Comment

Andrew GELMAN and Frederic Y. BOIS

This article is admirable in that it presents a general framework that allows the user to focus on the pharmacokinetic modeling rather than on the computational and statistical details of estimation. The authors set up a seemingly reasonable model and obtain a good fit to the pain-relief data with plausible parameter values, while accounting for the kind of complications that typically arise in clinical trials. We would like to focus on the following features of this work: (1) analysis of the missing data (dropouts), (2) the pharmacokinetic model, (3) checking the fit of the model, (4) modeling and display of individual variation, and (5) extrapolation to other dosing strategies.

1. ANALYSIS OF THE DROPOUTS

Although the analysis of dropouts appears to be a major complication in the analysis, and is certainly a problem of general interest in this field (see, e.g., Ten Have, Pulkstenis, Kunselman, and Landis 1997), it is, we believe, a minor concern in this particular example. As noted in the article, the pain-relief level just preceding dropout was 0 (no pain relief) in nearly all of the cases. Considering the design of the study (with dropouts allowed to switch to an effective analgesic), it seems reasonable to impute pain-relief scores of 0 for the responses that were unobserved because of dropout. Our Figure 1 shows the aggregate results for the completed dataset (including both observed and imputed responses). In addition to the imputation, we have altered Figure 1 of the article by separating the "dose = 0" respondents into three groups corresponding to the three different experiments. We believe that our Figure 1, with imputations included, gives a more direct picture of the pain-relief results of the experiment without being contaminated by the dropout behavior. For example, the diminishing of pain relief at later times is clear.

Because of the simplicity of the dropout behavior in this experiment (i.e., essentially all of the unobserved pain relief measurements can be assumed to be 0), we expect that fitting the pain-relief model directly to the completed dataset would yield nearly identical results to that of the full model fit to the censored dataset in the article.

But what about a more complicated scenario, perhaps including a substantial number of subjects who drop out early with nonzero pain-relief scores? In that case, we would recommend fitting the full model as described in the article, and then going back and imputing the censored responses based on their probabilities under the model conditional on the observed data. For example, if an unobserved pain relief score has predictive probabilities of .6, .3, and .1 of having the values 0, 1, or 2, then this would be imputed into the corrected Figure 1 as .6 observations in the pain relief index (PRL) = 0 total, .3 in the PRL = 1 total, and .1 in the PRL = 2 total.

In general, completed-data displays such as the corrected Figure 1, although not necessary for parameter estimation, could be quite useful in understanding the experimental results because the completed data show the best estimate of the pain-relief results had all subjects stayed with the assigned doses.

1.1 The Pharmacokinetic Model

We found Figures 2 and 3 in the article very helpful in understanding the pharmacokinetic model. In fact, it would have been nice to see more of the workings of the model; for example, it would be interesting to see a display of estimates of C, Ce, and $f_{\mathrm{PD}}(Ce)$, over time, for different patients. We also have a specific question about the model. Figure 3b shows the drug having no effect for the first hour. Is this contradicted by the bottom row of our Figure 1, which seems to show that patients with dose = 100 and 200 have more pain relief than the dose = 0 patients, even after times as short as .25 and .5 hour?

The pharmacokinetic model used in the article is somewhat complicated (at least compared to many standard models in statistics) and would be difficult, if not impossible, to fit solely from the given dataset. The authors deal with this problem by estimating only some of the parameters from the current data, while setting the other parameters to fixed values, based on results from other experiments. Keeping control over the parameters in this way is a key step in obtaining reasonable model estimates. Continuing in this direction, we expect further progress could be made by adapting a Bayesian approach with informative prior distributions on the parameters about which other experimental data are available (see, e.g., Bois, Gelman, Jiang, Maszle, and Alexeef 1996 and Gelman, Bois, and Jiang 1996 for an example from toxicokinetics).

From the substantive viewpoint of the ketorolac analysis, it would be useful to see the data used by the authors to get their pharmacokinetic parameter values, or at least the references to those data. Ketorolac kinetics are an interesting case, because the commercial product is in fact a mixture of two enantiomers—molecules of the same atomic composition but different symmetries—with different pharmacokinetic behavior (see Hayball, Wrobel, Tambly and Nation 1994; Vakily, Corrigan, and Jamali 1995). It would also be nice to see some reference in the article to the state-

Andrew Gelman is Associate Professor Department of Statistics, Columbia University, New York, NY 10027. Frederic Y. Bois is Staff Scientist, Lawrence Berkeley Laboratory, Berkeley, CA 94720. The authors thank the National Science Foundation for grant DMS-9404305 and Young Investigator award DMS-9457824; the Research Council of Katholieke Universiteit Leuven for fellowship F/96/9; and Lewis Sheiner for providing the data and model predictions for the ketorolac analysis.

© 1997 American Statistical Association

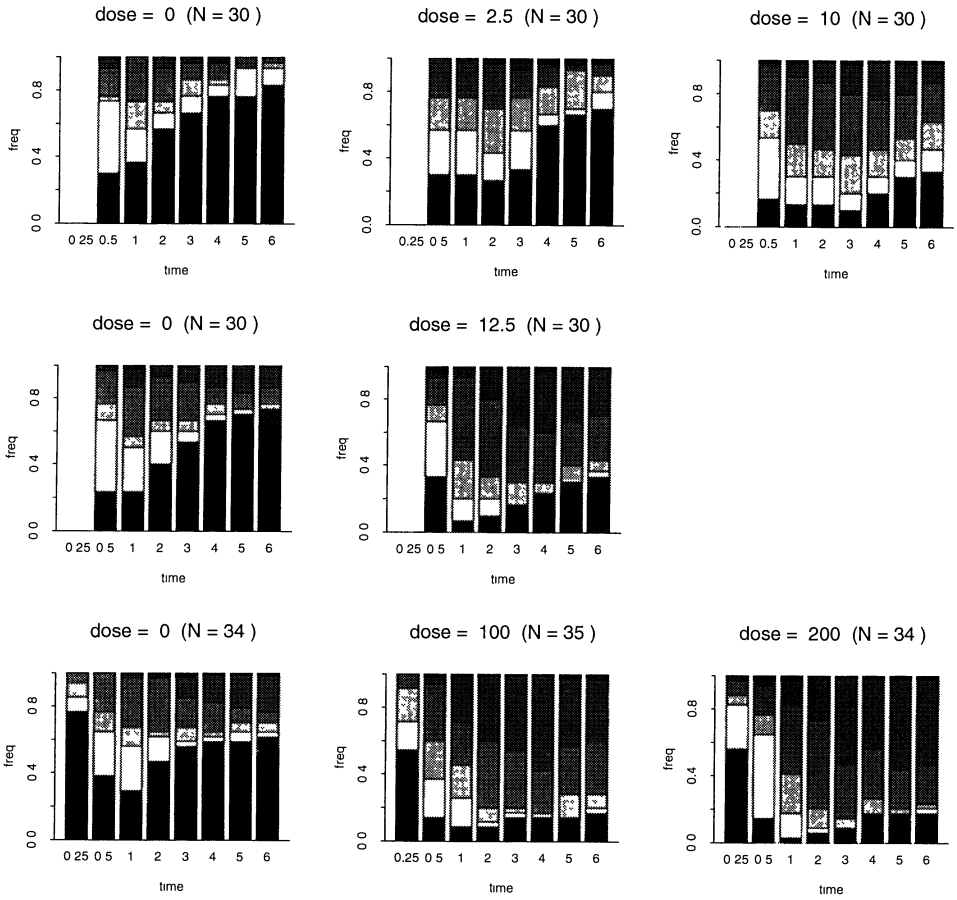

Figure 1. Frequencies of Pain Relief Responses at Different Doses and Different Times. This is a revised version of Figure 1 of Sheiner, Beal, and Dunne, with the following changes: (1) The three experiments at dose = 0 are separated, so that each row of the figure shows the results of a single experiment; (2) responses after dropout are imputed as 0 (no pain relief); (3) the time scales of the graphs are aligned.

ment that two-compartment kinetics are appropriate here (the cited Gibaldi and Perrier 1982 reference is not specific to ketorolac). Such details are perhaps not necessary for a general article in *JASA*, but it is worth noting that estimating the pharmacokinetic parameters in a new application requires substantive research. The variability of the parameters for ketorolac might be assessed using several studies (e.g., Hayball et al. 1994, Jung, Mroszczak, and Bynum 1988, Jung et al. 1989, and Lucker, Bullingham, Hooftman, Lloyd, and Mroszczak 1994). Consideration of covariates such as age would also strengthen the results, as the elderly tend to show reduced elimination of the drug (see Jallad, Garg, Martinez, Mroszczak, and Weidler 1990). Along similar lines, the internal workings of the model could be better understood if blood samples could be taken from some of the subjects as well—although perhaps this would not be recommended in a pain-relief study.

1.2 Model Checking

A crucial part of working with a complicated model, especially when communicating with statisticians without detailed subject matter knowledge, is to build confidence by understanding the behavior of the parameters and the fit to data. Figures 3–8 serve these purposes excellently. On the theory that one cannot have too much of a good thing, we would like to suggest further steps to understanding and checking the model.

First, as discussed earlier, we believe that Figure 1, along with Figures 5, 6, and 7, would be more relevant to the pain-relief aspect of the experiment if the censored responses were imputed appropriately rather than excluded. We emphasize that there is nothing wrong with the existing figures, but they are tricky to interpret because they mix pain-relief and dropout behavior.

Second, there should be a way to check the fit of the model to individual pain-relief responses. For example, for each observation, one can obtain an expected pain-relief score under the model and compare it to the actual response. Our Figure 2 shows a plot of the average residual (actual–predicted) versus average predicted response for all of the individual responses aggregated, binned into 20 equal-sized categories. From the plot, it is clear that when the model predicts very low pain or very high pain scores, the predictions tend to be too low or too high. (Separate plots did not show this pattern to be concentrated at any particular dose.) This particular pattern seems fairly minor, and it is not clear what affect it has on the substantive questions of interest—but it suggests another direction in model checking.

1.3 Individual Variation

A related issue is the display of individual variation. The model has random effects and is fit to individual-level data—but the plots in the article (and our discussion also)

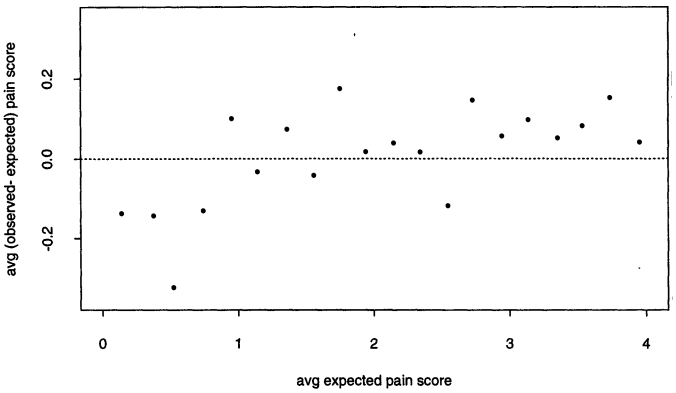

Figure 2. Plot of Average Residuals Versus Expected Pain Scores, With Responses Divided Into 20 Equal-Sized Bins Defined by Ranges of Expected Pain Scores. The prediction errors are relatively small, but with a consistent pattern that low predictions are too low and high predictions are too high.

display only aggregate results. At the very least, it would be informative to see the spread of the estimated random effects in the context of the estimated aggregate placebo and drug effects. In addition, the random effects could be used for further model checking. For example, an analysis of variance could be performed on the estimated random effects to see whether the differences seem to be real between the three experiments at dose = 0 (see our Fig. 1). Perhaps of more general interest, individual-level covariates could be included in the model as well, as discussed by Wakefield (1996). With regard to all of these model checks, a Bayesian analysis that summarizes results by posterior distributions of parameter estimates would also allow one to test the model using more general discrepancy measures (see Gelman, Meng, and Stern 1996).

1.4 Extrapolation

As the authors stress, a key reason for going to the trouble of fitting a pharmacokinetic model is that such models (if their parameters have reasonable values) should allow reasonable extrapolation to untried dose patterns, in a way that traditional hypothesis-testing methods cannot. Of course, once the model is fit and made a prediction made, it would be gratifying to see it tested. The cross-validation shown in Figure 7 is an excellent start in that direction. We wonder whether any tests are planned for ketorolac at the authors' recommended dosing regimen (25 mg at time 0 and at 4 hours). This would be a sensible check on the model and, to the extent that the model does not fit, would suggest ways in which it could be refined.

ADDITIONAL REFERENCES

Bois, F. Y., Gelman, A., Jiang, J., Maszle, D., and Alexeef, G. (1996), "Population Toxicokinetics of Tetrachloroethylene," *Archives of Toxicology*, 70, 347–355.

Gelman, A., Bois, F. Y., and Jiang, J. (1996), "Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions" *Journal of the American Statistical Association*, 91, 1400–1412.

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.

Hayball, P. J., Wrobel, J., Tamblyn, J. G., and Nation, R. L. (1994), "The Pharmacokinetics of Ketorolac Enantiomers Following intramuscular Administration of the Racemate," *British Journal of Clinical Pharmacology*, 37, 75–78.

Jallad, N. S., Garg, D. C., Martinez, J. J., Mroszczak, E. J., and Weidler, D. J. (1990), "Pharmacokinetics of Single-Dose Oral and Intramuscular Ketorolac Tromethamine in the Young and Elderly," *Journal of Clinical Pharmacology*, 30, 76–81.

Jung, D., Mroszczak, E., and Bynum, L. (1988), "Pharmacokinetics of Ketorolac Tromethamine in Humans after Intravenous, Intramuscular and Oral Administration," *European Journal of Clinical Pharmacology*, 35, 423–425.

Jung, D., Mroszczak, E. J., Wu, A., Ling, T. L., Sevelius, H., and Bynum, L. (1989), "Pharmacokinetics of Ketorolac and *p*-hydroxyketorolac Following Oral and Intramuscular Administration of Ketorolac Tromethamine," *Pharmaceutical Research*, 6, 62–65.

Lucker, P., Bullingham, R., Hooftman, L., Lloyd, J., and Mroszczak, E. (1994), "Tolerability, Central Effects and Pharmacokinetics of Intravenous Ketorolac Tromethamine in Volunteers," *International Journal of Clinical Pharmacology and Therapeutics*, 32, 409–414.

Ten Have, T. R., Pulkstenis, E., Kunselman, A., and Landis, J. R. (1997), "Mixed Effects Logistic Regression Models for Longitudinal Binary Response Data With Informative Dropout," technical report, Pennsylvania State University, Dept. of Biostatistics.

Vakily, M., Corrigan, B., and Jamali, F. (1995), "The Problem of Racemization in the Stereospecific Assay and Pharmacokinetic Evaluation of Ketorolac in Humans and Rats," *Pharmaceutical Research*, 12, 1652–1657.

Wakefield, J. C. (1996), "The Bayesian Analysis of Population Pharmacokinetic Models," *Journal of the American Statistical Association*, 91, 917–927.