

REFERENCES AND NOTES

1. NTP. Toxicology and Carcinogenesis Studies of 60-Hz Magnetic Fields in F344/N Rats and B6C3F1 Mice (Whole-Body Exposure Studies)-Draft. Research Triangle Park, NC:National Toxicology Program, 1998. <http://ehis.niehs.nih.gov/ntp/docs/tr488/488full.pdf> [cited 11 March 1998].
2. Frey AH, ed. On the Nature of Electromagnetic Field Interactions with Biological Systems. Austin, TX:R.G. Landes Co., 1994.
3. Portier CJ, Wolfe MS, eds. EMF Science Review Symposium Breakout Group Reports for Theoretical Mechanisms and *In Vitro* Research Findings, 24–27 March 1997, Durham, NC. Research Triangle Park, NC:National Institute of Environmental Health Sciences, 1998.
4. Frey AH. Is a toxicology model appropriate as a guide for biological research with electromagnetic fields? *J Bioelectr* 9 (2):233–234 (1990).
5. NTP. Studies of Magnetic Field Promotion (DMBA Initiation) in Sprague-Dawley Rats (Gavage/Whole-Body Exposure Studies)-Draft. Research Triangle Park, NC:National Toxicology Program, 1998. <http://ehis.niehs.nih.gov/ntp/docs/tr489/489full.pdf> [cited 11 March 1998].

The Importance of Protocol Design and Data Reporting to Research on Endocrine Disruption

Several recent articles have discussed the doubts expressed by some scientists regarding the validity of endocrine disruption studies conducted by industrial scientists or sponsored by the chemical industry (1–3). The last of these articles (3) recounted personal attacks made on the integrity of Stephen Safe of Texas A&M University.

It is in the nature of any new branch of toxicology that, at least initially, adverse effects may be discovered for chemicals by those academic laboratories working in the new area. The chemical industry is then left to confirm and extend the findings of others. Such confirmatory studies are usually necessary because the initial publications often describe the results of limited or unreplicated experiments (4). The articles mentioned above (1–3) concerned the prospect that repeat studies conducted by or sponsored by the chemical industry are designed in order not to confirm the original observation. We wish to discuss the complementary concern that many new findings in this area are either inadequately described or are based on data derived using inadequate test protocols. This makes it difficult to conduct faithful repeat experiments, however well-motivated the responsible scientists are.

We recently decided to confirm and extend adverse endocrine toxicities reported for nonylphenol (NP) and bisphenol-A (BPA). For both of these chemicals, we experienced problems when attempting to design repeat experiments due to inadequacies of the original publications. These inadequacies may seem to be relatively minor, but when the

outcomes of the repeat experiments are likely to be challenged, they become important.

Three influential papers using the Noble rat have been published by Colerangle and Roy over the past 4 years (5–7). The papers in question report the results of implanting the estrogens estrone, diethylstilbestrol (DES), NP, or BPA into Noble rats and monitoring the consequent changes in cell growth in the mammary gland. Either pellets or mini-pumps were used to deliver the test chemicals over 11 days. In each case growth of the mammary gland was reported. A significant aspect of these results is that estrogenic effects were found for NP and BPA at much lower dose levels than would have been expected based on the results of earlier studies (8–11), in particular, rat uterotrophic assays conducted using three daily administrations of the test chemicals. To resolve the uncertainties created by this apparent difference in assay sensitivities, we embarked on full repeats of the DES and NP Noble rat mammary gland assays (5–7). We also conducted rat uterotrophic assays utilizing the dosing protocol used by Colerangle and Roy [the test compound administered over 11 days via subcutaneously implanted mini-pumps (5–7)] and multiple strains of rats, including the Noble strain.

The following inadequacies of the three published studies in Noble rats have complicated their interpretation and the design of our own studies (5–7).

In the first study (5), DES was administered over 11 days as a subcutaneously implanted pellet. Cell labeling indices and growth fractions of mammary gland cells were determined using the methods of Foley et al. (12). DES was reported to increase the labeling index from 11% (controls) to 71%. Recalculation (12) of these indices from the primary data presented (5) gave values of 11% and 21% for controls and DES, respectively. Likewise, the growth fractions were reported to be 21% for controls and 158% for the DES animals; recalculation (12) from the primary data presented (5) gave values of 21% and 47%, respectively. The estrone figures were also in error. These errors, which have not been formally corrected by the authors, make it difficult to be certain of the magnitude of the effects expected in our repeat experiments. Subsequent data from these authors (6,7) appear to have been correctly calculated based on cell number estimates derived from the bar charts presented.

In the second and third papers (6,7), the activities of NP and BPA in the mammary gland of Noble rats were compared to that of DES. DES was shown as a positive control agent in both of these papers, and in each case the test data were identical to those reported to the original study (5), including use of the incorrect labeling indices and

growth fractions. The wrong impression was thereby given that the DES study had been replicated three times. Further, in the BPA paper (7), the DES is described as being administered via a mini-pump, whereas in the initial paper (5) it is reported to have been given as a pellet. No experimental details were provided for the administration of DES in the NP paper (6). Thus, after three separate publications, the test data for DES have apparently yet to be replicated.

Despite being published separately and a year apart, the vehicle control data for the NP (6) and the BPA (7) studies are the same in each paper and different from those in the original study (5). Either the data for NP and BPA were derived from a single study that was then published in two isolated parts or a vehicle control group was absent from one of the two studies (6,7). This created an unacceptable level of uncertainty.

Thus, while attempting to repeat these significant new findings in the Noble rat, we were presented with uncertainties in the original papers that could have been regarded as intentional had they occurred in our own (industrial) studies.

Nagel et al. (13) reported that BPA increased the weight of the prostate gland in mature CF-1 mice exposed *in utero*. In a subsequent paper from the same laboratory, vom Saal et al. (14) reported the induction of similar effects by DES. When designing a repeat of the CF-1 mouse experiment with BPA, we decided to include DES as a positive control chemical, despite the absence of such a control in the original BPA study (13). However, we were presented with a problem: the BPA animals described by Nagel et al. (13) were terminated at 6 months and the DES animals described by vom Saal et al. (14) were terminated at 8 months. No explanation for this difference in test protocol was given. It was therefore impossible to mount a faithful concurrent report of the BPA and DES experiments; thus, we decided to terminate both of our groups at 6 months. However, that means that we will not have faithfully repeated the original study on DES.

Another aspect of the study by Nagel et al. (13) caused us concern. In that study, two control groups were used: a vehicle control group and a group of animals that were not handled throughout the study. It was stated (13) that these two control groups gave similar data (not shown) and that they were therefore combined into a single, larger control group and used as such for the subsequent statistical analysis of the BPA test data. That represents bad statistical practice. We decided to include two such control groups in our own experiment and to maintain their separate identities during the statistical analysis of our data. Each of these small changes in

our experimentation could eventually be used to discredit our findings, should they happen not to agree with the original observations.

It seems important that all experiments in the rapidly expanding area of endocrine disruption toxicology should be carefully designed and fully reported. The use of concurrent positive and negative control groups also seems to be prudent. These needs are independent of who conducts or sponsors studies. Good science is good science. Finally, it should be noted that the only formal retraction of endocrine disruption data currently encountered derived from an academic laboratory (15), a salutary counterbalance to the assertions that stimulated this letter (1-3).

John Ashby
Jenny Odum

Zeneca Central Toxicology Laboratory
Alderley Park, Cheshire, United Kingdom

REFERENCES AND NOTES

1. Anonymous. Industry and scientists in cross-fire on endocrine-disrupting chemicals. ENDS Reports 268:26-29 (1997).
2. Anonymous. Controversy over bisphenol A research. Endocrine/Estrogen Newsletter 3(12):1-4 (1997).
3. Anonymous. Industry funding provokes controversy. Endocrine/Estrogen Letter 4(4):5-6 (1998).
4. Ashby J, Elliott HM. Reproducibility of endocrine disruption data. Regul Toxicol Pharmacol 26:94-95 (1997).
5. Colerangle JB, Roy D. Perturbation of cell cycle kinetics in the mammary gland by stilbene estrogen, diethylstilbestrol (DES). Cancer Lett 94:55-63 (1995).
6. Colerangle JB, Roy D. Exposure of environmental estrogenic compound nonylphenol to Noble rats alters cell-cycle kinetics in the mammary gland. Endocrine 4:115-122 (1996).
7. Colerangle JB, Roy D. Profound effects of the weak environmental estrogen-like chemical bisphenol A on the growth of the mammary gland of Noble rats. J Steroid Biochem Mol Biol 60:153-160 (1997).
8. Lee PC, Lee W. In vivo estrogenic action of nonylphenol in immature female rats. Bull Environ Contam Toxicol 57:341-348 (1996).
9. Odum J, Lefevre PA, Tittensor S, Paton D, Routledge EJ, Beresford NA, Sumpter JP, Ashby J. The rodent uterotrophic assay: critical protocol features, studies with nonylphenols, comparison with a yeast estrogenicity assay. Regul Toxicol Pharmacol 25:176-188 (1997).
10. Ashby J, Tinwell H. Uterotrophic activity of bisphenol-A to the immature rat. (Submitted).
11. Dodds EC, Lawson W. Synthetic oestrogenic agents without the phenanthrene nucleus. Nature 137:996 (1936).
12. Foley J, Ton T, Marenpot R, Butterworth B, Goldsworthy TI. Comparison of proliferating cell nuclear antigen to tritiated thymidine as a marker of proliferating hepatocytes in rats. Environ Health Perspect 101(suppl 5):199-206 (1993).
13. 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 105:70-76 (1997).
14. vom Saal FS, Timms BG, Montano MM, Palanza P, Thayer KA, Nagel SC, Dhar MD, Gunjam 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 dose levels. Proc Natl Acad Sci USA 94:2056-2061 (1997).
15. McLachlan JA. Synergistic effect of environmental estrogens: report withdrawn. Science 277:462-463 (1997).

Response

In a paper we published last year (1), we described biological effects *in vivo* on the rodent prostate caused by fetal exposure to very low doses of the environmental estrogen bisphenol A; this low-dose effect was predicted by a new *in vitro* assay. For the *in vivo* end point of prostate enlargement, the effect produced by bisphenol A mimicked the effect of fetal exposure to low doses of the natural and synthetic estrogens estradiol and DES, which were reported in another paper (2). Fetuses were exposed to bisphenol A by feeding pregnant female mice at average maternal doses of 2 and 20 $\mu\text{g}/\text{kg}$ maternal body weight per day (2 and 20 ppb); these exposure levels produced enlarged prostates measured in subsequent adulthood. Our conclusion was that these doses of bisphenol A, up to 25,000 times lower than the previously reported NOAEL (no observed adverse effect level) for bisphenol A (3), were near and within reported ranges of current human exposures from different sources of this chemical (4,5). Three subsequent reports by two other groups have confirmed our finding of high estrogenic bioactivity of bisphenol A *in vivo* using end points (pituitary and mammary gland responses) that were different from ours (6-8).

We find perplexing the statement of Ashby and Odum that "many new findings in this area are either inadequately described or are based on inadequate test protocols. This makes it difficult to conduct faithful repeat experiments." The information that went into our experimental design is based on more than 50 years of combined experience in hormone action and control of development. It is impossible to put all of this information in any one paper, and experimental details that have been published previously are typically not repeated [for example, see (9,10)]. For these reasons, when we are interested in replicating an experiment, we contact the original authors, and other scientists have often contacted us for the same reason. For example, Ashby has contacted us on numerous occasions concerning experimental procedures for the replication of our studies. In addition, we recently ran a training session for laboratory personnel from a contract laboratory hired by the Society of the Plastics Industry to replicate our study with bisphenol A. Given this degree of cooperation with Ashby and others associated with the chemical industry, which is also true for Richard Sharpe (11), we are puzzled as to why Ashby and Odum would make the above statement. Considering the many questions they raise above in understanding the procedures of Colerangle and Roy (8), we would hope that they would also have contacted the original authors in that study.

Ashby and Odum also raised two specific questions about our studies (1,2). The first question concerned examination of prostate

weight at 8 months of age in one study with prenatal exposure to estradiol and DES, while bisphenol A-exposed animals were examined at 6 months old of age. We had conducted a preliminary study comparing prostate weight in control CF-1 male mice (five to nine males/group) at 6, 7, 8, 9.5, and 12.5 months of age, which resulted in the following mean (\pm standard error) prostate weights (in milligrams): 42.1 ± 2.5 , 40.8 ± 2.7 , 45.1 ± 3.8 , 41.1 ± 2.8 , and 61.3 ± 2.8 , respectively. These unpublished findings showed that between 9 and 12 months of age, male CF-1 mice experienced a significant increase in prostate weight, but between 6 and 9 months of age, there was no significant difference in prostate weight. We had initially waited until males were 8 months old to examine effects of prenatal treatment with estradiol and DES on the prostate due to concern that effects might only be seen in middle age (12). However, we have sought to reduce the age at organ collection in these studies to reduce costs. Relative to control males, an increase in prostate weight was seen at 6 months of age in the bisphenol A study and, more recently, was also found in 50-day-old CF-1 male mice exposed prenatally to low doses of ethinyl estradiol (13).

The second technical question concerned the combination of vehicle control and unhandled control animals into a single control group in our studies. In all of our experiments we conduct an initial analysis just with these two control groups. In every study that we have conducted, this initial analysis has revealed no statistical difference between the two groups (the F value was 0.7 and $p > 0.4$ for this comparison in the bisphenol A study on prostate weight); these animals were then combined into one control group for comparison to chemical treatment groups. Ashby and Odum state, "that represents bad statistical practice." However, an initial comparison of multiple control groups is a common and appropriate procedure, although from some perspectives, there would be a decided advantage in not taking this approach. Specifically, the F ratio in analysis of variance is calculated as the product of variation between groups divided by variation within groups. The greater the number of groups with the same mean that are placed into an analysis of variance, the greater the reduction in the F ratio, and therefore the greater the probability of failing to find statistical significance. The procedure recommended by Ashby and Odum would thus increase the likelihood of falsely concluding that the test chemical had no effect.

The initial point made by Ashby and Odum involves the discovery of adverse effects for chemicals by academic laboratories and that the chemical industry is left to confirm unreplicated findings. It seems inappropriate to com-