Human and Ecological Risk Assessment: Vol. 5, No. 6, pp. 1255-1262 (1999)

# LETTER TO THE EDITOR

## Beware Missing Data And Undernourished Statistical Models: Comment On Fairbrother *et al.'s* Critical Evaluation

Joseph P. Skorupa
Division of Environmental Contaminants, U.S. Fish and Wildlife Service, Sacramento, CA*

Editors' Note: This letter follows from prepublication circulation by the editors to interested parties of the papers in this volume. It is clear from this letter and the papers, that selenium is an extremely controversial subject. It is equally clear that the issues raised in this letter and other contentious issues will only be resolved by well-designed studies that, ideally, are endorsed by key scientists on both "sides" of the issues. We hope that publishing these papers, this letter, and any subsequent letters will assist in the resolution of outstanding issues.

### CONVENTIONAL WISDOM

During the mid-1980s it was discovered at California's Kesterson Reservoir that unnaturally high dietary exposure to selenium (Se) had caused substantive reproductive toxicity to aquatic birds (*e.g.,* Ohlendorf, 1989). That discovery spawned a broad research program on Se sponsored by the Federal Government (*e.g.,* Dinar and Zilberman 1991), including both laboratory and field components. One objective of the federal research program was to investigate the threshold exposures associated with avian reproductive toxicity. The program's controlled laboratory studies were most recently summa-

* 2800 Cottage Way, Rm. W-2605, Sacramento, CA 95825-1846; Tel: (916) 414–6593; Fax: (916) 414–6712; E-mail: Joseph_Skorupa@fws.gov

rized by Heinz (1996). The program's field studies were most recently summarized by Skorupa (1998a).

Heinz (1996) concluded that Se content of avian eggs is a sensitive predictor of reproductive toxicity, and that the toxicity threshold was about 3 parts per million [ppm] (whole egg, wet weight [ww]; equivalent to about 10 ppm dry weight [dw] given the 70% moisture reported by Heinz *et al.,* 1989). Heinz based his conclusion on multiple lines of evidence, including that 3.4 ppm ww was the highest 'no effect' concentration of Se in mallard eggs and a value between 2.73 and 2.95 ppm ww was the lowest 'effect' concentration for chicken eggs. Heinz cautioned readers that the 3 ppm ww threshold should not be confused with a 'safe' level for regulatory purposes because it did not incorporate any margin of safety to account for uncertainties (such as the very limited number of species tested). Additional caution is warranted because at 3.4 ppm Se ww, the hatchability (viability) of fertile mallard eggs was depressed by 10.4% relative to controls (Heinz *et al.*, 1989). Due to insufficient statistical power, it is impossible to know whether 3.4 ppm Se ww was a true 'no effect' level or was in fact an approximation of the $EC_{10.}$

Based on a graphical plot of field data for black-necked stilts (a large shorebird), Skorupa (1998a) concluded that the incidence of nests containing inviable eggs showed a response threshold when Se concentrations in randomly collected sample eggs were >6 ppm dw, but <7 ppm dw. Initial analysis of 184 randomly sampled stilt nests with sibling eggs monitored to full term (successful hatching of at least one egg) suggested a response threshold at an egg concentration between about 4 to 10 ppm Se dw (Ohlendorf *et al.*, 1993; Skorupa, 1994). By 1998, data for >400 randomly sampled stilt nests were available, including >100 nests whose random sample egg fell within the critical threshold region of 4 to 10 ppm Se dw. This allowed plotting of response data at single ppm intervals in the critical threshold region and a very precise delineation of the 6 ppm dw threshold point (Skorupa, 1998a). Between 6 and 7 ppm Se dw in stilt eggs the nestwise incidence of inviable eggs jumps from a 'background' range of about 6 to 9% to about 16% and continuously increases in an exposure-responsive manner thereafter as egg exposures increase above 7 ppm dw. The initial jump in the nestwise incidence of inviable eggs is approximately equivalent to an eggwise viability depression of 3% (based on 4-egg clutches and corrected for background). Thus, the 6 ppm dw response threshold reported by Skorupa (1998a) is approximately the $EC_{03}$ for viability of stilt eggs. Based on more than 600 randomly sampled stilt eggs containing developed embryos, Skorupa (1998a) reported a logistic regression equation for external overt terata that predicts an $EC_{03}$ for terata of 25 ppm Se dw. A 3% rate of overt terata is about 10 to 20 times above normal for stilts (Skorupa *et al.,* 1996). The comparative $EC_{03}$ values, 6 vs. 25 ppm dw, indicate that egg viability is a more sensitive endpoint than overt terata.

## A CRITICAL REVIEW OF CONVENTIONAL WISDOM

Fairbrother *et al.* (1999) reexamined the above federal research results and concluded (in part) that Skorupa's field data for viability of stilt eggs as a function of Se exposure are too unreliable (confounded by non-chemical field effects) to be utilized for risk assessment, and have not been sufficiently peer reviewed. Thus, Fairbrother *et al.* recommend that regulators disregard the stilt data, even though it is the single most extensive existing set of exposure-response data for Se.

By fitting data points from two of the government's controlled laboratory studies (Heinz *et al.*, 1989; Stanley *et al.*, 1994) to logit functions Fairbrother *et al.* estimate that the $EC_{10}$ and $EC_{20}$ for mallard duckling production (a composite of egg fertility, viability of fertile eggs, and early post-hatch duckling survivorship) are 16 and 21 ppm Se dw respectively (whole egg concentrations). Fairbrother *et al.* recommend that these substantially higher values, compared with the Heinz (10 ppm dw) and Skorupa (6 ppm dw) response thresholds, be utilized for developing regulatory thresholds.

Finally, by fitting both teratogenicity and duckling production data from Heinz *et al.* (1989) and Stanley *et al.* (1994) to a 4-parameter full logit model and to various reduced-parameter partial models, Fairbrother *et al.* conclude that no difference in response sensitivity can be demonstrated statistically between the endpoints of duckling production (related to egg viability) and teratogenicity.

## ARE THE STILT DATA CONFOUNDED BY NON-CHEMICAL EFFECTS?

Fairbrother *et al.* generically suggest that factors such as disease, climate, parental nutrition, and other contaminants may confound field measures of egg viability. About 95% of the >400 stilt nests monitored occurred within California's San Joaquin Valley during a period when there were no major epizootic events substantively involving black-necked stilts (California Department of Fish & Game file reports). The valley floor is also relatively homogeneous climatologically (Preston, 1981). There is also no evidence of confounding effects from other contaminants (*e.g.*, Moore *et al.*, 1990; Ohlendorf *et al.*, 1993). Additionally, the graphical approach employed by Skorupa (1998a) explicitly illustrates the level of potentially confounding background noise (6 to 9% background response rate) as well as explicitly illustrating any response threshold that clearly rises above background noise. If background noise were significantly confounding the data, it should weaken the power to illustrate sensitive threshold points, not spuriously increase sensitivity. Finally, in the early years of data collection (1988 to 1989), and in collaboration with Dr. Ursula Abbott (U.C. Davis) and her then graduate student Paul Martin (whose unpublished graduate thesis is cited by Fairbrother *et al.*), field estimates of egg viability were compared to estimates from artificial incubation of field-collected fresh eggs (Skorupa *et al.*, 1989). The results from field (uncontrolled external environment) and hatchery (controlled external environment) were nearly identical for each of three study sites (95 vs. 93%; 96 vs. 98%; and 79 vs. 74% hatchability for field vs. hatchery, respectively). These results indicate that uncontrolled environmental factors external to the egg were not influencing field measures of egg viability.

## HAVE THE DATA BEEN SUFFICIENTLY PEER REVIEWED?

Fairbrother *et al.* cite the internal Fish and Wildlife Service report by Skorupa *et al.* (1996), but fail to cite the same data from the extensively peer-reviewed follow-up publication of the National Irrigation Water Quality Program [NIWQP] (Skorupa, 1998b). Skorupa (1998b), where the basic field datasets are presented as they existed in 1996, was peer-reviewed by more than 10 peers and was in review for more than 2 years (a list of peer reviewers is presented on p. 6 of Martin and Larsen, 1998). Elsewhere, Fairbrother *et al.* (different order of authorship however, *e.g.*, Adams *et al.*, 1998) rely heavily on data from NIWQP publications, so presumably Fairbrother and coauthors view the NIWQP peer-review process as sufficiently reliable. Fairbrother *et al.* are also curiously willing to reject the peer-reviewed finding of a very low response threshold published by Ohlendorf *et al.* (1986; embryotoxicity $EC_{20}$ of 5 ppm Se dw for stilt eggs) citing the unpeer-reviewed Skorupa *et al.* (1996) as overriding justification. The issue of peer-review is, in fact, a red herring argument raised by Fairbrother *et al.* The mainstream peer-reviewed results of research summarized by Heinz (1996) have substantively the same regulatory implications as Skorupa's stilt field data (*i.e.*, that 'safe' Se exposure for eggs is <10 ppm Se dw), so there are no regulatory decisions that must rely solely on the stilt data.

## WHAT THRESHOLD POINTS DO LAB STUDIES REALLY SUPPORT?

Fairbrother *et al.* assert that lab studies of mallards support an estimated $EC_{10}$ of 16 ppm Se dw. However, there are several serious flaws in their analysis. The most important flaw is the missing data. Fairbrother *et al.* derived the 16 ppm threshold point by fitting 6 treatment data points garnered from Heinz *et al.* (1989) and Stanley *et al.* (1994) to a logit function that is forced through the origin by normalizing control responses to zero. However, an additional treatment data point from Heinz *et al.* (1987) is intentionally discarded as anomalous, and two additional treatment data points presented by Stanley *et al.* (1996) are excluded without comment. Thus, only 6 of 9 available treatment data points are employed for Fairbrother *et al.*'s critical evaluation. What difference do the three missing data points make? Based on the complete 9-point set of laboratory results, the estimated logistic $EC_{10}$ is 7.9 ppm Se dw, much lower than the 6-point estimate of 16 ppm.

The logistic function is very sensitive to the data point from Heinz *et al.* (1987) that Fairbrother *et al.* explicitly discarded. They discarded that data point because the diet-to-egg bioaccumulation factor [BAF] was deemed anomalous. However, the logistic function does not model bioaccumulation, it models response to measured exposures independent of the BAF that produced the measured exposure. Thus, it is not clear why a data point should be excluded because of an anomalous BAF. There are many between-study inconsistencies in the lab results, not just the low BAF for the Heinz *et al.* (1987) data point. As Heinz (1996) noted there is a tremendous amount of individual variability among mallards in their sensitivity to a given Se treatment level and in the degree to which individuals may avoid treated diets (creating BAF variability), such that it cannot be expected that replicated treatments of small groups will necessarily be in any better than rough agreement. Under such circumstances it is at best a dubious ride down a slippery slope to start practicing *post hoc* data cleansing.

Another important flaw is that Fairbrother *et al.* pool fundamentally incompatible response endpoints. Fairbrother *et al.* noted the high inconsistency of their response variable (depression of duckling production) where 37 ppm dw in eggs was associated with 51% depression (Heinz *et al.*, 1989) and the nearly identical exposure of 42 ppm dw was associated with 99% depression of duckling production (Stanley *et al.*, 1994). However, Fairbrother *et al.* don't seem to be aware that they are comparing two qualitatively different endpoints. Most of the difference in duckling production is attributable to the 81% post-hatch duckling survivorship from the 37 ppm eggs compared to only 10% duckling survivorship from the 42 ppm eggs. This 8-fold difference at nearly equal egg Se levels is almost certainly due to the fact that in the former case ducklings *were immediately put on a clean diet upon hatching*, whereas in the latter case ducklings *were left on the same treatment diets their mothers were exposed to*. Clearly the more sensitive result of Stanley *et al.* (99% depression at 42 ppm egg Se) most realistically mimics the real world where ducklings are indeed exposed to the same contaminated environment as their mothers. Heinz *et al.* (1987), Stanley *et al.* (1994), and Stanley *et al.* (1996) all fed ducklings treated diets; only Heinz *et al.* (1989) fed ducklings clean diets. Thus, for Fairbrother *et al.*'s analysis, the Heinz *et al.* (1989) data points should not have been pooled with the other studies. Excluding the Heinz *et al.* (1989) data points, the estimated logistic $EC_{10}$ for duckling production is 6.9 ppm Se dw.

## IS EGG INVIABILITY MORE SENSITIVE THAN TERATOGENICITY?

Fairbrother *et al.* report failing to find statistical support for such a finding. That is not surprising, because in their full logit model they are estimating 4 parameters (coefficients $b_0$ - $b_3$) from just 6 independent data points. Such a data-starved statistical model has little discriminatory power.

To answer the above question, one need not resort to inappropriately sophisticated statistics. One merely needs to examine the lab data for mallards:

| Dietary Treatment | Control-adjusted teratogenesis | Control-adjusted egg inviability | Source |
|---|---|---|---|
| 7 ppm dw | 0% | 34% | S-96 |
| 8 ppm dw | 4% | 38% | H-89 |
| 10 ppm dw | 12% | 53% | H-87 |
| 10 ppm dw | 53% | 91% | S-94 |
| 16 ppm dw | 66% | 96% | H-89 |

*Note:* Sources above are Heinz *et al.*, 1987; Heinz *et al.*, 1989; Stanley *et al.*, 1994; Stanley *et al.*, 1996.

If inviability and teratogenicity were truly equally sensitive endpoints the odds of percent inviability exceeding percent teratogenesis at any given treatment level would be 50:50. However, inviability clearly exceeds teratogenesis in all 5 cases. For equally sensitive endpoints that outcome has an expected probability of $(0.5)^5$, or 0.03. Since the probability of the observed outcome is less than 0.05, the assertion that inviability and teratogenicity are equally sensitive endpoints is statistically rejected.

## SUMMARY

Fairbrother *et al.* (1999) do not provide scientifically credible evidence for avian response thresholds significantly different than the thresholds summarized by Heinz (1996) and Skorupa (1998). If the reproductive $EC_{10}$ truly were as high as 16 ppm dw, then the biological appropriateness of an $EC_{10}$ as a regulatory criterion would be open to question because 16 ppm dw is >3 times the dietary threshold for toxicity of 5 ppm dw recognized by Fairbrother *et al.*, and thus eggs with 16 ppm Se dw would pose an unacceptable secondary hazard to egg-predators. Fairbrother *et al.*'s recommendation for additional lab testing is worthy of support, especially if such studies are designed with sufficient statistical power to resolve treatment effects as low as 5 to 10% difference from controls. Past lab studies were unable to statistically detect treatment effects lower than about 40% difference from controls.

## REFERENCES

Adams, W. J., Brix, K. V., Cothern, K. A., Tear, L. M., Cardwell, R. D., Fairbrother, A., and Toll, J. 1998. Assessment of selenium food chain transfer and critical exposure factors for avian wildlife species: need for site-specific data. In: *Environmental Toxicology and Risk Assessment: Seventh Volume. ASTM STP 1333* (Little, E. E., DeLonay,

A. J., and Greenberg, B. M., Eds.). American Society for Testing and Materials, Philadelphia, PA.

Dinar, A. and Zilberman, D. 1991. *The Economics and Management of Water and Drainage in Agriculture.* Kluwer Academic Publishers, Boston, MA. 946 p.

Fairbrother, A., Brix, K. V., Toll, J. E., McKay, S., and Adams, W. J. 1999. A critical evaluation of egg selenium concentrations as predictors of avian toxicity. *Health Ecol. Risk Assessment* 5(6), 1229–1253.

Heinz, G. H., Hoffman, D. J., Krynitsky, A. J., and Weller, D. M. G. 1987. Reproduction in mallards fed selenium. *Environ. Toxicol. Chem.* 6, 423–433.

Heinz, G. H., Hoffman, D. J., and Gold, L. G. 1989. Impaired reproduction of mallards fed an organic form of selenium. *J. Wildl. Manage.* 53, 418–428.

Heinz, G. H. 1996. Selenium in birds. In: *Environmental Contaminants in Wildlife: Interpreting Tissue Concentrations*. pp. 453–464. (Beyer, W. N., Heinz, G. H., and Redmon-Norwood, A. W., Eds.). CRC Press, Lewis Publishers, Boca Raton, FL.

Martin, P. L. and Larsen, D. E. 1998. *Guidelines for Interpretation of the Biological Effects of Selected Constituents in Biota, Water, and Sediment.* National Irrigation Water Quality Program Information Report No. 3. U.S. Department of Interior, Denver, CO. 198 p.

Moore, S. B., Winckel, J., Detwiler, S. J., Klasing, S. A., Gaul, P. A., Kanim, N. R., Kesser, B. E., DeBevec, A. B., Beardsley, K., and Puckett, L. K. 1990. *Fish and Wildlife Resources and Agricultural Drainage in the San Joaquin Valley, California* (two volumes). San Joaquin Valley Drainage Program, U.S. Bureau of Reclamation and California Dept. of Water Resources, Sacramento, CA.

Ohlendorf, H. M., Hothem, R. L., Bunck, C. M., Aldrich, T. W., and Moore, J. F. 1986. Relationships between selenium concentrations and avian reproduction. *Trans. N. Am. Wildl. Nat. Resour. Conf*. 51, 330–342.

Ohlendorf, H. M. 1989. Bioaccumulation and effects of selenium in wildlife. In: *Selenium in Agriculture and the Environment.* pp. 133–177. (Jacobs, L. W., Ed.). Am. Soc. Agronomy, and Soil Sci. Soc. of America, Madison, WI.

Ohlendorf, H. M., Skorupa, J. P., Saiki, M. K., and Barnum, D. A. 1993. Food-chain transfer of trace elements to wildlife. In: *Management of Irrigation and Drainage Systems: Integrated Perspectives*. pp. 596–603. (Allen, R. G. and Neale, C. M. U., Eds.). Am. Soc. Civil Eng., NY.

Preston, W. L. 1981. *Vanishing Landscapes: Land and Life in the Tulare Lake Basin*. University of California Press, Berkeley, CA. 278 p.

Skorupa, J. P., Abbott, U. K., Martin, P., Roster, D. L., and Welsh, D. 1989. *Hatchery estimates of embryo viability among recurvirostrids: a potential biomonitoring technique for subsurface agricultural drainage ponds.* Abstract of platform presentation at the 1989 Annual Meeting of the Western Section of The Wildlife Society, January 26–28, 1989, Redding, CA.

Skorupa, J. P. 1994. Impacts of selenium on the biological systems of the Salton Sea. In: *Proceedings of the Salton Sea Symposium, January 13, 1994, Indian Wells, California*. Salton Sea Authority, Imperial, CA.

Skorupa, J. P., Mormon, S. P., and Sefchick-Edwards, J. S. 1996. *Guidelines for interpreting selenium exposures of biota associated with nonmarine aquatic habitats.* Report to the National Irrigation Water Quality Program (NIWQP). U.S. Fish and Wildlife Service, Sacramento, CA.

Skorupa, J. P. 1998a. Selenium poisoning of fish and wildlife in nature: lessons from twelve real-world examples. In: *Environmental Chemistry of Selenium*, pp. 315–354 (Frankenberger, W. T. and Engberg, R. A., Eds.). Marcel Dekker, Inc., NY.

Skorupa, J. P. 1998b. Selenium. In: *Guidelines for Interpretation of the Biological Effects of Selected Constituents in Biota, Water, and Sediment,* pp. 139–184 (Martin, P. L., and Larsen, D. E., Eds.). National Irrigation Water Quality Program Information Report No. 3. U.S. Department of Interior, Denver, CO.

Stanley, T. R., Jr., Spann, J. W., Smith, G. J., and Rosscoe, R. 1994. Main and interactive effects of arsenic and selenium on mallard reproduction and duckling growth and survival. *Arch. Environ. Contam. Toxicol.* 26, 444–451.

Stanley, T. R., Jr., Smith, G. J., Hoffman, D. J., Heinz, G. H., and Rosscoe, R. 1996. Effects of boron and selenium on mallard reproduction and duckling growth and survival. *Environ. Toxicol. Chem.* 15, 1124–1132.