Comment on the article by Hodges

Andrew Gelman, Department of Statistics, Columbia University, New York, NY 10027
Phillip N. Price, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

October 23, 1997

This paper makes the interesting and important contribution of viewing the pseudo-data expression of the hierarchical linear model as a unifying tool in model checking. (The analytical and computational use of the formulation for Bayesian inference is well known; see, e.g., Dempster, Rubin, and Tsutakawa, 1991, and Gelman et al., 1995, chapter 13.) More generally, the author notes that hierarchical regression models are more difficult to understand than we might imagine, especially if predictors appear at more than one level in the hierarchy. We would like to echo that point with a statistical anecdote of our own (see Price, Nero, and Gelman, 1996).

We fit a hierarchical linear model to the logarithms of home radon measurements in Minnesota, with random effects for the counties of measurement. The model thus had variance components at the within- and between-county levels, which we examined to get an idea of the model's precision. We then added an individual-level predictor—an indicator for whether each house had a basement. Adding the predictor caused the estimated (posterior median) within-county variance to decrease (as expected; houses with basements tend to have higher radon levels), but the estimated between-county variance substantially *increased*, which was completely unexpected. What was going on? After some thought, we realized that the counties with more basements happened to have higher random effects (in the second model). In the first model, much of the variation in county radon

levels was cancelled by an opposite variation in the proportion of basements. The increased between-county variance in the second model indicates true variation among counties that happened to be masked by the first model.

This sort of pattern, caused by correlation between individual-level variables and random effects, does not occur in non-hierarchical regression. We suspect that the tools developed in the paper under discussion will be useful in understanding this sort of data structure, and we look forward to future work in this area.

Finally, we disagree with the claim in Section 3.2.2 that "Bayesian diagnostics ... place an extra layer of mathematics between the analyst and the data." Posterior predictive checks—comparisons of observed data to their predictive distribution under an assumed model—can be performed graphically and, in fact, can be simpler to interpret than classical methods such as Studentized residual plots (see Rubin, 1984, and Gelman, Meng, and Stern, 1996).

References

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association* **76**, 341–353.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis.* London: Chapman and Hall.

Gelman, A., Meng, X. L., and Stern, H. S. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6**, 733–807.

Price, P. N., Nero, A. V., and Gelman, A. (1996). Bayesian prediction of mean indoor radon concentrations for Minnesota counties. *Health Physics* **71**, 922–936.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12**, 1151–1172.