30	108.25	2.50	28	118	16	28	128	16		165.5	8.9
Age at puberty onset (males)	P1										25	677	133.7
	P2	30	40.9	0.4	28	45.0	1.9	28	44.8	2.3	25	763	108.9
Body wt at puberty onset (males)	P1										25	125	16.2
	P2	30	208.17	2.28	28	228	27	28	235	24	25	119	11.5
Conc. epid. sperm (10^6 sperm/g)	P1	30	946.71	20.94							25	88	11.0
	P2	30	825.59	38.16							25	89	7.7
Conc. test. sperm. (10^6 sperm./g testis)	P1	30	104.23	7.22	28	141	62	27	108	33	25	1.4	0.9
	P2	30	47.22	6.12	28	127	46	28	122	43	25	1.3	1.1
Sperm motility (%)	P1	30	77.6	1.0	28	77	8	27	76	8	25	96	
	P2	30	68.6	4.0	28	73	10	28	77	8	25	96	
Abnormal sperm (%)	P1	30	2.12	0.21							25	100	
	P2	30	5.99	3.18							25	100	
Normal sperm (%)	P1				28	96	3	27	96	3			
	P2				28	95	6	28	91	9			
Ovarian follicle count (mean)	P1	30	281.1	34.1									
	P2	30	368.4	26.3							25	224	
Estrous cycle length (d)	P1	30	4.39	0.20							25	4.1	
	P2	30	4.38	0.10							25	4.1	
Time to mating (d)	P1	30	3.1	0.6							24	2.6	
	P2	30	2.8	0.2							24	3.0	
Gestational length (d)	P1	26	22.2	0.1	28	22.6	0.4	26	22.8	0.5	24	22.1	
	P2	29	22.0	0.1	26	22.6	0.4	28	22.6	0.4	24	22.0	
Postimplantation loss (%)	P1	27	15.79	4.52							24	8.1	
	P2	29	10.02	1.21							24	8.8	
Litter size, day 0 or 1	P1	26	13.3	0.9	28	14.4	3.2	25	14.3	3.2	24	11.0	
	P2	29	14.2	0.3	26	13.0	3.6	28	14.4	2.8	24	11.4	
Sex ratio (% males)	P1				28	52		25	48		24	46	
	P2				26	42		28	47		24	47	
Anogenital distance, male PND 0 (mm)	P1	26	2.06	0.03									
	P2	29	2.05	0.01									
Body wt/litter, male PND 0 (g)	P1	26	6.76	0.10									
	P2	29	6.63	0.11									
Anogenital distance, female PND 0 (mm)	P1	26	0.96	0.02									
	P2	29	0.98	0.01									
Body wt/litter, female PND 0 (g)	P1	26	6.35	0.10									
	P2	29	6.21	0.10									
Male pup body wt, day 1 (g)	P1				28	6.1	0.9	25	6.1	0.8			
	P2				26	6.5	1.0	28	6.2	0.7			
Male pup body wt, day 21 (g)	P1				28	51.8	6.6	25	53.4	7.5			
	P2				26	53.7	9.7	28	56.0	5.6			
Female pup body wt, day 1 (g)	P1				28	5.8	0.9	25	5.7	0.7			
	P2				26	5.9	0.8	28	5.8	0.6			
Female pup body wt, day 21 (g)	P1				28	49.6	7.1	25	51.9	7.5			
	P2				26	51.5	10.1	28	53.3	5.8			

Table 17
Continued

Reference		Tyl et al. (2004)			Willoughby et al. (2000)			Willoughby et al. (2000)			Schneider et al. (2005)		
Strain		Sprague-Dawley			Sprague-Dawley			Sprague-Dawley			Wistar		
Endpoint	Gen.	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Male adult organ weights													
Terminal body weights (g)	P1	30	599.25	9.46	28	600	81	27	639	71			
	P2	30	601.30	9.12	28	583	76	28	611	72			
Relative adrenal wt (paired) (%)[a]	P1	30	0.0103	0.0003									
	P2	30	0.0097	0.0003									
Adrenal wt (paired) (g)	P1	30	0.0615	0.0017	28	0.061	0.010	27	0.054	0.013			
	P2	30	0.0583	0.0018	28	0.055	0.009	28	0.061	0.012			
Relative brain wt (%)	P1	30	0.365	0.005									
	P2	30	0.369	0.006									
Brain wt. (g)	P1	30	2.177	0.014	28	2.10	0.11	27	2.15	0.09			
	P2	30	2.205	0.016	28	1.99	0.14	28	2.18	0.09			
Relative epididymal wt (paired) (%)	P1	30	0.238	0.004	28	0.22	0.03	27	0.21	0.02			
	P2	30	0.226	0.006	28	0.22	0.03	28	0.22	0.03			
Epididymal wt (paired) (g)	P1	30	1.424	0.019	28	1.30	0.10	27	1.34	0.12			
	P2	30	1.351	0.028	28	1.24	0.10	28	1.32	0.10			
Relative pituitary wt (%)	P1	30	0.0028	0.0000									
	P2	30	0.0028	0.0001									
Pituitary wt (g)	P1	30	0.0169	0.0003									
	P2	30	0.0168	0.0003									
Relative prostate wt (%)	P1	30	0.142	0.006	28	0.10	0.03	27	0.10	0.03			
	P2	30	0.126	0.006	28	0.10	0.02	28	0.09	0.03			
Prostate wt (g)	P1	30	0.846	0.034	28	0.615	0.140	27	0.653	0.169			
	P2	30	0.756	0.037	28	0.554	0.130	28	0.600	0.198			
Relative seminal vesicle wt (%)	P1	30	0.396	0.010	28	0.43	0.08	27	0.39	0.05			
	P2	30	0.358	0.009	28	0.42	0.07	28	0.39	0.07			
Seminal vesicle wt (g)	P1	30	2.362	0.059	28	2.55	0.32	27	2.49	0.29			
	P2	30	2.145	0.052	28	2.39	0.32	28	2.35	0.30			
Spleen wt (g)	P1				28	0.879	0.117	27	0.864	0.132			
	P2				28	0.885	0.142	28	0.864	0.168			
Relative testes wt (paired) (%)	P1	30	0.578	0.010	28	0.63	0.09	27	0.60	0.07			
	P2	30	0.600	0.010	28	0.63	0.08	28	0.63	0.08			
Testes wt (paired) (g)	P1	30	3.447	0.058	28	3.72	0.26	27	3.84	0.42			
	P2	30	3.598	0.050	28	3.64	0.34	28	3.78	0.32			
Thymus wt (g)	P1				28	0.335	0.081	27	0.328	0.065			
	P2				28	0.390	0.101	28	0.364	0.096			
Female adult organ weights													
Terminal body wt (g)	P1	29	342.21	3.76	27	354	30	24	351	28			
	P2	30	338.70	5.00	24	340	32	26	341	31			
Adrenal wt (paired) (g)	P1				27	0.089	0.014	24	0.088	0.017			
	P2				24	0.087	0.014	26	0.087	0.011			
Brain wt (g)	P1				27	1.94	0.09	24	1.95	0.20			
	P2				24	1.96	0.09	26	2.00	0.06			
Relative ovary wt (paired) (%)	P1	29	0.0495	0.0002	27	0.0341	0.0062	24	0.0315	0.0067			
	P2	30	0.043	0.001	24	0.0331	0.0044	26	0.0319	0.0041			
Ovary wt (paired) (g)	P1	29	0.153	0.007	27	0.120	0.021	24	0.110	0.022			
	P2	30	0.146	0.004	24	0.112	0.016	26	0.110	0.015			
Spleen wt (g)	P1				27	0.644	0.076	24	0.626	0.099			
	P2				24	0.690	0.079	26	0.663	0.092			
Thymus wt (g)	P1				27	0.215	0.066	24	0.228	0.059			
	P2				24	0.301	0.085	26	0.250	0.072			
Relative uterine wt (%)	P1	29	0.169	0.007	27	0.13	0.02	24	0.15	0.03			
	P2	30	0.169	0.006	24	0.15	0.04	26	0.15	0.04			
Uterine wt (g)	P1	29	0.575	0.021	27	0.46	0.08	24	0.53	0.11			
	P2	30	0.570	0.021	24	0.49	0.01	26	0.52	0.13			
Male weanling organ weights													
Terminal body wt (g)	P1	68	49.45	1.24									
	P2	86	51.78	0.90									

Table 17
Continued

Reference Strain Endpoint	Gen.	Tyl et al. (2004) Sprague-Dawley *n*	Mean	SD	Willoughby et al. (2000) Sprague-Dawley *n*	Mean	SD	Willoughby et al. (2000) Sprague-Dawley *n*	Mean	SD	Schneider et al. (2005) Wistar *n*	Mean	SD
Relative brain wt (%)	P1	68	2.987	0.061									
	P2	86	2.896	0.040									
Brain wt (g)	P1	68	1.450	0.010									
	P2	86	1.485	0.012									
Relative spleen wt (%)	P1	68	0.412	0.011									
	P2	86	0.406	0.009									
Spleen wt (g)	P1	68	0.205	0.009									
	P2	86	0.211	0.007									
Relative thymus wt (%)	P1	68	0.435	0.003									
	P2	86	0.454	0.009									
Thymus wt (g)	P1	68	0.215	0.007									
	P2	86	0.236	0.008									
Female weanling organ weights													
Terminal body wt (g)	P1	73	47.86	1.10									
	P2	87	48.45	0.87									
Relative brain wt (%)	P1	73	3.007	0.056									
	P2	87	2.978	0.044									
Brain wt (g)	P1	73	1.419	0.012									
	P2	87	1.430	0.013									
Relative spleen wt (%)	P1	73	0.414	0.011									
	P2	87	0.404	0.008									
Spleen wt (g)	P1	73	0.200	0.009									
	P2	87	0.197	0.006									
Relative thymus wt (%)	P1	73	0.485	0.013									
	P2	87	0.493	0.010									
Thymus wt (g)	P1	73	0.232	0.008									
	P2	87	0.239	0.008									
Relative uterine wt (%)	P1	73	0.224	0.010									
	P2	87	0.164	0.007									
Uterine wt (g)	P1	73	0.108	0.006									
	P2	87	0.080	0.004									

[a]Organ weights relative to the terminal body weight in percent.

be a more sensitive indicator of male reproductive toxicity than fertility because of excess sperm production in rats (Meistrich, 1989). Decreased sperm production may be associated with decreased fertility, decreased testicular weight (Blazak et al., 1993), or testicular histopathologic changes, but this is not always the case.

Sperm Morphology: Sperm morphology involves the assessment of sperm for changes in shape of the tail, head, or midpiece. It was considered a ''high variability'' endpoint as mean CVs were approximately 185% in Sprague-Dawley rats and approximately 38% in Wistar rats. When examining the distribution of sperm morphology values, it appears that some atypically high values (e.g., maximum percent abnormal sperm values of 15.9 and 30% in P1 and P2 Sprague-Dawley rats, respectively) skewed the mean for percent abnormal sperm. These high values may have resulted from methodological differences across laboratories (Table 2). For example, slide preparation can result in artifactual increases in the percent abnormal sperm (Chapin and Conner, 1999). Furthermore, laboratories are expected to define the parameters for data collection such that a reasonably sensitive assessment is conducted. In the case of sperm morphology, the level of magnification is not specified, but the minimum number of sperm examined/male is defined as greater than or equal to 200, with 500 sperm/male recommended, if possible.

In rodents, background levels of abnormal sperm in control mice and rats typically range from 0.3–6% (Chapin and Conner, 1999). Thus, >75% of the Sprague-Dawley values and all of the Wistar values are within this previously reported range. Morphology is considered a sensitive endpoint with which to detect treatment-related effects because the probability of detecting a doubling of abnormal sperm forms (e.g., 3 to 6%) is 90% when $n = 20$ samples/group (Chapin and Conner, 1999). Chapin et al. (1997) showed altered fertility when sperm morphology alterations exceeded 10%, but it seems likely that when morphology alterations achieved a certain threshold, decreases in sperm count also could be noted due to multiple apoptotic pathways in the rodent testis responsible for the maintenance of sperm integrity (Sasagawa et al., 2001; Stumpp et al., 2004). Sperm morphology alterations also

have been associated with decreased reproductive organ weights, sperm motility changes, and altered testicular histopathology (Chapin et al., 1985; Linder et al., 1988; Gray et al., 1990); however, changes in sperm shape need not be correlated with alterations in other sperm endpoints (Chapin and Conner, 1999; Morrissey et al., 1988).

Estrous Cycle Evaluation. Data on estrous cycle length were within the expected range (e.g., 4–5 days) across most studies that reported this endpoint. There was slightly greater variability in Wistar rats, where some control animals had cycles as short as 3.6 days. While the typical estrous cycle in rats lasts 4–5 days, there is a background incidence of altered estrous cycles in rats, which can vary from 5% to 20–30% with some rat strains or shipments (Cooper and Goldman, 1999). This inter-animal variability may impact the sensitivity of this endpoint to detect reproductive toxicants. Thus, estrous cycle data are used to complement other data and do not typically indicate an adverse effect alone (Cooper and Goldman, 1999).

The Reproductive Health Effects test guideline states that the test report should include ''an analysis of P and F1 female cycle pattern and mean estrous cycle length.'' While estrous cycle length was frequently reported across laboratories, estrous cycle pattern was either not reported or reported in different manners, making it difficult to determine typical evaluation and reporting practices. The difficulty in reporting estrous cycle pattern may relate to the lack of an agreed-upon method for evaluating estrous cycles across a population (e.g., over the 3-week monitoring period, is an animal's cycle abnormal if it has 3 normal 4–5-day estrous cycles and 1 7-day cycle?). Cooper and Goldman (1999) reported that estrous cycle pattern is an important parameter in order to detect changes that might be masked if only examining estrous cycle length. Estrous cycle reporting is further complicated by the brevity of estrus (E) and proestrus (P) stages in rats, (typically 1–2 d and 0.5–1 d in duration, respectively; Zarrow et al., 1964). These stages can be missed with once-daily vaginal smears. The missing stages have to be inferred to correctly assess cycle normalcy and duration.

Time to Mating and Gestational Length. Time to mating (precoital interval) ranged from 1.8–4.4 and 2.4–4.0 days in first- and second-generation Sprague Dawley rats, respectively. Similar intervals were seen in Wistar rats (2.3–3.1 days for P1 animals and 1.9–3.4 days for P2), indicating that the majority of animals mate during the first estrous cycle.

Mean gestational length was reported in nearly all laboratories. Across both strains, gestational length ranged from 21.5–23.2 days with the majority of values below 23 days. Altered gestational length may be associated with or related to dystocia (difficulty in delivering), litter size, altered postnatal survival rates, and/or altered pup sex ratio and body weights, because male rodents are heavier than female rodents from birth.

Primordial Follicle Counts. Quantitative follicle counts are only required in P2 females, who have been exposed to the test chemical during gestation, lactation, and into adulthood. Primordial follicle counts were among the most variable of the endpoints measured across laboratories. In Sprague-Dawley-derived rats, total primordial follicle counts varied from 10.7 versus 384.0 per sample. The mean primordial follicle count across studies was 116.2 ± 80.1 ($X \pm SD$). Slightly less variability was noted in Wistar-derived rats with a range of 64.0 to 250.0 for total primordial follicle counts (mean = 173.2 ± 49.8).

There are numerous factors that may have contributed to the variability across laboratories. For example, different sampling methods likely contribute to this variability as the number of animals, number of sections, and the selection of ovarian sections for examination are not specified in the current EPA OPPTS or OECD guidelines (US EPA OPPTS, 1998; OECD, 2001), and therefore vary across laboratories (Regan et al., 2005). The test guideline required each laboratory to establish a procedure that was ''statistically valid'' based on references describing this procedure in mice (Bolon et al., 1997; Bucci et al., 1997) with some guidance (e.g., Heindel, 1999) as to how to apply these practices to rats. With each laboratory defining its own procedure, some variability between laboratories was not unexpected. However, greater variability was attributed to inherent biological differences (Table 2). There is inherent intra- and inter-animal variability in the ovarian follicle pool. Bucci et al. (1997) reported 30–40% variability in follicle counts between ovaries taken from the same mice with even greater variability between ovaries from different animals. Thus, large variability between individuals and groups has been reported with primordial follicle counts, making interpretation of follicle count data difficult. Furthermore, quantification of primordial follicles in the F1 females is very time-consuming and expensive. The Society of Toxicologic Pathology has recommended that quantification of primordial follicles be reserved for a Tier 2 evaluation of ovarian effects and triggered in each case based on qualitative histopathology data as well as data from other reproductive endpoints (Regan et al., 2005). This recommendation is supported by the inherent variability in the current data set. As part of this Tier 2 assessment, a full spectrum count of the various follicular stages would provide more useful data to interpret changes in the number of primordial follicles.

Reproductive Organ Weights. Male and female reproductive organ weights, relative and absolute, are presented in Tables 7 (Sprague-Dawley rats) and 8 (Wistar rats), respectively. With bilateral organs (e.g., testes), some laboratories presented unilateral weights; thus, values in Tables 7 and 8 represent unilateral weights either as reported or derived (half of bilateral weights). Overall, CVs were relatively low for reproductive organ weights. Variability was likely related to differences in age/body weight at the time of necropsy, relatively minor differences in dissecting/trimming techniques, and/or inherent biological variability. *Testis Weights*: Differences in sperm production rate often are associated with alterations in testis weight. Furthermore, testis weights in rats have some inherent variability that appears to be independent of spermatogenesis (Blazak et al., 1985; Robb et al., 1978). Even within the normal physiological range, levels of androgen production vary between male rats, which may contribute to variability in other androgen-dependent reproductive organ weights. *Prostate Weights*: The prostate is composed of dorsolateral and ventral lobes, with different profiles (numbers and distribution pattern) of androgen (AR) and estrogen (ER) receptors and, therefore, with different susceptibilities and responsiveness. The prostate is, therefore, weighed whole or with the lobes separated, whereby a total

weight can be derived from the separate lobe weights. Absolute and relative prostate weights in Sprague-Dawley rats were designated as a ''high variability'' endpoint with CVs $\geq$29%. Given that variability was lower in the Wistar studies (CV $<$14% with fewer studies collected in fewer laboratories), it is possible that either strain differences and/or inter-laboratory differences in dissecting techniques contributed to this variability. Notably, a relatively large inter-animal variability in prostate weights has been documented (Elswick et al., 2000) and the variability analysis (Table 2) supports some inter-animal variability in this endpoint in P2 males. *Seminal Vesicle Weights*: The seminal vesicles are usually weighed with coagulating glands (and their fluids) as a single unit. Variability in seminal vesicles weights may be related to differences in the production of seminal fluid. In addition, loss of some fluid during dissection may contribute to some variability in seminal vesicle weights. To avoid this, some laboratories weigh the prostate and seminal vesicles together to prevent seminal fluid loss, then remove the prostate and weigh it separately. The seminal vesicle weight is then determined by subtraction. *Ovarian and Uterine Weights:* A likely contributor to differences in female reproductive organ weights is the stage of the estrous cycle during necropsy. The current US EPA OPPTS (1998) and OECD (2001) guidelines do not require that all females be euthanized during the same stage of the estrous cycle; rather, a determination of estrous stage at the time of necropsy can be used to explain terminal differences in hormone-influenced endpoints. In addition, collection of wet (imbibed) and blotted uterine weights can address some issues related to variability.

Non-Reproductive Organ Weights. Weights for adrenal glands, brain, pituitary, spleen, thymus. and thyroid glands in P1 and P2 adult animals appear in Tables 9A, 9B (male and female Sprague-Dawley rats) and 10A, 10B (male and female Wistar rats). Across organs and strains, CVs were $\leq$20% for all organ weights with the exception of relative brain weights in P2 Sprague-Dawley males and thymus weights. The relative brain weight in P2 males had a CV of 22%, despite low CVs ($\leq$9%) for other brain weight measurements in Sprague-Dawley adult males (i.e., P1 absolute and relative brain weights and P2 absolute brain weights). This variance may be a reflection of relatively stable absolute brain weights in the presence of variable terminal body weights. Absolute and relative thymus weights were considered ''highly variable'' endpoints as CVs ranged from 25–42% in male rats of both strains and approximately 31% in the second-generation female Sprague-Dawley rats. Interestingly, the CVs were $\leq$17% for absolute and relative thymus weights for female Wistar rats and first-generation female Sprague-Dawley rats. The reason for the variability in thymus weights is not known, but both intra- and inter-laboratory factors contribute to this variability (Table 2). Relative organ weight parameters are affected by and dependent on terminal body weight (i.e., a ''normal'' or significantly reduced absolute organ weight and a reduced body weight can result in a significantly increased relative organ weight). Dr. J. Haseman (NIEHS) has suggested that for an organ weight change to be considered ''real.'' both the absolute and relative weights of that organ have to be significantly different and both in the same direction (both increased or both decreased).

Litter size. Data for litter sizes are presented in Tables 11 (Sprague-Dawley rats) and 12 (Wistar rats). For Sprague-Dawley rats, data are separated into litters that were standardized (i.e., culled) or not standardized; all litters were standardized in the Wistar studies. While indices of birth, viability, lactation, and offspring sex ratio are mentioned in the test guideline, these indices often were not reported as independent values in the studies submitted. While some laboratories reported live-birth and survival indices, these variables were reported so infrequently that it was not appropriate to present control data from such small sample sizes. It is possible that laboratories rely on litter size to evaluate live birth rates and survival, making litter size a more robust value for inclusion in this inter-laboratory data set. Litter sizes, sex ratio (% males/litter), and body weights per sex per litter should be collected and reported frequently throughout lactation to detect not just an effect on these parameters, but when this change(s) occurred. From the submitted data, it appears that Sprague-Dawley rats have larger litters than Wistar rats (mean values of $\sim$13.6 vs. $\sim$11.7 pups on PND 0, respectively, when comparing data from litters that were subsequently culled). Based on litter size measurements, pup survival appeared to be similar between the two strains.

Pup Body Weights and Anogenital Distance. Sprague-Dawley and Wistar pup body weights and anogenital distance appear in Tables 13 and 14, respectively. Aside from slightly smaller litter sizes, Wistar pups also weigh slightly less than Sprague-Dawley pups.

The sample size for anogenital distance (AGD) measurements is less than other endpoints, because AGD is a triggered endpoint in the second-generation (F2) offspring, and therefore, not reported in all studies. Triggers for AGD include a change in sex ratio or puberty onset in the F1/P2 offspring. AGD measurements differ by pup age, and AGD was measured on different days across laboratories (i.e., postnatal days 0–4). Data in Tables 13 and 14 reflect AGD measurements collected on PND 1 (day of birth = PND 0). AGD was considered a ''high variability'' endpoint with CVs ranging from 28–42%. Some of this variability can be attributed to the difficulty in accurately measuring small distances as well as the low sample size ($n = 9$) included in this analysis. Furthermore, pup size/body weight also can affect AGD measurements (Gallavan et al., 1999) and some laboratories failed to report body weight at the time AGD was collected. In laboratories reporting body weight, different methodologies were used to adjust for body weight differences, although cubed root of body weight was most frequently used (Gallavan et al., 1999). With these inter-laboratory differences, it was difficult to accurately determine a mean control value for relative AGD measurements. One laboratory also reported AGD measurements in the first-generation (F1) offspring, which may be advantageous for subsequent interpretation of F1/P2 endpoints and comparison with the F2 data (Tyl et al., 2004).

Pup Organ Weights. According to the OPPTS 870.3800 guidelines, brain, spleen, and thymus weights in weanling rats are required. These organ weights are reported in Tables 15 (Sprague-Dawley rats) and 16 (Wistar rats). Uterine weight in weanling females is not

required by EPA OPPTS 870.3800 or OECD 416, but is required per Japan's Ministry of Agriculture, Forestry and Fisheries (JMAFF) testing guidelines as of 2002. With the exception of weanling uterine weights, CVs were $\leq$16% for all pup organ weights. Absolute and relative uterine weights were considered ''highly variable'' endpoints as CVs ranged from 16–52% in weanling female Sprague-Dawley rats. This high variability is likely related to the low sample size for this endpoint ($n = 3$ for Sprague-Dawley rats), although differences in the onset of endogenous estrogen secretion also could contribute to this variability.

Strain Differences: Sprague-Dawley Versus Wistar

Throughout Tables 3–16, control values for Sprague-Dawley and Wistar rats appear to be similar, although in many cases, Wistar data were less variable. This may reflect the number of laboratories that contributed to the respective data sets, because nine companies contributed to the Sprague-Dawley data, whereas the Wistar data originated from only two companies. With respect to male reproductive parameters, Sprague-Dawley rats were noted to have heavier seminal vesicle weights than Wistar rats in the present dataset, a finding that was reported previously (Wilkinson et al., 1999). Wilkinson et al. (1999) also reported greater epididymal sperm counts in Sprague-Dawley rats, which was not confirmed in the current data set.

The Use of Inter-Laboratory Vehicle Control Data

These inter-laboratory vehicle control data are provided to allow laboratories to examine their mean control values in the context of mean control values obtained at other laboratories. With these data, laboratories may gain some assessment of whether their procedures for measuring endpoints are appropriate and if their mean values fall within or outside of the range of values typically obtained for a given endpoint. These data also illustrate endpoints having greater variability due to either inherent variability and/or methodological differences in endpoint assessment across laboratories. It is important to note that while there is a normal range for most parameters, mean control values may periodically fall outside this range due to chance (at $P = 0.05$, there is the recognition that one in 20 significant differences is, in fact, due to chance, i.e., a type I error). Thus, these inter-laboratory control data may be useful in interpreting data from treated animals when a control value is aberrant. However, these data are not intended to supplant each laboratory's own historical control data.

When presenting the inter-laboratory control data, the committee agreed to present the mean (mean of study means) and SD, as well as the minimum, 25th, 50th, 75th percentile, and maximum of all mean values reported for each endpoint. It was thought that these values accurately convey the distribution of mean data for each endpoint across studies; however, the committee members felt that values for some endpoints exceeded the range of biologically plausible variability. Thus, caution should be exercised when using some values at the extremes of the distribution to interpret multi-generation study data, because some variables include maximum control values that overlap with potential treatment-related effects (e.g., 30% abnormal sperm).

When a laboratory's values for specific endpoints are consistently outside the ranges reported (or frequently located on the peripheral ranges of the distribution), a review of laboratory procedures may be warranted. In the event that a laboratory is confident that their practices are scientifically valid and sufficiently sensitive to detect compound-mediated changes, the laboratory must rely on its own historical control data. For some highly variable endpoints, performance criteria may be needed.

Missing Data

Interestingly, control data were not reported for all endpoints in all 44 of the submitted reports. There are three possible explanations to explain these missing data points: (1) the endpoints were examined, but the data were not reported, (2) the endpoints were examined, but the data were reported in a different format, or (3) the endpoints were not evaluated in all of the submitted studies.

The possibility that endpoints were examined but the data were not reported would not be surprising. To compile the data for analyses, companies were asked to submit the report text and the summary tables of the study results. Due to the size of each multi-generation study report, individual animal data and appendices for the reports were typically not requested. Thus, if values for a specific endpoint were omitted from the report text or summary tables, these data would be absent in the dataset. In an effort to minimize the amount of missing data, data gaps were identified and the text of the report was reviewed for evidence that the endpoint in question was examined in the submitting laboratory. If the endpoint was mentioned in the text, the laboratory was contacted and asked to provide the missing data (e.g., submission of individual animal data, if needed). However, in many cases, the missing endpoint(s) were not mentioned in the text. It is possible that these data were collected, but in the absence of treatment-related effects, the data for all endpoints were not presented in either the text or summary tables, depending on the practice of each laboratory.

In some cases, the endpoints may have been examined, but the data were reported in a different manner. For example, instead of presenting ''mean estrous cycle length,'' some laboratories report the ''number or percent of females with normal/abnormal cycles''; thus, while these laboratories evaluated the estrous cycle, they reported it in a manner that was not readily incorporated into this analysis. In actuality, both mean cycle duration and numbers (and types) of abnormal cycles within the evaluation period are useful to better characterize female cyclicity. ''Days to mating'' could be reported as the ''number of animals mating during different cycle intervals'' (e.g., number mated during the first cycle on days 1–4, the second cycle on days 5–8, etc.). Some laboratories report ''number of implantations'' and ''live litter size on day 0'' as opposed to ''percent postimplantation loss.'' Epididymal sperm concentration could be reported as ''number of sperm per sample volume'' (e.g., 10^7 sperm/ml) as opposed to per g epididymal tissue weight, which would still allow comparison of sperm

concentrations across treatment groups, assuming that the same sample volume was used for each sample. An estimate of sperm count/g cauda epididymal tissue could be derived using the individual cauda epididymal weights associated with each male's sperm count (preferred method as it is most accurate), using each whole epididymis weight if the cauda were not weighed separately (less accurate), or using the mean cauda or mean epididymal weights, although there would be imprecision in this calculated value. Overall, these examples illustrate different practices by which data on the same functional parameters may be reported. Thus, these studies may, in fact, provide the necessary data to assess effects on the reproductive system, but due to formatting differences, these data were not amenable for entry into the database. In appropriate cases, some values were derived from the data reported in an alternate format (e.g., percent abnormal sperm based on the percent normal sperm, unilateral organ weights based on bilateral organ weights, etc.).

It also is possible that some endpoints were not examined in all laboratories. Some endpoints are triggered based on other results in the study (e.g., anogenital distance); therefore, if the triggering criteria were not met, these endpoints may not be included in a given study. Some reports from 1997 and 1998 also were included in the database. These early reports were likely initiated prior to finalizing the test guidelines in 1998, and, thus, may not have included all endpoints. Furthermore, the test guidelines may be interpreted differently by different laboratories, such that some laboratories collected all "suggested" endpoints, whereas other laboratories did not. Lastly, it is possible that laboratories chose not to collect and/or report all required endpoints.

Despite some data gaps where data for each endpoint were not received from every laboratory, the sample sizes for the inter-laboratory control values were reasonable. Thus, the steering committee felt that the mean values were a reasonable representation of control means in most cases.

Comparison With Other Published Studies

To further evaluate the inter-laboratory control data set, data were compared with reproductive toxicity data reported in the scientific literature. Table 17 shows data from 4 reproductive toxicity studies that were used for this comparison. The published data from these studies were evaluated against the database to insure that these studies were not included in the inter-laboratory control data set. Overall, the data in the inter-laboratory control data set were comparable to the values reported in the scientific literature.

CONCLUSIONS

This work presents a compilation of inter-laboratory control data that can be used by investigators to examine variability in reproductive endpoint values. For endpoint values identified as highly variable, further consideration should be given to identifying and limiting potential source(s) of the variability. Procedural variability may be due to: Presence/absence of SOPs; inter-laboratory differences in SOPs; differences in training, experience, and/or equipment; inter-technician variability in data collection (are technicians "blind" to treatment group?); time of day for data collection; and order of data collection (i.e., are they distributed across dose groups rather than completing collection in each dose group sequentially?). These inter-laboratory control data also provide a means for laboratories to review their performance on reproductive toxicity measures, as well as providing some perspective for interpreting their own control data and data from treated animals.

ACKNOWLEDGMENTS

The ILSI HESI DART Committee gratefully acknowledges the voluntary submission of data from the following companies, without whose assistance this project would not have been possible: The American Chemistry Council (Arlington, VA), BASF (Ludwigshafen, Germany), Bayer Cropsciences (Stilwell, KS), Charles River Laboratories (Horsham, PA), The Dow Chemical Company (Midland, MI), Dow-Corning Corporation (Midland, MI), DuPont Haskell Laboratories (Newark, DE), Exxon-Mobil (Annandale, NJ), Syngenta (Alderley Park, Macclesfield, UK), Valent (Dublin, CA), and WIL Laboratories (Ashland, OH).

REFERENCES

Blazak WF, Ernst TL, Stewart BE. 1985. Potential indicators of reproductive toxicity: testicular sperm production and epididymal sperm number, transit time and motility in F344 rats. Fundam Appl Toxicol 5:1097–1103.

Blazak WF, Treinen KA, Juniewicz PE. 1993. Application of testicular sperm head counts in the assessment of male reproductive toxicity. In: Chapin RE, Heindel JJ, editors. Methods in toxicology, Part A. Male Reproductive Toxicology. New York: Academic Press. p 86–94.

Bolon B, Bucci TJ, Wargritton AR, Chen JJ, Mattison DR, Heindel JJ. 1997. Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. Fundam Appl Toxicol 39:1–10.

Bucci TJ, Bolon B, Warbritton AR, Chen JJ, Heindel JJ. 1997. Influence of sampling on the reproducibility of ovarian follicle counts in mouse toxicity studies. Reprod Toxicol 11:689–696.

Chapin RE, Conner MW. 1999. Testicular Histology and Sperm Parameters. In: Daston G, Kimmel C, editors. An Evaluation and Interpretation of Reproductive Endpoints for Human Health Risk Assesssment. Washington, DC: International Life Sciences Institute. p 28–41.

Chapin RE, Dutton SL, Ross MD, Lamb 4th JC. 1985. Effects of ethylene glycol monomethyl ether (EGME) on mating performance and epididymal sperm parameters in F344 rats. Fundam Appl Toxicol 5:182–189.

Chapin RE, Filler RS, Gulati DK. 1992. Methods for assessing rat sperm motility. Reprod Toxicol 6:267–273.

Chapin RE, Sloane RA, Haseman JK. 1997. The relationships among reproductive endpoints in Swiss mice using the reproductive assessment by continuous breeding database. Fundam Appl Toxicol 38:129–142.

Cooper RL, Goldman JM. 1999. Vaginal cytology. In: Daston G, Kimmel C, editors. An evaluation and interpretation of reproductive endpoints for human health risk assesssment. Washington, DC: International Life Sciences Institute. p 42–56.

Elswick BA, Welsch F, Janszen DB. 2000. Effect of different sampling designs on outcome of endocrine disruptor studies. Reprod Toxicol 14:359–367.

Foster PMD. 1999. Reproductive Performance. In: Daston G, Kimmel C, editors. An evaluation and interpretation of reproductive endpoints for human health risk assesssment. Washington, DC: International Life Sciences Institute. p 5–9.

Gallavan Jr RH, Holson JF, Stump DG, Knapp JF, Reynolds VL. 1999. Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. Reprod Toxicol 13:383–390.

Goldman JM, Laws SC, Balchak SK, Cooper RL, Kavlock RJ. 2000. Endocrine-disrupting chemicals: Prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. Crit Rev Toxicol 30:135–196.

Gray Jr LE, Ostby J, Linder R, Goldman J, Rehnberg G, Cooper R. 1990. Carbendazim-induced alterations of reproductive development and function in the rat and hamster. Fundam Appl Toxicol 15: 281–297.

Heindel JJ. 1999. Oocyte quantitation and ovarian histology. In: Daston G, Kimmel C, editors. An evaluation and interpretation of reproductive endpoints for human health risk assesssment. Washington, DC: International Life Sciences Institute. p 57–74.

Linder RE, Rehnberg GL, Strader LF, Diggs JP. 1988. Evaluation of reproductive parameters in adult male Wistar rats after subchronic exposure (gavage) to benomyl. J Toxicol Environ Health 25:285–298.

Meistrich ML. 1989. Evaluation of reproductive toxicity by testicular sperm head counts. J Am Coll Toxicol 8:551–567.

Morrissey RE, Schwetz BA, Lamb 4th JC, Ross MD, Teague JL, Morris RW. 1988. Evaluation of rodent sperm, vaginal cytology, and reproductive organ weight data from National Toxicology Program 13-week studies. Fundam Appl Toxicol 11:343–358.

Odum J, Tinwell H, Tobin G, Ashby J. 2004. Cumulative dietary energy intake determines the onset of puberty in female rats. Environ Health Perspect 112:1472–1480.

OECD (Organisation for Economic Cooperation and Development). 2001. Guideline for Testing of Chemicals, No. 416, ''Two-generation Reproduction Toxicity Study,'' p 1–13.

Regan KS, Cline JM, Creasy D, Davis B, Foley GL, Lanning L, Latendresse JR, Makris S, Morton D, Rehm S, Stebbins K. 2005. STP position paper: ovarian follicular counting in the assessment of rodent reproductive toxicity. Toxicol Pathol 33:409–412.

Robb GW, Amann RP, Killian GJ. 1978. Daily sperm production and epididymal sperm reserves of pubertal and adult rats. J Reprod Fertil 54:103–107.

Sasagawa I, Yazawa H, Suzuki Y, Nakada T. 2001. Stress and testicular germ cell apoptosis. Arch Androl 47:211–216.

Schneider S, Deckardt K, Hellwig J, Küttler K, Mellert W, Schulte S, van Ravenzwaay B. 2005. Octyl methoxycinnamate: Two generation reproduction toxicity in Wistar rats by dietary administration. Food Chem Toxicol 43:1083–1092.

Seed J, Chapin RE, Clegg ED, Dostal LA, Foote RH, Hurtt ME, Klinefelter GR, Makris SL, Perreault SD, Schrader S, Seyler D, Sprando R, Treinien KA, Veeramachaneni DNR, Wise LD. 1996. Methods for assessing sperm motility, morphology and counts in the rat, rabbit, and dog: A consensus report. Reprod Toxicol 10:237–244.

Slott VL, Perreault SD. 1993. Computer-assisted sperm analysis of rodent epididymal sperm motility using the Hamilton-Thorn motility analyzer. In: Chapin RE, Heindel JJ, editors. Methods in toxicology, Part A. Male Reproductive Toxicology. New York: Academic Press. p 319–333.

Stoker TE, Parks LG, Gray LE, Cooper RE. 2000. Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC recommendations. Crit Rev Toxicol 30:197–252.

Stumpp T, Sasso-Cerri E, Freymüller E, Miraglia SM. 2004. Apoptosis and testicular alterations in albino rats treated with etoposide during the prepubertal phase. Anat Rec (Part A) 279:611–622.

Suter L, Clemann N, Koch E, Bobadilla M, Bechter R. 1998. New and traditional approaches for the assessment of testicular toxicity. Reprod Toxicol 12:39–47.

Tyl RW, Myers CB, Marr MC, Fail PA, Seely JC, Brine DR, Barter RA, Butala JH. 2004. Reproductive toxicity evaluation of dietary butyl benzyl phthalate (BBP) in rats. Reprod Toxicol 18:241–264.

US EPA OPPTS (Office of Prevention, Pesticides and Toxic Substances). 1998. Health Effects Test Guidelines. OPPTS 870.3800. Reproduction and Fertility Effects, August 1998.

Wilkinson JM, Halley S, Towers PA. 1999. Comparison of male reproductive parameters in three rat strains: Dark Agouti, Sprague-Dawley and Wistar. Lab Anim 34:70–75.

Willoughby CR, Fulcher SM, Creasy DM, Heath JA, Priston RAJ, Moore NP. 2000. Two-generation reproduction toxicity studies of di-(C_7-C_9 alkyl) phthalate and di-(C_9-C_{11} alkyl) phthalate in the rat. Reprod Toxicol 14:427–450.

Zarrow MX, Yochim JM, McCarthy JL, Sanborn RC. 1964. Experimental endocrinology. A sourcebook of basic techniques. New York: Academic Press.