# COMMENTS

*We welcome comments by our readers reflecting agreement or disagreement with the material published in this section and, at the discretion of the Editor-in-Chief, will publish such comments. We would like to emphasize that material published anywhere in EBM, including this section, does not necessarily reflect the opinions of the Society or the Editor-in-Chief.*

# Comments to the Editor Concerning the Paper Entitled "Reproductive Malformation of the Male Offspring Following Maternal Exposure to Estrogenic Chemicals" by C. Gupta

BARBARA A. ELSWICK,[1] FREDERICK J. MILLER, AND FRANK WELSCH
*CIIT, Research Triangle Park, North Carolina 27709*

Potential effects on the endocrine system elicited by the weak estrogen bisphenol A (BPA) in laboratory rats and mice are highly controversial. The increased prostate gland weights in adult male offspring whose mothers were exposed to very low doses of BPA during pregnancy (1) were not found by other investigators (2, 3). An association of the low-dose effects on mouse prostate glands of BPA based on the effects of other much more potent estrogenic compounds with human prostatic disease has been postulated (4). Humans are exposed to exceedingly low levels of BPA from ingestion of canned foods and beverages where container linings contain BPA, raising the possibility that human health may be at risk. These discrepancies in animal study data require resolution based on the weight of scientific evidence, and, if deemed necessary, action needs to be taken to protect human health. To determine whether exposure to BPA and other weak estrogenic chemicals elicits detectable adverse biological effects in rodents, stringent study design and careful data analysis are necessary.

Examination of the data presented in a recently published paper by C. Gupta (5) raises concern regarding statistical analysis of the observations and the resulting conclusions. Several questions should be addressed so that the usefulness of the data in assessing the potential effects of BPA as an environmentally important chemical with weak estrogenic properties in male mouse offspring can be more definitively determined.

The data analysis section of Gupta's article (5) states that the data were tested by an analysis of variance (ANOVA). However, the descriptions are insufficient as to whether additional statistical analyses were performed. The reader is left wondering which post-hoc test (for example, Dunnett's, multiple *t*-tests, or Fisher's LSD) was used and whether corrections were made for multiple comparisons. The footnote in Table I states $P < 0.05$ (larger) or $P < 0.05$ (smaller) when compared with the vehicle control. Determining means and then performing a one-tailed, post-hoc statistical test based on whether a mean is larger or smaller introduces potential bias in the statistical analysis. A closer examination of the data in several other tables of the publication by Gupta (5) also suggests that the assumption of homogeneity of variance, which is a prerequisite for an ANOVA, is not met. We performed a Cochran's test for homogeneity of variances for the anogenital distance (AGD) measurements made on day 3 in Table I and the prostate size data in Table III, and the outcome revealed that the variances are not homogeneous ($\alpha = 0.01$). Therefore the question arises as to whether homogeneity testing was done and whether the data were transformed prior to ANOVA to account for a lack of homogeneous variances among treatment means.

There are additional concerns about the data analysis. When a chemical is administered to pregnant dams, the dam has long been recognized to constitute the experimental unit, and this practice is commonly accepted in develop-

[1] To whom requests for reprints should be addressed at CIIT, PO Box 12137, Research Triangle Park, NC 27709. E-mail: elswick@ciit.org

0037-9727/01/2262-0074$15.00/0

mental toxicity studies (6, 7). Therefore, given the description that 15 dams per group were treated, the sample size on days 3 and 21 in Table I should be only 15, not 45 or 30, respectively. This oversight alone makes it impossible to draw valid conclusions from the data summary in Table I concerning estrogenic effects of BPA and the other chemicals on AGD. Additional biometry-related questions arise. If the effects of a chemical on AGD is to be measured in the same animal at three different time points, then a repeated-measures ANOVA is the appropriate statistical analysis of the data in Table I in view of the correlation among the observations on a given animal (8). Additionally, if AGD is normalized by body weight, then using the cube root of the body weight eliminates the problem of overcorrection for body size that can occur if body weight alone is used (9, 10). However, this method may introduce errors if a linear relationship does not exist between the biological end point and body weight or body weight function (11, 12). An analysis of covariance (ANCOVA) with body weight as a covariate is a better way to account for differences in AGD due to body weight (12, 10), and a nested ANCOVA (dam within treatment) will provide an even more precise measure. Another consideration in reviewing the data in Gupta's article (5) is whether sampling one pup per litter provides a reliable estimate for the same sex littermates, especially on postnatal day 60. The AGD of individuals within a given rodent litter is affected by the sex of the neighboring fetus or fetuses in the uterine horn (13–15). Therefore variability among individuals within a litter is likely, and sampling only one pup from a given litter may not provide a reliable estimate of AGD for the same sex littermates. Whether sampling an organ weight from one offspring per sex per litter provides a reliable estimate for the litter mean is also questionable. Simulated statistical random sampling using fresh weights of the ventral prostate lobe of one adult male per litter from control and chemical-exposed rat offspring showed there was a high likelihood of reaching false conclusions concerning treatment-induced effects when considerable intralitter variability for the measured end point exists (16). Thus Gupta's sampling approach could be a potential confounder in the AGD data in Table I and organ weight data in Table II. While we could have used the technique of Larson (17) to perform an ANOVA from the summary statistics shown in the article (5), we feel the author should be able to respond to our concerns and provide more details about those data. Further clarification will help to better assess the relevance of Gupta's observations in the evaluation of effects of prenatally induced estrogenic chemicals in mice.


1. Nagel SC, vom Saal FS, Thayer KA, Dhar MG, Boechler M, Welshons WV. 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**(1):70–76, 1997.
2. Ashby J, Tinwell H, Haseman J. Lack of effects for low dose levels of bisphenol A and diethylstilbestrol on the prostate gland of CF1 mice exposed in utero. Regul Toxicol Pharmacol **30**:156–166, 1999.
3. Cagen SZ, Waechter JM Jr, Dimond SS, Breslin WJ, Butala JH, Jekat FW, Joiner RL, Shiotsuka RN, Veenstra GE, Harris LR. Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A. Toxicol Sci **50**:36–44, 1999.
4. vom Saal FS, Timms BG, Montano MM, Palanza P, Thayer KA, Nagel SC, Minati DD, Ganjam VK, Parmigiani S, Welshons WV. 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**:2056–2061, 1997.
5. Gupta C. Reproductive malformation of the male offspring following maternal exposure to estrogenic chemicals. Proc Soc Exp Biol Med **224**:61–68, 2000.
6. Weil CS. Selection of the valid number of sampling units and a consideration of their combination in toxicological studies involving reproduction, teratogenesis or carcinogensis. Fd Cosmet Toxicol **8**:177–182, 1970.
7. Haseman JK, Hogan MD. Selection of the experimental unit in teratology studies. Teratology **12**(2):165–171, 1975.
8. Tamura RN, Buelke-Sam J. The use of repeated measures analysis in developmental toxicology studies. Neurotox Teratol **14**:205–210, 1992.
9. Clarke RL. Endpoints of reproductive system development. In: Daston G, Kimmel C, Eds. An Evaluation and Interpretation of Reproductive Endpoints for Human Health Risk Assessment. Washington, DC: ILSI Press. pp10–27, 1998.
10. Gallavan RH Jr, Holson JF, Stump DG, Knapp JF, Reynolds VL. Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. Reprod Toxicol **13**:383–390, 1999.
11. Shirley E. The analysis of organ weight data. Toxicology **8**:13–22, 1977.
12. Shirley E, Newnham P. The choice between analysis of variance and analysis of covariance with special reference to the analysis of organ weights in toxicology studies. Stat Med **3**:85–91, 1984.
13. Meisel R, Ward I. Fetal female rats are masculinized by male littermates located caudally in the uterus. Science **213**:239–242, 1981.
14. vom Saal FS, Bronson FH. In utero proximity of female mouse fetuses to males: Effect on reproductive performance during later life. Biol Reprod **19**:842–853, 1978.
15. Vandendbergh JG, Huggett CL. The anogenital distance index, a predictor of the intrauterine position effects on reproduction in female house mice. Lab Anim Sci **45**(5):567–573, 1995.
16. Elswick BA, Welsch F, Janszen DB. Effect of different sampling designs on outcome of endocrine disruptor studies. Reprod Toxicol **14**:359–367, 2000.
17. Larson DA. Analysis of variance with just summary statistics as input. Amer Statist **46**:151–152, 1993.