Comments on *Guidance for the use of Bayesian Statistics in Medical Device Clinical Trials*

David Spiegelhalter (Council Member) on behalf of the

Royal Statistical Society
12 Errol Street
LONDON
EC1Y 8LX
UK

August 20th 2006

Written in consultation with Sheila Bird and Deborah Ashby.

These guidelines are well-written and very welcome. We appreciate the efforts that have been made to communicate complex ideas without being too prescriptive or too technical. We shall not comment in detail on every aspect of the guidance, but would particularly like to welcome the explicit recognition of important but potentially controversial issues, for example

1. the central role of assumptions of exchangeability;
2. the notion of 'acceptable prior information';
3. the possible need to discount prior information
4. the need to investigate operating characteristics of Bayesian designs
5. the potential role for predictive probability calculations

While not suggesting any fundamental changes to the guidance, we do feel that some clarification and expansion may be appropriate on a range of issues. We list these below, with some suggested additions that we hope may be considered appropriate, and make more minor comments in an Appendix.

1. **New evidence:** *page 8, line –9, and page 10, line –6*. The current guidance appears to expect that all prior information needs to be 'on the table' before the start of the confirmatory study. However there must surely be an explicit mechanism for evidence to become available during the course of the study and to be incorporated into the prior. To quote David Cox (1999*) "I was surprised to read that priors must be chosen before the data have been seen. Nothing in the formalism demands this. Prior does not refer to time, but to a situation, hypothetical when we have data, where we assess what our evidence would have been if we had had no data."*

2. **Sensitivity analysis:** *page 9,line –6.* Currently sensitivity analysis to alternative prior specification is only recommended 'in some cases'. We feel this should be an essential aspect of any Bayesian report, in particular when 'discounting' prior information or using supposedly 'non-informative' priors for variance components or other non-location parameters. There may also be a role for expressing a formal degree of scepticism concerning large treatment effects.

3. **Exchangeability***: page 13, Section 4.6 andf page 19, line –17.* An operational definition of exchangeability would be useful. While the formal definition in terms of invariance of joint probability distributions is not appropriate, perhaps a phrase such as "exchangeability is essentially a judgement that, although we do not expect the observations to be equal, we see no reason to expect any systematic difference between them".

4. **Indirect comparisons:** *page 16, section 5.4, last para.* Perhaps more discussion due here on the potential dangers of essentially non-randomised comparisons, and references to the growing literature on `indirect' and `mixed-treatment' comparisons: see, for example, Lu and Ades (2004).

5. **Permissible evidence:** *page 17 and page 30.* It is perhaps surprising there is no mention made of `pre-clinical' studies, using `engineering knowledge, animal studies and so on. It would be better if a clear statement made either allowing this evidence or not.

6. **Use of historical information:** *page 18.* There are a number of different models which lead to different forms discounting of historical information. Possibly refer to Spiegelhalter et al (2004) p 148.

7. **Discounting prior information:** *page 18, para 3. page 31, page 35.* Page 31 advises "adding a constant to the (between)-study variance until the predictive probability of the claim is relatively low", while page 35 suggests you "increase the variance of the prior distribution until the Type 1 error rate reduces sufficiently". These ideas are somewhat ad-hoc but seem a reasonable way of preventing the claim being a foregone conclusion, and so allow the final study to have a chance of disproving the claim. It may be better to have a more unified approach expressed solely in terms of inflating the prior variance. The requirement of controlling the Type 1 error is quite strong: according to some quick calculations, it could result in the elimination of all prior information if the prior mean is more than 50% of the classical critical value of the treatment effect (details on request). Some more calculations of this sort might be useful to check this guidance is not too prescriptive.

8. **Non-informative priors:** *page 18.* It would be good to add that "non-informative" priors for between-study variances are not fully established, and sensitivity analysis is appropriate. Perhaps "non-informative" should always be in quotes.

9. **Sample size:** *page 21, Section 5.7, last para.* The phrase "posterior distribution of the sample size" is not clear. Presumably this is the predictive distribution for the eventual sample size required? *Page 21, last para.* It is not clear what is meant by this sentence.

10. **Operating characteristics:** *page 22, Section 5.8: page 31, 'predictive probability': page 34, Section 9.4.* A distinction should be made between operating characteristics *conditional* on specified parameter values (traditional Type 1 and Type II error), and *unconditional* values averaged over prior distributions. Currently the Sections on operating characteristics appear to only use conditional values (although including prior evidence in the analysis), while the Section on predictive probability suggest unconditional analysis. It would perhaps be better to integrate this discussion.

11. **Summaries of the posterior distribution:** *page 22, Section 6.1*. We would suggest plots of the full posterior distribution, and also full numerical and graphical summaries of the prior distribution.

12. **Interval hypotheses:** *Page 22, section 6.2* Could add that hypothesis of "equivalence" and "non-inferiority" are easily handled by reporting the posterior probability content of the region of interest.

13. **Stopping criteria:** *page 23 'deciding when to stop a trial'.* While it may be reasonable to use the predictive probability to help in the stopping decision for futility, we are doubtful of its value in stopping for efficacy, since the crucial summary to report is the current posterior probability of the hypothesis of superiority, and so this should guide the stopping rule.

14. **Predictive probabilities:** *page 31, 'predictive probability', para 1*. We are not sure what this guidance means: does it mean that the predictive (average) power should not exceed the traditional power evaluated at a suitable alternative hypothesis?

15. **Model selection:** *page 32, section 9.2*. We feel that in general the Bayes factor approach to model selection is difficult to implement in MCMC and potentially very sensitive to apparently innocuous choices of parameters in supposedly "non-informative'" priors. Mention could be again be made to the DIC approach in Spiegelhalter *et a*l (2002)

Appendix – minor points

1. *page 5 line- 12* `making the effects local and not systemic' needs some clarification
2. *page 6 and 33*. The address in the footnotes should be Medical Research Council Biostatistics Unit, Cambridge
3. *page 7.* An important additional forthcoming reference is the major review conducted by Ashby (2006)
4. *page 10, para 3* Different *frequentist* analyses can also reach different conclusions!
5. *page 12, Section 4.4* Writing *p(x|data, prior)* perhaps is more appropriate, and emphasises that all is conditional on the prior.
6. *page 12, Section 4.5* Predictive inference can also be carried out within frequentist analyses, but are more natural in a Bayesian framework
7. *page 14 Section 4.8, para 2* This statement may not be generally accessible.
8. *page 21, line 9* 'a single observation is sufficient' seems too extreme an illustration of a principle.
9. *page 22, Section 5.8, line 4* Surely a distinction should be made between 'ineffective' and 'unsafe' ? Safety monitoring is a whole different issue.
10. *page 23, final, para.* Could add that final posterior distribution becomes the prior for post-marketing surveillance.

Ashby D (2006) Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine,* **23**, (in press)

Lu G and Ades AE (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, **23**, 3105-3124