NOVARTIS

06d-0191-emc0001-02.doc 23-Aug-2006 (15:47)B&SR

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Comments by Novartis Pharma Biostatistics and Statistical Reporting

Author(s): Beat Neuenschwander, Amy Racine

Document type: Review Document

Document status: Final

Release date:

Number of pages: 9

1 General Comments

This is a well-written document addressing the main aspects of Bayesian design and analysis in Medical Device Clinical Trials.

It would be good to have a subsection in Section 3 about foundations, because in various places it is said that the Bayesian approach is coherent, consistent and scientifically valid. Bayesian statistics is not just another complementary approach to statistics, it has its strength in the sound foundations, and since recently, in the success in applications.

In general, more references might be helpful. We have added a list of additional references, some of which might be useful. Please feel free to add whatever you think would help the understanding of certain aspects.

2 Specific Comments

2.1 Section 3

Section 3.1.
"…consistent, mathematically formal method called Bayes Theorem…"
This seems to be a somewhat narrow perspective. Something should be said about full probability calculus (by using this calculus rigorously the approach is internally consistent and a lot of adhockery can be avoided), Bayes theorem is just one consequence of this. For example, other quantities of interest (like predictive distributions) are derived via the probability calculus. A good reference might be enough to support this.

Section 3.2, 1st paragraph
"informative prior information" sounds a bit strange and could be replaced by "prior information" without changing the meaning.

Last sentence. An example might be useful, e.g. complex hierarchical models (?).

Section 3.2, 3rd paragraph
A word of caution might be in place here. It is usually not straightforward to properly quantify prior information due to the problem of between-trial variability (sometimes it can be estimated, but not always).

Section 3.3

References: more references would be useful, in order to stress the fact that MCMC is an extremely well-developed and important field of Bayesian statistics. For additional references see Section 3 of this document.

Section 3.5.

WinBUGS is not the only commonly available computer program, but it's probably the only generic one, and it is used by a majority of people (15000 registered users by the end of 2005, Spiegelhalter, personal communication).

Section 3.6

References: 2nd edition by Gamerman & Lopez (see below).

Section 3.7

This section appears a bit weird. It cannot push aside sound science, because if properly applied it is sound science. Maybe it's just the title of this subsection that needs to be changed.

Section 3.8

With regard to interim looks, frequentists would claim that they can stop the trial as well (e.g. with group sequential methods). Is it easier to do things like this in the Bayesian setting?

Last paragraph. Maybe add something like: by using MCMC or other simulation methods, asymptotic results can be avoided.

Section 3.9

Bullets 2 and 3 are used in the frequentist approach as well. The main difference is the pre-specification of the prior.

The clinical agreement of the appropriateness of the prior, what does this mean? Especially later we get into down-weighting the prior and page 9 checking the robustness of priors. Does this require that we perform the evaluation over a range of prior weight and assumptions to check whether we lead to similar posterior conclusions?

2.2 Section 4

Section 4.1

The terminology endpoint instead of parameter is misleading. For example, the endpoint is time to healing, and the parameter might be the mean of time to healing in the statistical model.

The word probability is used a lot, but what is really meant are probability distributions over quantities of interest. We think the distinction is relevant, otherwise people think that the outcome of a Bayesian analysis is a probability (one number), which is clearly not the case. Maybe the final outcome of a Bayesian analysis will be one number, but the outcomes of Bayesian inference are probability distributions (posterior, predictive).

"If absolutely nothing is known …, something called a non-informative prior distribution may be specified". This is sloppy wording: the area of non-informative priors is not as simple as it might look like.

"If nothing is known about a parameter, a prior distribution reflecting this ignorance should be used".

Bayes' theorem and posterior probabilities

"…scientifically valid way". Why? Only if model and prior are okay; refer to probability calculus.

The Bayesian paradigm

Some references to the foundations of statistics should be given. Also, probabilities are used differently in the Bayesian and frequentist paradigm.

Decision rules

"For Bayesian trials, hypotheses are assessed with decision rules that are based on posterior probabilities."

Section 4.2

The notation **x** for parameter is unusual and confusing. Why not just use the standard Greek letter θ for parameters, and X or Y for the data.

3rd paragraph: a figure might help. The prior has more mass between 0.2 and 0.3 that between 0.7 and 0.8. This is particular useful for rare events. There are many priors that put more weight on small values. Any guidance?

Section 4.3

"The likelihood is the statistical model… " is good enough.

Section 4.4

Maybe here it would be good to have Bayes theorem written down…? It's nowhere stated in mathematical form.

Section 4.5

Predictive distribution instead of predictive probability.

Here it would be nice to have the derivation of the predictive distribution as the weighted average of the sampling distribution of the future data (weighted over the posterior). Just to illustrate how simple and straightforward the Bayesian approach is when it comes to predictions.

Last bullet: model checking (via posterior predictive model checks). A reference (e.g. Gelman) might be in place.

Section 4.6

The 1st paragraph is a challenge. It's difficult to make recommendations here. Is there no good reference that explains the concept of exchangeability more clearly?

Section 4.7

Why not use the binomial example with fixed sample size vs. negative-binomial to illustrate the principle.

Maybe it would be worthwhile to mention the pure likelihood approach to statistics propagated by Edwards, Royall, and Blume (see Blume's tutorial in Statistics in Medicine, and the Royall in Section 3 of this document). To make the point, there are non-Bayesians who are committed to the likelihood principle.

2.3 Section 5

Section 5.5

Part about informative priors.

Last paragraph. We think the amount of prior information should not depend on how many patients will be enrolled in the study. If the prior information is very informative, maybe the study is not needed at all.

Maybe it would be good to say something about evidence synthesis and discounting of historical information (see Spiegelhalter, Abrams, Myles (2004)).

Page 17, last paragraph

We know little about device trials. Do we need two independent trials? We are not sure how results of two trials are presented. I one of the first trial used as the prior (or down-weighted prior) of the second trial and provide a combined a posterior for decision making?

Page 18, top

"However, if differences…are large". Does this mean that one has to check compatibility of prior and current study after the data in the current study are available? And how exactly? And in the 2nd paragraph "properly calibrated" historical control? Quantitatively how do we decide that the historical control is properly calibrated after we included loads of covariates. The same issue arises on page 21, 2nd paragraph from bottom, verification of prior: how?

Page 18, middle

"…increasing stringency of the decision rule". An example might help.

Section 5.7

2nd paragraph: "If there were no variability…"

Special considerations when sizing a Bayesian trial: the 2nd paragraph about the minimum sample size is unclear, an example would help.

2.4 Section 6

Section 6.1

Are Bayes factors an option?

Section 6.2

Why "other types" of Bayesian intervals? Aren't highest posterior density intervals and central posterior intervals examples of credible intervals.

Section 6.4

Model checking. The terminology “posterior predictive checks” should appear here. And DIC cannot be used for model checking, only for comparing models (model selection).

Deciding when to stop a trial. We have little problem of stopping a trial for futility. For stopping a trial for success, do we need to use a pessimistic (very unfavorable) prior to avoid early stopping for success?

3 Additional References

Apologies for the fact that some of these might already me mentioned in the Guidance document.

3.1 Books

Berger, J.O. (1985) Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Carlin, P.C. and Louis, T.A. (1996) Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall, London.

Chen, M.-H., Shao, Q.-M., Ibrahim, J.G. (2000) Monte Carlo Methods in Bayesian Computation. Springer, New York.

Gamerman, D. and Lopes, H.F. (2006) Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. 2nd Edition, Chapman & Hall/CRC, Boca Raton.

Lindley, D.V. (1972) Bayesian Statistics, a Review. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

O’Hagan. A. (1994) Bayesian Inference, Kendall’s Advanced Theory of Statistics, Volume 2B. University Press, Cambridge.

Pratt, J.W., Raiffa, H., and Schlaifer, R. (2001) Introduction to Statistical Decision Theory. Massachusetts Institute of Technology.

Press, S.J. (1989) Bayesian Statistics: Principles, Models, and Applications. Wiley, New York.

Robert, C.P. (1994) The Bayesian Choice: A Decision-Theoretic Motivation. Springer, New York.

Robert, C.P. and Casella, G. (1999) Monte Carlo Statistical Methods. Springer, New York.

Royall, R. (1997) Statistical Evidence: A Likelihood Paradigm, Chapman & Hall, London.

Sivia, D.S. (1996) Data Analysis: A Bayesian Tutorial. Clarendon Press, Oxford.

3.2 Papers

Special issue of Clinical Trials

The workshop ”Can Bayesian Approaches to Studying New Treatments Improve Regulatory Decision Making?”, was jointly organized by the FDA and Johns Hopkins University at the

FDA in May 2004 and included contributions from health authorities, academia and industry; see the special issue of *Clinical Trials*, 2005:2.

Introduction to Bayesian Statistics in Clinical Trials

Berry, D.A. (2005) Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis, Clinical Trials 2:295-300.

Goodman, S.N. (2005) Introduction to Bayesian methods I: measuring the strength of evidence, Clinical Trials 2:282-290.

Louis, T.A. (2005) Introduction to Bayesian methods II: fundamental concepts, Clinical Trials 2:291-294.

Non-statistical journals

Berry. D.A. (2006) Bayesian Clinical Trials, Nature Reviews, 5:27-36.

Couzin, J. (2004) The New Math of Clinical Trials. Science 303:784-786

Davidoff, F. (1999) Standing Statistics Right Side Up. Annals of Internal Medicine, 130(12):1019-1021.

Goodman, S.N. (1999) Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy, Annals of Internal Medicine 130(12):995-1004.

Goodman, S.N. (1999) Toward Evidence-Based Medical Statistics. 2: The Bayes Factor, Annals of Internal Medicine 130(12):1005-1013.

Malakoff, D. (1999) Bayes offers a “new” way to make sense of numbers. Science 19(286):1460-1464

Sterne, J.A.C. and Smith, G.D. (2001) Sifting the evidence – what’s wrong with significance tests? British Medical Journal, 322:226-231.

Applications

Berry. S.M., Berry, D.A., Natarajan, K., Lin, C.-S., Hennekens, C.H., and Belder,R. (2004) Bayesian survival analysis with nonproportional hazards: metaanalysis of combination Pravastatin-Aspirin. Journal of the American Statistical Association, 99(465):36-44.

Goodman, S.N., and Sladky, J.T. (2005) A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barré syndrome, Clinical Trials 2.305-310.

Grieve, A.P., and Krams, M. (2005) ASTIN: a Bayesian adaptive dose-response trial in acute stroke, Clinical Trials 2:340-351.

Lipscomb, B., Ma, G., and Berry, D.A. (2005) Bayesian predictions of final outcomes: regulatory approval of a spinal implant, Clinical Trials 2:325-333.

Clinical Trials and Biostatistics (general)

Ashby, D. and Smith, A.F.M. (2000) Evidence-based medicine as Bayesian decision-making, Statistics in Medicine 19:3291-3305.

Breslow, N. (1990) Biostatistics and Bayes. Statistical Science 5(3): 269-298.

Matthews, R. (2001) Methods for Assessing the Credibility of Clinical Trial Outcomes. Drug Information Journal, 35:1469-1478.

Data Monitoring, Interim Analysis

Ashby, D. and Tan, S.-B. (2005) Where's the utility in Bayesian data-monitoring of clinical trials? Clinical Trials 2:197-208.

Fayers, P.M., Ashby, D. and Parmar, M.K. (1997) Bayesian Data Monitoring in Clinical Trials, Tutorial in Biostatistics, Statistics in Medicine 16:1413-1430.

Heitjan, D. (1997) Bayesian interim analysis of phase II cancer clinical trials, Statistics in Medicine, 16:1791-1802.

Sydes, M.R., Spiegelhalter, D.J., Altman, D.G., Babiker, A.B., Parmar, M.K.B, and the DAMOCLES Group (2004) Systematic qualitative review of the literature on data monitoring committees for randomized controlled trials, Clinical Trials, 1:60-79.

Frequentist and Bayesian Statistics

Efron, B. (2005) Bayesians, Frequentists, and Scientists, Journal of the American Statistical Association. 100(469):

Health Authorities

Campbell, G. (2005) The experience in the FDA's Center for Devices and Radiological Health with Bayesian Strategies, Clinical Trials 2:359-363.

Meta-Analysis

Warn, D.E., Thompson, S.G. and Spiegelhalter, D.J. (2002) Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales, Statistics in Medicine, 21:1601-1623.

Model averaging, model selection, model criticism

Draper, D. (1995) Assessment and Propagation of Model Uncertainty. Journal of the Royal Statistical Society, 57(1):45-97.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999) Bayesian Model Averaging: A Tutorial. Statistical Science 14(4):382-417.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002): Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B, 64(4):583-639.

P-values and Bayes Factors

Berger, J.O, and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of P values and evidence. Journal of the American Statistical Association, 82(397):112-139.

Casella, G., and Berger, R.L. (1987). Reconciling Bayesian and Frequentist evidence in the one-sided testing problem. Journal of the American Statistical Association, 82(397):106-111.

Han, C. and Carlin, B.P. (2001) Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review, Journal of the American Association 96(455):

Kass, R.E. and Raftery, A.E. (1995) Bayes Factors, Journal of the American Statistical Association, 90(430):1773-1795.

Priors

Chaloner, K. and Rhame, F.S. (2001) Quantifying and documenting prior beliefs in clinical trials. Statistics in Medicine 20:581-600.

Safety

Berry, S.M. and Berry, D.A. (2004) Accounting for Multiplicities in Assessing Drug Safety: a Three-level hierarchical mixture model. Biometrics 60:418-426.

Sample size determination

Cheng, Y., Su, F., and Berry, D.A. (2003) Choosing sample size for clinical trial using decision analysis. Biometrika 90(4):923-936.

Inoue, L.Y.T., Berry, D.A. and Parmigiani, G. (2005) Relationship Between Bayesian and Frequentist Sample Size Determination. The American Statistician, 59(1):79-87.

Joseph, L., Wolfson, D.B., and Du Berger, R. (1995) Some comments on Bayesian sample size determination. The Statistician, 44(2):167-171.

Joseph, L., Du Berger, R. and Belisle, P. (1997) Bayesian and Mixed Bayesian/Likelihood Criteria for Sample Size Determination. Statistics in Medicine, 16:769-781.

Pezeshk, H. (2003) Bayesian techniques for sample size determination in clinical trials: a short review. Statistical Methods in Medical Research, 12:489-504.