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/

