Bayesian Analysis (2006) **1**, Number 3, pp. 543–546

Comment on article by Browne and Draper

Paul C Lambert*
University of Leicester, UK

I would like to congratulate the authors on a clearly written and detailed paper. Large scale simulation studies are important to understand the properties of complex models which we are increasingly able to fit. The amount of computing time needed for the simulation studies performed by Browne and Draper (stated in the Appendix) demonstrates that this can be a time consuming task.

As stated by the authors, the use of multilevel models has grown substantially over the last few years. However, as listed in the first paragraph of section 1, there are a number of competing methods proposed for their estimation, both Bayesian and likelihood based. Within the Bayesian framework there is of course the added issue of the choice of prior distributions for the various model parameters. It is worth noting here that the increased use of Bayesian methods over the last decade or so has not necessarily been due to a philosophical shift, but rather a desire to fit complex models, with software such as WinBUGS enabling users to do this. Many of these users want their 'data to dominate' and therefore want all prior distributions to be non-informative. However, this is rarely straightforward and in hierarchical models it is the choice of prior distribution for the hierarchical variance parameters that has been shown to be most crucial, particularly in small samples. In earlier work we conducted a simulation study on the choice of prior distribution for the variance component (between study variance) in a meta-analysis of aggregated data (Lambert et al. 2005). One of the advantages of using aggregated data is that models are quicker to fit and we were able to compare 13 different prior distributions for 9 different scenarios. When the number of level 2 units is large the choice of prior distribution becomes less important. However, for many real applications in medicine one would expect the number of level 2 units to be small, for example meta-analysis (Sutton and Abrams 2001) and cluster randomised trials (Turner et al. 2001). It is to the situations where there are only a small number of level 2 units that I wish to address most of my comments.

- The inverse-gamma (ϵ, ϵ) distribution is by far the most common prior distribution used for variance components. One reason for this is that in the set of BUGS examples (Spiegelhalter et al. 1996a,b) it is the only prior distribution used for variance components, with $\epsilon = 0.001$. As Browne and Draper point out, the inverse-gamma (ϵ, ϵ) distribution has a spike near zero and that this can create problems for low values of σ_u^2 or when the number of level 2 units is small. These problems have recently been demonstrated by Gelman (Gelman 2006). My view is that there is a need to educate users to move away from tradition and avoid using this prior distribution for hierarchical variance parameters, particularly when the number of level 2 units is small.

*University of Leicester, UK, http://www.hs.le.ac.uk/personal/pl4/

- One of the problems with both the prior distributions investigated by Browne and Draper is that with a small number of level 2 units, the posterior distribution may include implausibility large values for σ_u^2. The use of weakly informative prior distributions that will give low (or zero) probability to values that are clearly implausible are likely to produce more realistic estimates (Gelman 2006; Lambert et al. 2005).

- A disadvantage of the two prior distributions chosen by Browne and Draper is that interpretation on the variance or precision scales is less obvious and for this reason I prefer prior distributions on the standard deviation scale, particularly if using informative or weakly informative prior distributions, as these will be on the same scale as the model and thus provides greater transparency. Two such prior distributions are the uniform or half-normal distributions. In addition the half-Cauchy distribution used by Gelman looks particularly promising for situations with a small number of level 2 units (Gelman 2006)

- Another important point illustrated in the paper is that the choice of summary statistic (mean, median or mode) can lead to very different point estimates, particularly in small samples. This is of course to be expected when the posterior distribution is skewed, but does illustrate the importance in reporting which summary measure has been used. It is also worthwhile noting that the majority of WinBUGS users rarely report the mode for the simple reason that the standard output does not report it.

- The results of the simulation for the random effects logistic regression (RELR) are particularly interesting with the quasi-likelihood methods performing poorly even with a large number of level 2 and level 3 units. It is for these types of models that the Bayesian approach is particularly advantageous. This is demonstated by their use in genetic epidemiology where complex random effects models are used to model genetic and enviromental associations in pedigree data (Burton et al. 1999). The RELR simulation study has a large number of units in comparison to the variance components simulations and one would expect similar problems to occur regarding the choice of prior distributions when the number of level 2 (or level 3) units are small. I agree with Browne and Draper that other likelihood based approaches need further investigation, in particular the use of adaptive quadrature based methods (Pinheiro and Bates 1995) and hierarchical generalized linear models (Lee and Nelder 1996). However, due to flexibility and potential to extend the models I think it is likely that a Bayesian approach is the most sensible in these situations.

- It is clear is that for any Bayesian hierarchical model involving a small number of units, the role of the prior distribution for the hierarchical variance parameters is crucial and that there is unlikely to be an 'off-the-shelf' vague prior distribution suitable for all scenarios. Therefore a sensitivity analysis should routinely be performed. Finally, it is worth reiterating the importance of reporting all prior distributions used, in both the main and sensitivity analyses, and their impact on results.

References

Burton, P. R., Tiller, K. J., Gurrin, L. C., Cookson, W. O. C. M., Musk, A. W., and Palmer, L. J. (1999). "Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMS) and Gibbs sampling." *Genetic Epidemiology*, 17: 118–140. 544

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, ??: ???–??? 543, 544

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and D. R. Jones, D. (2005). "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS." *Statistics in Medicine*, 24: 2401–2428. 543, 544

Lee, Y. and Nelder, J. A. (1996). "Hierarchical generalized linear models (with discussion)." *Journal of the Royal Statistical Society Series B*, 58: 619–656. 544

Pinheiro, J. C. and Bates, D. M. (1995). "Approximations to the log-likelihood function in the nonlinear mixed-effects model." *Journal of Computational and Graphical Statistics*, 4: 12–35. 544

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996a). *BUGS Examples Volume 1, Version 0.5*. Cambridge: MRC Biostatistics Unit. 543

— (1996b). *BUGS Examples Volume 2, Version 0.5*. Cambridge: MRC Biostatistics Unit. 543

Sutton, A. J. and Abrams, K. R. (2001). "Bayesian methods in meta-analysis and evidence synthesis." *Statistical Methods in Medical Research*, 10: 277–303. 543

Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001). "Bayesian methods of analysis for cluster randomized trials with binary outcome data." *Statistics in Medicine*, 20: 453–472. 543