Critique of Vreugdenhil et al.'s Study Linking PCBs to the Play Behaviors of Dutch Girls and Boys

Vreugdenhil et al. (2002) administered the Pre-School Activities Inventory (PSAI) (Golombok and Rust 1993a, 1993b) to 158 Dutch girls and boys and concluded that higher prenatal exposure to polychlorinated biphenyls (PCBs) in boys was related to less masculinized play, and in girls was related to more masculinized play. They further concluded that prenatal exposure to PCBs and related compounds caused prenatal steroid hormone imbalances, leading directly to sex-inappropriate play behaviors. However, this study has many flaws that preclude reaching these conclusions.

The PSAI has weak psychometric properties. The test–retest reliabilities in the 0.60s are based on tiny samples of 15–18. The 1-year interval is too long to measure the stability of test scores; long intervals confound the test's stability with real changes in children's behavior. The split-half reliability is adequate for girls (0.80) but poor for boys (0.66). Golombok and Rust (1993b) provided split-half reliability for the total sample, and they hailed the value of 0.88 as "robust." However, that value is spuriously high because of the bimodal distribution of test scores when combining sexes. Interestingly, Golombok and Rust (1993b) noted that the value of 0.84 for the test-retest reliability of the combined sample of 33 boys and girls is spuriously high because of bimodality; that argument also applies to split-half coefficients.

The construct validity of the PSAI is suspect. Golombok and Rust (1993b) failed to match socioeconomic background of mothers of boys with that of mothers of girls, and they based the final instrument on data from 32 boys and 43 girls, samples too small to yield generalizable data. They provided only a single PSAI validity study, despite subsequent testing of > 2,000 parents of young children; also, validity data from London may not generalize to data from the Netherlands.

Also, Vreugdenhil et al. (2002) reported no studies to validate the parents' perceptions—they did not actually observe children's play activities, an essential aspect of test validity, as Golombok and Rust (1993b) noted for previous play-preference instruments. Parents' perceptions may be biased and require validation.

Vreugdenhil et al. (2002) interpreted the PSAI composite score (male items minus female items); however, PSAI psychometric data are provided only for the total score. Consequently, the reliability and validity data do not even apply to the "difference" score used by the authors; indeed, difference scores are notoriously unreliable. Golombok and Rust (1993a, 1993b) provided no data on the separate male and female items; Vreugdenhil et al. (2002) presented no rationale for analyzing data for the so-called masculine and feminine scales.

Vreugdenhil et al. (2002) inappropriately used the interaction term "sex × exposure" in their regression analysis. This analysis involves combined groups of boys and girls, a serious problem because bimodal distributions spuriously inflate correlations; that includes coefficients involving the sex × exposure interaction, conceivably explaining the "significant" PCB results. Also, Vreugdenhil et al. used age-unadjusted scores, fine for comparing boys with girls, according to Golombok and Rust (1993b), but requiring care "in the choice of appropriate statistics when [combining] data from both sexes … in the same analysis" (p. 134). Vreugdenhil et al. (2002) did not heed this warning, again ignoring bimodal distributions.

The PSAI is likely not valid for school-age children. The average age of the Dutch children was 7.5 years, whereas the PSAI was developed and designed for preschool children; its oldest norms group is 60–71 months, much younger than the average age of the Dutch children.

Vreugdenhil et al. (2002) made many multiple comparisons, but they did not take the chance errors into account, a problem that affects many PCB studies (Cicchetti and Kaufman 2002; Kaufman 2002) and studies of lead level as well (Kaufman 2001, Phelps 1999), further compromising the meaningfulness of any reported significant effects. The authors related parents' perceptions of their young children's behaviors to sex steroid hormones but offered no direct evidence that PSAI scores are in any way related to sex steroid hormones—nor have Golombok and Rust made such claims for their test.

Vreugdenhil et al. (2002) referred to the Yu-Cheng sample and inferred that gender differences on Raven's Matrices tests are evidence of sex-specific differences in spatial ability. This is an error; Raven's tests primarily assess reasoning ability, not spatial ability. Spatial tests such as Block Design do produce sex-specific results (Jensen 1980), but sex differences are not usually found with Raven's tests or similar matrices tests [see Table 4.33 in Kaufman and Kaufman (1983)].

Overall, Vreugdenhil et al. (2002) used a flawed instrument and made other methodological errors that should cause them and other researchers to question their significant findings and their conclusions.

Alan S. Kaufman
Yale Child Study Center
Yale University School of Medicine
New Haven, Connecticut
*Current address: Escondido, California
E-mail: Alanadeen@att.net

REFERENCES AND NOTES

Cicchetti DV, Kaufman AS. 2002. Does PCB exposure have any effect on later psychomotor and mental development? [Letter]. Lancet 359:1438.

Golombok S, Rust J. 1993a. The measurement of gender role behaviour in pre-school children: a research note. J Child Psychol Psychiatry 34:805–811.

———. 1993b. The Pre-school Activities Inventory: a standardized assessment of gender role in children. Psychol Assess 5:131–136.

Jensen AR. 1980. Bias in Mental Testing. New York:Free Press.

Kaufman AS. 2001. Do low levels of lead produce IQ loss in children? A careful examination of the literature. Arch Clin Neuropsychol 16:303–341.

———. 2002. PCB-induced impairments in older adults: critique of Schantz et al.'s methodology and conclusions [Letter]. Environ Health Perspect 110:A70–A71.

Kaufman AS, Kaufman NL. 1983. K-ABC Interpretive Manual. Circle Pines, MN:American Guidance Service.

Phelps L. 1999. Low-level lead exposure: implications for research and practice. Sch Psychol Rev 28:477–492.

Vreugdenhil HJI, Slijper FME, Mulder PGH, Weisglas-Kuperus N. 2002. Effects of perinatal exposure to PCBs and dioxins on play behavior in Dutch children at school age. Environ Health Perspect 110:A593–A598.

"Linking PCBs to Play Behaviors of Dutch Girls and Boys": Vreugdenhil et al.'s Response

Kaufman's comments on the findings we presented in our article "Effects of Prenatal Exposure to PCBs and Dioxins on Play Behavior in Children at School Age" (Vreugdenhil et al. 2002) is based on three points: the instrument, the age of the children, and the use of an interaction term in our analysis. We would like to respond to these questions.

The Pre-School Activities Inventory (PSAI) is a very simple parent questionnaire. We included the 24 questions in our paper in the form of an appendix. Our data on the PSAI show that boys scored significantly higher on the masculine scale than girls, and correspondingly, girls scored significantly higher on the feminine scale than boys. The composite scores for boys also indicated an overall masculine score, and correspondingly, the composite scores for girls indicated an overall feminine score. These data show that the PSAI is a valuable instrument in assessing masculine and feminine play behavior in Dutch boys and girls at 7 years of age. We

agree with Kaufman that parent's perceptions might be biased. That we found significant differences in the effects of prenatal PCB exposure with such a relatively simple parent questionnaire in our exploratory study makes our findings even more relevant.

Because we were interested in effects of PCB exposure on masculine and feminine play behavior, we presented our results for the masculine, feminine, and composite (defined as the within-subject difference; feminine score minus masculine score) PSAI scores. We estimated the effect of PCB exposure on these scores and the difference in effect between males and females by fitting one regression model for each outcome variable (masculine, feminine, and composite score) in the combined data set of males and females. Because this was an observational study, we included a number of confounding variables in the linear model. The set of covariates taken along with exposure (the variable of interest) included type of feeding in infancy, duration of breast-feeding, sex, parity, parental education level, parental IQ, the home environment, and the age of the child. First, this means that through the variable "sex" the model allows for a bimodal distribution of the scores, so that the results are neatly and automatically adjusted for that bimodality. Indeed, as expected, the distribution of the residuals estimated from the regression analysis no longer shows that bimodal property. Second, through the variable age in the model, the estimated effects of exposure on the outcome variables have automatically become age-adjusted. Of course, this also means that the difference between boys and girls in the exposure effect (as represented by the coefficient of the sex by the exposure–interaction term) is automatically adjusted for the sex-bimodality and for age, as well as for any of the other covariates in the model. Therefore, we do not see any of the methodological flaws mentioned by Kaufman, and we hope that we have hereby reassured Kaufman on this matter.

We presented the p-values of the statistical tests as calculated to three decimal places. Because of the exploratory nature of this observational study, we did not apply multiplicity correction on an overall significance level to obtain a significance level per test. We think that the scientific audience should be free to decide how much significance to attach to each result, given the corresponding reported p-value and prior knowledge of the results of other publications on this subject. We do not understand how the absence of a rather arbitrary multiplicity correction per test on an (also rather arbitrary) overall significance level has led Kaufman to try to convince others to place little confidence in

our findings. How else can scientists build evidence unless we present our results, including exploratory studies.

The authors declare they have no conflict of interest.

Hestien J.I. Vreugdenhil
Erasmus University Rotterdam
and Sophia Children's Hospital
Rotterdam, the Netherlands

Paul G.H. Mulder
Erasmus University Rotterdam
Rotterdam, the Netherlands

Nynke Weisglas-Kuperus
Erasmus University Rotterdam
and Sophia Children's Hospital
Rotterdam, the Netherlands
E-mail: weisglas@alkg.azr.nl

REFERENCES

Vreugdenhil HJI, Slijper FME, Mulder PGH, Weisglas-Kuperus N. 2002. Effects of perinatal exposure to PCBs and dioxins on play behavior in Dutch children at school age. Environ Health Perspect 110:A593–A598

Dutch Girls and Boys, PCB Levels, and Play Behavior: What Do the Data Really Tell Us?

In their article in the October 2002 issue of *EHP*, Vreugdenhil et al. (2002) concluded that

Childhood play behavior shows marked sex differences and is likely to be influenced by the prenatal steroid hormone environment,

and, more specifically, that

Higher prenatal exposure to PCBs was associated with less masculinized play behavior in boys and with more masculinized play behavior in girls.

The data of Vreugdenhil et al. (2002), based on the Pre-School Activities Inventory (PSAI), showed (all at $p < 0.001$) that boys score significantly higher on the masculine scale (24.2 ± 5.3; mean ± SD) than on the feminine scale (9.6 ± 3.3); correspondingly, girls (26.4 ± 6.2) score almost three times as high as boys (9.6 ± 3.3) on the feminine scale. Further, boys' masculine scores (24.2 ± 5.3) are nearly twice as high as girls' masculine scores (12.6 ± 4.5). Finally, composite scores for boys indicate an overall masculine score of –14.6 (9.6 – 24.2); correspondingly, composite scores for girls indicate an overall feminine score of about 14, since 26.4 – 12.6 = 13.8.

These data indicate clearly that the major outcome is not at all what the authors report. Further analyses indicate that there are no

meaningful differences whatsoever in maternal cord PCB levels for boy and girl samples, or 0.42 µg/L vs. 0.40 µg/L, respectively, when weighted for differences in sample sizes.

To discover the actual relationship between the maternal cord PCB levels and the scores on the masculine scale of the PSAI, we need to examine the data.

Vreugdenhil et al. (2002) are to be faulted for not defining precisely how both the PCB cord levels were transformed into a scale ranging between –2.0 and +2.0 (the x-axis) and, correspondingly, how the masculine scale was transformed so that it could range between –20 and +20. More specifically, how does this transformed scale translate back into the real or actual scores that produced the overwhelming evidence that boys behaved like boys and girls behaved like girls?

That said, the raw data, whatever the scores actually mean, indicate that the preponderance of the data points center close to zero on both PCB levels and scores on the masculine scale, indicating that for the preponderance of mother–child pairs there was a correlation of approximately zero between PCB levels and scores on the masculine scale. This lack of a meaningful effect is very pronounced for boys, but even more pronounced for girls, where the resulting partial correlation is close to zero (+0.17) and fails to even approach statistical significance.

The correlation for boys at –0.29 is still exceedingly low, namely, statistically significant, but clinically quite meaningless. For girls the amount of variation in PCB maternal cord levels that is explained by the variation in masculine scale scores is a dismal 0.17^2, or 2.89%, leaving 97.11% of unexplained variation in masculine scores as a function of variation in PCB maternal cord levels. Correspondingly, the amount of explained variation in masculine scores is 0.29^2, or 8.41%, leaving 91.59% unexplained variation in masculine scores as a function of variation in PCB maternal cord levels.

Clearly then, the results presented by Vreugdenhil et al. (2002) are misleading and totally inaccurate. These inaccuracies came about because *a*) the authors performed 39 multiple regressions (although it is likely that many more were performed, proved negative, and were therefore not reported), and *b*) they did not consider the number of results that would be statistically significant by chance alone.

Vreugdenhil et al. (2002) should have controlled for the number of comparisons that could have occurred by chance alone by dividing 0.05 by the number of reported comparisons, or 39, to produce an adjusted p level of 0.001. Had they performed this

absolutely necessary correction for Type I error (e.g., Toothaker 1991), they would have discovered that of the 39 analyses, only 1 is statistically significant, this at 0.001. Technically there are two, if one chooses to ignore the fact that the sex × exposure variable (very poorly defined) is redundant with masculine scale scores, since each comprises one of the two components of this interaction term. Failing to institute this necessary correction, the authors falsely reported six (25%) of the obtained results as statistically significant, rather than only one, which is less than the two that would have been expected by chance alone. The authors' reported findings lack scientific merit and should therefore be dismissed by the scientific community.

The author has previously consulted for General Electric but did not receive remuneration for any of the work involved in preparing this manuscript.

Domenic V. Cicchetti
Yale University School of Medicine
New Haven, Connecticut
E-mail: dom.Cicchetti@yale.edu

REFERENCES

Toothaker, LE. 1991. Multiple Comparisons for Researchers. Newbury Park, CA:Sage Publications.
Vreugdenhil HJI, Slijper FME, Mulder PGH, Weisglas-Kuperus N. 2002. Effects of perinatal exposure to PCBs and dioxins on play behavior in Dutch children at school age. Environ Health Perspect 110:A593–A598.

"Dutch Girls and Boys: Response from Vreugdenhil et al.

Cicchetti's comments about our article "Effects of Prenatal Exposure to PCBs and Dioxins on Play Behavior in Children at School Age" (Vreugdenhil et al. 2002) are based on a multitude of misinterpretations and misunderstandings.

Cicchetti correctly noted that our data show that on the Pre-School Activities Inventory (PSAI) boys score significantly higher on the masculine scale than girls, and correspondingly, girls score significantly higher on the feminine scale than boys. The composite scores for boys also indicate an overall masculine score and, correspondingly, the composite scores for girls indicate an overall feminine score. He also correctly noted that mean cord PCB levels are virtually the same in boys and girls; however, this does not rule out the possibility of a difference in effect of cord PCB levels on a PSAI scale between boys and girls. Cicchetti mistakenly concludes that our major outcome,

that a higher prenatal exposure to PCBs was associated with less masculinized play behavior in boys and with more masculinized play behavior in girls,

is not at all what we report. The interaction sex × exposure represents the difference in effect of, for example, $\ln\Sigma PCB_{cord}$ (the log-normal concentration of the sum of PCBs in cord plasma) on a PSAI scale between girls and boys; it represents the effect among girls minus the effect among boys, which is shown in Tables 2 and 3 of our paper (Vreugdenhil et al. 2002). The estimate of this difference is based on the assumptions that the other explanatory variables have the same effects in boys and girls and that the variance of the residual term is the same in boys and girls. There is no suspicion that these assumptions would not hold true for our data.

Cicchetti also has problems with our Figure 1 (Vreugdenhil et al. 2002). This figure includes two ordinary partial regression plots of the residuals of the masculine scale and $\ln\Sigma PCB_{cord}$, when both these variables are regressed upon the other independent (confounding) variables. The slopes of these regression lines coincide with the regression coefficients given in our Table 2 (Vreugdenhil et al. 2002). The null hypothesis of these regression coefficients being zero coincides with the partial correlations in Figure 1 being equal to zero, so that the *p*-values coincide. In order for a multiple linear regression analysis to be valid, it is not necessary to make assumptions on the distribution of the independent variables (e.g., $\ln PCB_{cord}$). Also, it is well known that, in a multiple regression analysis, the percentage of explained variability is a different concept than validity of the estimated regression coefficients. A small percentage explained by a fitted model does not necessarily invalidate the estimated coefficients of that model.

Cicchetti claims that we performed 39 (and likely many more) regressions. This is again obviously due to a misunderstanding or misinterpretation of our paper (Vreugdenhil et al. 2002). For the effect in boys, the effect in girls, and the difference in effect between boys and girls, we used only one regression model. Moreover, the composite score is the difference between the feminine score and the masculine score. Hence, the regression results of the composite score can be essentially derived from the regressions of the feminine and masculine score, up to one parameter (the correlation of the residuals of the feminine score and the masculine score). The multiplicity correction suggested by Cicchetti for the significance level per test is based on mutual independence of the statistical tests and is therefore known to be rather

conservative. This holds more strongly for the results of our exploratory study. Because of the nature of our study, we presented the *p*-values as calculated to three decimal places and we did not propose a multiplicity correction for the significance level per test. Control for chance finding (or actually lack of control) is, in our opinion, an improper criterion to disregard the results of a study or even to leave them unpublished. This is because it is based on a rather arbitrarily and externally (to the study) chosen method to lower the already arbitrarily chosen threshold for the *p*-value. Moreover, strictly applying this criterion as proposed by Cicchetti would prevent science from storing and building up evidence. Certainly it does not make sense to divide the overall significance level by the spurious number 39 as he proposed for the reasons given above.

Cicchetti proposes that our findings should be dismissed from the scientific community because not all of our *p*-values are smaller than his proposed spurious multiplicity correction for the significance level per test. We think however, that the scientific community itself is perfectly able to judge the merits of our results by its own means.

The authors declare they have no conflict of interest.

Hestien J.I. Vreugdenhil
Erasmus University Rotterdam
and Sophia Children's Hospital
Rotterdam, the Netherlands

Paul G.H. Mulder
Erasmus University Rotterdam
Rotterdam, the Netherlands

Nynke Weisglas-Kuperus
Erasmus University Rotterdam
and Sophia Children's Hospital
Rotterdam, the Netherlands
E-mail: weisglas@alkg.azr.nl

REFERENCES

Vreugdenhil HJI, Slijper FME, Mulder PGH, Weisglas-Kuperus N. 2002. Effects of perinatal exposure to PCBs and dioxins on play behavior in Dutch children at school age. Environ Health Perspect 110:A593–A598.

Adverse Health Effects of Bisphenol A in Early Life

In their paper "Parent Bisphenol A (BPA) Accumulation in the Human Maternal–Fetal–Placental Unit," Schönfelder et al. (2002) suggested that "long-term follow-up studies are needed to assess the adverse effects of BPA exposure in early life." Two long-term exposure studies (multigenerational reproductive and developmental studies) have recently been published (Ema et al.

2001; Tyl et al. 2002). Neither provides evidence of any effect of BPA at the levels reported by Schönfelder et al. (2002).

In the long-term study by Ema et al. (2001), conducted by the Chemical Compound Safety Research Institute of Japan, Crj:CD (SD) IGS rats were dosed each day with BPA (0, 0.2, 2.0, 20, or 200 µg/kg/day) by stomach tube over two generations. Assessments included parental growth rate, food intake, reproductive performance, sperm production and motility, gross pathology and histopathology, organ weight, litter size, pup survival and growth, and anogenital distance. In addition, Ema et al. measured levels of several hormones related to reproduction, reflex development, and maze performance. Upon analysis of the data for all of these end points for the parental generation and the F_1 and F_2 generations, no consistent evidence of a low-dose effect of BPA was found.

In the study by Tyl et al. (2002), conducted by the Research Triangle Institute in the United States, Sprague-Dawley rats were fed a diet containing BPA at levels from 0 to 7,500 ppm, yielding approximate intakes of 0, 1, 20, 300, 5,000, 50,000, and 500,000 µg/kg/day. Exposures were continued until adulthood of the third-generation offspring. The end points evaluated included parental growth rate, food intake, reproductive performance, sperm production and motility, gross pathology and histopathology, organ weights, litter size, pup survival and growth, and anogenital distance. In addition Tyl et al. measured the day of vaginal opening, preputial separation, and in males, the presence or absence of retained nipples. The lowest observed adverse effect level (LOAEL) in this study was 50,000 µg/kg/day, and the effects observed at the LOAEL were weight loss or reduction in weight gain. No effects were observed at lower doses.

Reassuringly, the results of the two available long-term studies provide no evidence—despite the exceptional power of the studies—of any effect of BPA exposure at levels near or orders of magnitude higher than those reported by Schönfelder et al. (2002).

The author declares he has no conflict of interest.

John E. Heinze
Environmental Health Research Foundation
Manassas, Virginia
E-mail: jheinze@ehrf.info

REFERENCES

Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazono A. 2001. Rat two-generation reproductive toxicity study of bisphenol A. Reprod Toxicol 15:505–523.

Schönfelder G, Wittfoht W, Hopp H, Talsness CE, Paul M, Chahoud I. 2002. Parent bisphenol A accumulation in the human maternal–fetal–placental unit. Environ Health Prospect 110:A703–A707.

Tyl RW, Myers CB, Marr MC, Thomas BF, Keimowitz AR, Brine DR, et al. 2002. Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. Toxicol Sci 68:121–146.

Adverse Health Effects of Bisphenol A: Chahoud's Response

Heinze stated in his letter that according to two long-term studies, bisphenol A (BPA) does not induce any effect on reproduction in offspring at the low dose level. Several other studies have reported adverse effects of BPA; I briefly discuss two examples below.

Markey et al. (2001) investigated the effect of fetal exposure to BPA [25 and 250 µg/kg body weight (bw)] on the development of the mammary gland in CD-1 mice. They concluded their results as follows:

> The altered relationship in DNA synthesis between the epithelium and stroma and the increase in terminal ducts and terminal end buds are striking, because these changes are associated with carcinogenesis in both rodents and humans.

Kawai et al. (2003) carried out a study to evaluate the effect of fetal exposure to BPA (2 and 20 ng/kg bw) on male offspring. They observed that *in utero* exposure at these dose levels resulted in significantly reduced relative testis weight and concluded that low doses of BPA interfered with the normal development of reproductive organs.

I would like to take the opportunity to discuss the problem of the interpretation of so-called negative studies.

Ashby et al. (1999) aimed to disprove studies published by vom Saal and colleagues (Nagel et al. 1997; vom Saal et al. 1997, 1998). Ashby et al. were not able to confirm the results described by vom Saal and colleagues; however, their study (2 and 20 µg BPA/kg bw) shows significantly elevated testis and epididymal weights, even after adjustment for body weight. Ashby et al. considered this clear effect "an equivocal finding."

Tyl et al. (2002) conducted a three-generation reproductive toxicity study on dietary BPA in CD Sprague-Dawley rats. The F_2 generation showed no statistically significant difference in body weight compared to the control. However, at doses of 1 µg, 300 µg, and 5,000 µg BPA/kg bw, the absolute and relative paired ovary weights exhibited a significant decrease in the F_2 generation compared to control. Tyl et al. considered these effects not biologically significant.

Investigators are in the position to interpret the adversity of their own data, and readers also have the freedom to build their own opinion regarding the adversity of the effects. In conclusion, I would like to emphasize the need for mechanistic experimental studies as well as follow-up studies in humans regarding low-dose effects.

The author declares he has no conflict of interest.

Ibrahim Chahoud
Institute of Clinical Pharmacology
and Toxicology
Free University Berlin
Berlin, Germany
E-mail: chahoud@zedat.fu-berlin.de

REFERENCES

Ashby J, Tinwell H, Hasemann J. 1999. Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF-1 mice exposed in utero. Regul Toxicol Pharmacol 30:156–166.

Kawai K, Nozaki T, Nishikata H, Aou S, Takii M, Kubo C. 2003. Aggressive behavior and serum testosterone concentration during the maturation process of male mice: the effects of fetal exposure to bisphenol A. Environ Health Perspect 111:175–178.

Markey CM, Luque EH, Munoz de Toro M, Sonnenschein C, Soto AM. 2001. In utero exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. Biol Reprod 65:1215–1223.

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

Tyl RW, Myers CB, Marr MC, Thomas BF, Keimowitz AR, Brine DR, et al. 2002. Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. Toxicol Sci 68:121–146.

vom Saal FS, Cooke PS, Buchanan DL, Palanza P, Thayer KA, Nagel SC, et al. 1997. A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production, and behavior. Toxicol Ind Health 14(1-2):239–260.

vom Saal FS, Timms BG, Montano MM, Palanza P, Thayer KA, Nagel SC, et al. 1998. Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc Natl Acad Sci USA 94(5):2056–2061.