Bayesian Analysis (2009) **4**, Number 1, pp. 41–44

Comment on Article by Craigmile et al.

David B. Dunson*

I would first like to congratulate the authors for taking on a very ambitious modeling project. It is increasingly the case in a broad variety of application areas that there is interest in combining data collected on different measurement scales, while accounting for complications such as censoring, missingness and spatial misalignment. Bayesian hierarchical modeling provides a natural paradigm for addressing such problems. However, difficulties in specifying the model and implementing the analysis provide a significant hurdle to many statisticians, limiting the use of Bayesian hierarchical models in applications such as exposure assessment. It is too often the case that ad hoc methods are used for combining information from different sources, with the results then used to make important regulatory decisions. For example, it is standard practice to summarize data from a particular source using a point estimate, which is then included without accounting for estimation uncertainty in a model for data from a different source. This type of exercise leads to a modeling house of cards, which can certainly produce highly misleading inferences and predictions. The Craigmile et al. article is a step in the right direction toward shifting the current standard practice.

All that said, I feel it necessary in my role as a discussant to raise a number of issues with their analysis. First, the primary goal of the article is to provide a behind-the-scenes look at the practical details involved in implementing a Bayesian hierarchical analysis in a complex setting. However, the authors focus on an arsenic exposure pathways application, which has been considered in a number of previous Bayesian analyses. The authors rely heavily on the models chosen in these previous analyses without much justification, allowing them to essentially bypass the challenging issue of model uncertainty. This luxury will be unavailable in most applications, and it is typically necessary to properly account for model uncertainty to obtain reasonable inferences and predictions, particularly in exposure pathway modeling.

I found the author's assessment of the Clayton et al. (2002) model insufficient. Even an eyeball analysis of the correlation table in Figure 4 showed a number of discrepancies, and it seems much more appealing to formally account for the obvious fact that the Clayton et al. (2002) model may not hold exactly. This can be done by allowing uncertainty in the directed acyclic graph (DAG), while using a prior centered on the Clayton et al. (2002) structure. There are many more modeling assumptions made, and it seems insufficient to simply rely on graphs of model fit and posterior predictive checks, because it is unlikely that the sparse data will provide evidence against even a poor model. In the authors' defense, accommodating model uncertainty does complicate the analysis, and one must make some pragmatic decisions to simplify the process.

A second major issue is the widespread use of diffuse but proper normal and gamma priors. Because data are sparse and there are many parameters to estimate, high variance priors seem to be a very bad idea. The first issue is that one does not want

*Department of Statistical Science, Duke University, Durham, NC, mailto:dunson@stat.duke.edu

 DOI:10.1214/09-BA401B

to choose a prior with very large variance if the limiting case as the variance $\to \infty$ corresponds to an improper posterior because the MCMC algorithm may behave as if the posterior is improper and have very poor mixing and convergence properties (Natarajan and McCulloch 1998). For normal-gamma hierarchical models, it is well known that such issues arise, and indeed this is exactly what occurred in the authors' analysis. Their response was to claim non-identifiability and simplify the model in a stepwise manner until the MCMC algorithm appeared to converge. From a Bayesian perspective, I find that this type of strategy should not be recommended in general. If the MCMC exhibits poor mixing and slow convergence, it is not necessarily due to the data containing insufficient information about the parameters. In many Bayesian hierarchical models, there is a high degree of posterior correlation in the parameters that leads to very poor mixing, and diffuse gamma priors can exacerbate this problem.

Instead of discarding pieces of data and simplifying the model, there are two preferred strategies that immediately come to mind (1) use a more efficient MCMC algorithm; and (2) use more informative priors. Regarding (1), a natural approach to follow in the types of normal-gamma hierarchical models considered in this article is to use parameter expansion Gibbs steps (Gelman 2006; Ghosh and Dunson 2009). Such algorithms are no more complicated to implement than typical Gibbs samplers and maintain conjugacy. Regarding (2), it is my view that high variance priors are almost always a bad idea in sparse data situations in which there are many parameters.

In such settings, even if there is essentially no prior information available about the parameters, heavy-tailed shrinkage priors (e.g., Cauchy) have been widely recommended for Bayesian robustness. Such priors allow parameters to effectively drop out of the model by setting their values close to zero, while maintaining conjugacy through the use of a scale mixture of normals specification. This is done adaptively, while appropriately accounting for uncertainty. For variance parameters, one can typically use a reparameterized model, as described in Gelman (2006), and then use half-Cauchy priors as a heavy-tailed default for random effect standard deviations. Conjugacy is maintained, and the analysis is often more efficient because mixing is improved, requiring fewer MCMC samples need to be collected. In the arsenic models considered in the paper, I suspect that one has substantial prior information to use in choosing an informative prior. Even rough prior information that would allow one to choose a plausible range of values for a coefficient would be very helpful.

A final issue of concern is possibly biased sampling and informative missingness. Regarding the biased sampling issue, the NHEXAS data were collected from a three-stage, population-based sampling design, and the 179 individuals represented a small subset of those households that were contacted. It seems that the sampling weights used in the NHEXAS survey possibly have an important impact on the results, but such weights were not included in the analysis. Biased sampling is a concern for other components of the data as well. For example, the authors rely on topsoil and stream sediment measurements from two surveys. One wonders how the monitoring sites were chosen. Often, there is a tendency to over-sample sites that are suspected to have high concentrations or to place a monitor at a location of concern. Certainly, such biased sampling will lead to problems in the analysis if not appropriately dealt with.

Missing data creates a related concern. Examining table 2, it is clear that there is a large proportion missing for many of the media. It seems that the authors have implicitly assumed missing at random (MAR) in their analyses. However, violations of this assumption may have a large impact on the results given the large amount of missing data.

References

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 3: 515–534. 42

Ghosh, J. and Dunson, D. B. (2009). "Default prior distributions and efficient posterior computation in Bayesian factor analysis." *Journal of Computational and Graphical Statistics*, to appear. 42

Natarajan, R. and McCulloch, R. E. (1998). "Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference?" *Journal of Computational and Graphical Statistics*, 7: 267–277. 42