Bayesian Statistics in Medicine: Where Are We and Where Should We Be Going?

Dalene Stangl and Don Berry
Institute of Statistics and Decision Sciences, Duke University, Durham, N.C.

Introduction

Advances in numerical methods and computation have made it possible to implement Bayesian analysis previously considered infeasible. One need only open a recent statistical journal to gaze upon an array of articles including Bayesian in the title or listed as a keyword. This is true even for journals devoted completely to statistical methods and applications in medicine such as *Statistics in Medicine* and *Statistical Methods in Health Research*. While clinical trials are the most prevalent area of application (Abrams et al. 1996, Berry 1993, Brophy et al. 1995, Carlin et al. 1995, Cowles et al. 1996, Gatsonis et al. 1992, Gray 1994, Greenhouse et al. 1992, Gustafson 1995, Kadane 1996, Lewis et al. 1994, Lewis 1996, Racine et al. 1986, Simon et al. 1996, Smith et al. 1996, Spiegelhalter et al. 1994, Stangl 1995, Verdinelli 1996), this is not the only area of medicine where Bayesian methods are being used. The fields of epidemiology (Ashby et al., 1996, Erkanli et al. 1997, Malec et al. 1993, Raftery 1996, Waller et al. 1997), pharmacokinetics (Gelman et al. 1996, Greive 1988, Racine-Poon et al. 1996, Wakefield 1996), genetics (Churchill 1995, Liu 1995), imaging (Johnson 1995, Weir 1997), and health policy (Berry 1992, Christiansen et al. 1996, Hasselblad 1996, Kahn et al. 1996, Normand et al. 1997, Stangl et al. 1997) are also rapidly turning to Bayesian methods to help solve complicated modeling issues.

Statisticians who propose the use of Bayesian methods in medical applications vary in their motivations. Some simply see the Bayesian approach as a more natural and coherent way of thinking about science and learning, while others see it as a choice to be used when it is suitable for the problem at hand. With equally wide motivations, many other statisticians are turning to Bayesian methods to design, analyze, and summarize medical studies. The goal of this article is to review the technology transfer of Bayesian methods to medicine. The paper will be broken into six sections: comparative inference, design, elicitation, estimation, prediction, and decision analysis. Each section will cite applications in diverse areas of medicine. After examining the current state and direction of Bayesian methods in these areas we will delineate the gaps and directions that will lead to the further development and use of Bayesian methods in medicine.

Comparative Inference

Classical statistics can be thought of as a theoretical paradigm by which probability theory, sampling and data are combined to make inferences about populations. The applied statistician learns distribution theory and develops a toolkit of methods which are applied in accordance with the variables' level-of-measurement and method of collection. In contrast, the Bayesian paradigm is much less theoretically complex, requiring not much more than a grasp of Bayes theorem, and is much more analytically intense, in that there are no automatically applicable algorithms by which to get from data to result except in simple conjugate cases. Careful analytic thought is required for each new application as the complete model including previous information (prior), current information (likelihood), and values on potential actions (utilities) is likely to be unique in some way. The Bayesian paradigm discourages template analyses that can be applied indiscriminately across problems. While some would argue that this is a criticism of the Bayesian paradigm, we shall argue that it is indeed a strength. More careful thought allows versatility in choosing the quantities of interest and types of statements that can be made in drawing conclusions. It encourages greater collaboration between the statistician, client, and consumer--clearly a desirable mode of operation in medical studies.

The most basic difference between the Classical and Bayesian paradigm is the distinction between defining probability as long-term frequency versus degree of belief. Using the latter definition, probability applies to parameters and hypotheses as well as to data. Prior probabilities articulate our beliefs about parameters and outcomes before seeing the current data, while posterior and predictive probabilities combine prior probabilities and the likelihood function to describe our updated beliefs about parameters and future outcomes. This basic difference in the two paradigms leads to distinct designs and analyses of medical studies.

Design of Medical Studies

The Bayesian paradigm progresses from the viewpoint that design is a decision problem solved by maximizing the expected utility of the experiment. A comprehensive review of Bayesian design can be found in Chaloner and Verdinelli (1994). They review design for both linear and nonlinear models. Although the paper does not specifically address design for medical applications, most of what appears there is directly applicable.

Papers that do address design issues in medical contexts, particularly clinical trials include Berger and Berry (1988), Berry (1985, 1987, 1988, 1989a, 1989b, 1991), Berry et al.(1994, 1988, 1992), and Lewis et al.(1992, 1993, 1994). These works are broad, touching upon sample size, adaptive allocation,

monitoring and interim analysis, and stopping rules. Most of this work is related to sequential sampling and adaptive allocation of subjects to treatments. In sequential sampling, observations are made sequentially, and the decision to continue the experiment is made after each observation or a set of observations. The sampling plan that maximizes expected utility can be found using dynamic programming and backward induction. Carlin et al.(1997) address this same problem using a forward sampling algorithm that eases the analytic and computational burdens associated with backward induction, and they apply their method to data from a recent AIDS clinical trial. In adaptive allocation designs, patients are not balanced equally between the treatment and control groups, but rather are assigned in a way that maximizes a utility such as the total number of patients who respond favorably, including patients in the sample and those in the future patient horizon. The goal is to find an optimal balance between treating the current patient effectively and getting information that will help determine the most effective treatment for future patients. Berry and Eick (1995) show that the Bayesian adaptive treatment assignment performs better than competing strategies and recommend it for use when the disease or condition is rare yielding a small patient horizon. A review of this work on sequential sampling and adaptive designs is included in Berry et al.(1996).

Other good examples of Bayesian design are presented in Flournoy (1993), Lewis (1996), Clyde et al.(1996), and Gelfand and Vlachos (1995). Flournoy discusses a Bayesian approach to a Phase I clinical trial to determine toxicity dosage. Dosage is chosen to minimize predicted posterior variance, and Bayesian updating provides information for implementing an up-and-down treatment allocation scheme. Lewis considers sample size determinations when using interim analyses of a trial to determine the effect of phenytoin for the prophylaxis of early post-traumatic seizures in children who have suffered moderate to severe blunt head injury. He calculates required sample sizes from both a frequentist and Bayesian perspective and demonstrates that the Bayesian design has the advantage of minimizing the expectation of the sample size and total cost. Clyde and co-authors illustrate the computation of Bayesian designs in binary regression experiments. Their medical application is a heart defibrillator experiment in which the number of observations that can be taken is limited, optimal energy levels are unknown, and it is important to incorporate all available prior information. By modeling individual-to-individual variation in the appropriate defibrillation setting they can use information on past subjects in formulating sensible prior distributions for designing experiments for current patients. Gelfand and Vlachos consider the use of multiple stopping criteria in the design of a sequential clinical trial permitting multiple arms with multiple patient endpoints. In this work response is a categorical outcome in the form of cells in a multi-way contingency table. Since data is acquired sequentially, the crucial ingredient in the specification of a design is the stopping rule.

Elicitation of Priors

Elicitation of the prior distribution is a considerable challenge in most analyses. Typically, experts do not agree on what is known apriori, they underestimate their own uncertainty, and they have difficulty articulating their beliefs in terms of probability distributions. Dissecting past studies to glean information is always a good idea but designs, analyses, and results will differ across studies, so there is no automatic algorithm detailing how to use this information.

Chaloner (1996) presents an example of eliciting beliefs from experts involved in a clinical trial to determine the appropriate dosage of trimethoprim/sulfamethoxazole for fighting pneumocystis carinii pneumonia (PCP) in patients with AIDS. Using graphical methods presented in Chaloner et al. (1993) beliefs were elicited from three experts: two infectious disease MDs and an investigator running the design of the AIDS clinical trial. The entire script used for elicitation is presented. The first question asked in the script is "What is your best guess of the percentage of people assigned to the placebo group who will experience PCP 2 years after enrollment?" The script then asks for a percentage and 95% interval for the treatment group conditional on the experts' placebo percentage being correct. After responding, the expert is presented with a parametric distribution that approximates his/her beliefs. The expert is allowed to use slide-dialogs to adjust the curve to more accurately represent their beliefs. An overwhelming advantage of this methodology is that it does not restrict beliefs to parametric families. Based on experience, Chaloner argues that a quality elicitation will provide elicitees with interactive feedback, a scripted interview, and a systematic literature review. She encourages the use of percentiles and suggests using as many experts from as many sources of expertise as practically feasible.

Another example of eliciting expert beliefs is presented in Kadane and Wolfson (1996, 1997). This example uses a method that elicits quantiles of the posterior predictive distribution, in the belief that experts can more easily express opinions about the predictive quantity which they experience, than about the parameters which they do not experience. The example presented demonstrates elicitation for a normal linear regression model with two covariates. This method does not have the advantage of graphical display or the ability to handle nonparametric or even nonconjugate priors as does Chaloner's method, but it does allow the expert to articulate opinion based on up to four covariates. Other papers demonstrating the elicitation of expert beliefs include O'Hagan (1997) and Craig et al. (1997); however their examples are not related to medical studies.

Brophy and Joseph (1995) present an example of developing an empirically based prior. Using results of two previous clinical trials comparing thrombolytic strategies in acute myocardial infarction they demonstrate the use of historical data to develop a prior distribution. They also demonstrate how the

historical information can be down weighted if for some reason it should not be believed on par with the current data.

Because statisticians disagree on the value of prior information, how to find it, and how to use it, a great deal of analyses in medical studies are done using reference priors. Reference priors are of several types ranging from those that articulate complete equality across the range of the parameter space to those that base the prior on current data. Jeffreys priors are a common reference prior. This class of priors arises from the principle that any rule for defining a prior density should yield an equivalent result if applied to a transformed parameter. To derive such a prior, one takes the prior proportional to the square root of the Fisher information for the parameter of interest. Another type of reference prior is referred to by various names including flat, constant, or uniform. These priors entail picking a plausible range for the parameter and placing uniform probability across this range. The last type of reference prior that we will mention is the unity information prior. This prior has much in common with empirical Bayes methods in that priors are estimated from the current data. However, unit information priors do improve upon empirical Bayes methods in that uncertainty in the parameters is modeled. The prior is centered near the maximum likelihood estimate and the variance is set to approximate the concentration in the likelihood function after one observation.

An alternative that bypasses much of the controversy between expert, empirically-based, and reference priors is a compromise that supports the use of a range of priors. This method is described and exemplified in Spiegelhalter et al. (1994). These authors argue for the use of three priors: a reference prior that represents minimal prior information, a skeptical prior that represents the belief that large treatment differences are unlikely, and an enthusiastic prior that represents the belief of individuals who are reluctant to stop a trial when results support the null hypothesis. They argue that studies should be reported acknowledging that different individuals will hold different beliefs, and hence the results should be presented for a range of priors. They demonstrate the use of a range of priors in a clinical trial comparing two treatments for patients with resected cancer of the colon or rectum and three clinical trials testing the effect of adding a drug thought to enhance the effects of radiotherapy in various types of cancer. Other authors that demonstrate the use of a range of priors are Kass et al. (1989), Lewis (1996), and Smith et al. (1996).

Attempts to bypass specification of specific priors includes the use of classes of priors (Berger, 1992, Wasserman, 1992, and Greenhouse and Wasserman, 1994) and partitioning of priors into subspaces which either support or do not support the null hypothesis (Carlin et al., 1995, Carlin and Louis, 1996, Sargent and Carlin, 1996). Greenhouse et al.(1994) demonstrates the replacement of a single prior by the

use of the class of epsilon-contaminated priors in monitoring a phase II safety trial and in an analysis of the ECMO trial. This class is defined by

$$\Gamma_\varepsilon = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q; q \in Q\}$$

where $\varepsilon \in [0,1]$ reflects the amount of uncertainty concerning the accuracy of π_0 and Q is the class of all prior distributions. Upper and lower bounds for the posterior expectation of quantities of interest are calculated with reference to the entire class of priors. These bounds are computed for all values of ε, and the results are plotted versus ε. The authors state that if ε is small, and the interval between the upper and lower bounds is large, then the inferences are sensitive to the choice of the prior. In this case one must either collect more data, refine the specification of the prior distribution, or consider a more restricted class of priors. Guidelines for choosing between these options have not been developed nor have guides for what constitutes small ε and/or large intervals between the upper and lower bounds.

Carlin et al.(1995) take the approach of giving a partial characterization of the class of priors leading to a given decision conditional on the observed data. They demonstrate their method in the analysis of a trial design to determine the effect of two drugs to prevent toxoplasmosis in patients with aids. The decision of concern is the stopping rule of the trial, and the authors focus on the sensitivity of the stopping decision to the elicited prior distribution for the treatment effect. Extensions to the approach appear in Sargent and Carlin (1996) and are reviewed in Carlin and Louis (1996).

While this paper references studies that make use of experts and past data, most published studies use some variant of reference priors relying on the justification of convenience and "letting the data speak for itself". Frequently these analyses are supplemented by a sensitivity analysis that checks the robustness of results to specification of the prior. In contrast to the range of priors suggested above, these sensitivity analyses tend to check results across changes in scale rather than location. For example the variance in a prior is changed rather than shifting the central location of the prior. Similarly, rarely are assumptions about parameter independence checked. Although sensitivity to the prior is important in Bayesian analyses, it should not serve as a substitute for carefully thinking through prior distributions, and reference priors should not serve as a substitute for trying to quantify what is known apriori. The same is true for robustness checks and partitioning of priors. For most problems of importance we will not have the luxury of postponing decisions until enough data is collected so that priors are irrelevant. Usually decisions must be made with uncertainty, hence effort needs to go into modeling prior information apriori as well as assessing the impact of possible priors after collecting the current data.

Estimation

Estimation is the area of Bayesian methods that has advanced most in the last decade. Computational advances have been dramatic and are due to the technology transfer of numerical integration techniques (Naylor and Smith, 1982), analytic approximations (Tierney et al., 1986, 1989) and Markov chain Monte Carlo methods(MCMC) (Hastings, 1970, Geman and Geman, 1984) to statistics. While numerical integration techniques and analytic approximations launched Bayesian methods in applied medical contexts (Racine et al., 1986, Grieve, 1987, 1988a, 1988b, and Albert, 1988), it was MCMC methods that fueled them around the world.

MCMC methodology has been so important to Bayesian modeling because it trades the difficult problem of dealing with complex joint distributions for a sequence of simpler problems dealing with a single unknown by sampling from its full conditional distribution. This methodology allows Bayesian statisticians to estimate increasingly complex and hence realistic models. Some of the early works that demonstrated the utility of iterated simulation methods in applied work with generalized linear and proportional hazards models were Gelfand et al. (1990), Zeger and Karim (1991), and Dellaportas and Smith (1993). Today there have been literally hundreds of papers showing applications of MCMC methods in medical contexts. For readers interested in trying MCMC methods excellent introductory reviews are presented in Gelman et al.(1996) and Carlin and Louis (1996). The former is written specifically for biostatisticians, while the latter is not; however, the case studies in both tend toward medical contexts. For example, Gelman and Rubin model the difference in reaction times between schizophrenics and nonschizophrenics, and Carlin and Louis have examples covering drug response and survival in patients with AIDS, interim monitoring of clinical trials, and spatio-temporal mapping of cancer rates.

To easily see the increased complexity that Bayesian methods can now handle, readers need only peruse the pages of *Markov Chain Monte Carlo in Practice* (ed. Gilks, Richardson, Spiegelhalter, 1996) , *Bayesian Biostatistics* (ed. Berry and Stangl, 1996), and *Bayes and Empirical Bayes* (Carlin and Louis, 1996). These books contain numerous practical applications of complex hierarchical models in medical contexts. Applications include longitudinal, spatial, covariate measurement and survival models and span the substantive areas of immunology, pharmacology, transplantation, screening, clinical trials, epidemiology, and genetics.

In addition to the medical applications covered in these volumes many solo manuscripts using complex modeling have been published. These examples include multicenter clinical trials, medical imaging, DNA sequencing, and disease progression. Etzioni and Carlin (1993), Gray (1994), Stangl (1995), and

Gustafson (1997) reanalyze data from collaborative, multi-institution trials to assess intra and inter-site variability using Bayesian methods. Using adaptive numerical integration techniques Etzioni and Carlin fit a nonlinear dose-response curve with varying degrees of complexity in the hierarchical structure. Using Gibbs sampling to fit hierarchical proportional-hazards and exponential-changepoint models respectively, Gray and Stangl model univariate time-to-event outcomes in multicenter trials, while Gustafson fits hierarchical random effects for bivariate time-to-event outcomes. Johnson et al. (1995) show how to use more costly, but higher-resolution magnetic resonance imaging in developing prior distributions in a hierarchical model for lower-resolution SPECT data. Churchill (1995) assesses the accuracy of a finished DNA sequence. Slate and Cronin (1997) fit hierarchical changepoint models to indicate recurrence of prostate cancer. Their changepoint rule works better to detect recurrence than a commonly used threshold rule. Craig and Newton (1997) fit a hidden Markov model of disease progression. Bayesian methods allow fitting of this model despite complications due to irregular data collection, treatment interventions, and death.

Several authors have used Bayesian methods in the analysis of complex health surveys. Malec et al.(1993) use a hierarchical logistic regression model to analyze health utilization data from the National Health Interview Survey. They frame their inferential questions as predictions for finite populations and use Gibbs in a fully Bayesian analysis. Crawford et al. (1993, 1995) uses Bayesian approaches to impute missing data in a health services context. In the 1995 paper, the authors use multiple imputation to analyze a longitudinal data set from a study of care giving for the elderly to determine characteristic of individuals and families that predict substitution of informal for formal care. In the earlier paper she uses Bayesian methods to model medical costs measured from the Harvard Medical Practice Study, a complex sample survey in which nonignorable nonresponse was a significant problem. Belin et al.(1995) present methods for combining data from several sources to derive more accurate estimates of effects in large studies of health care utilization, for example, effects of interventions aimed at improving screening rates for mammography and Pap smears. The authors use techniques based on multiple imputation for handling missing classifications and on propensity scores for adjusting for patient differences across different health care providers. Erkanli et al.(1997) estimate prevalence of psychiatric disorders in children based on a two-stage multi-cluster sampling design. A large sample of children were screened using a brief interview, and then depending on the screening score a fraction of the original sample were selected for an intensive follow-up interview.

In conjunction with the development of iterative simulation methods continually improving software is making this methodology more accessible to applied statisticians. The most versatile and advanced of such software is BUGS. (Spiegelhalter et al., 1994). The acronym stands for Bayesian inference using

Gibbs sampling. BUGS provides syntax for specifying graphical models and a command language for running Gibbs sampling. Indeed the BUGS manual itself is filled with many excellent pedagogic examples in medical contexts. To accompany BUGS is software called CODA (Best et al., 1995) that monitors convergence of the sampling algorthms. An overview of other available software is presented in Carlin and Louis (1996).

With the ability to fit more complex models comes increased complexity in model checking and selection. Raftery et al.(1996a, 1996b, 1997) propose the use of Bayes factors in model selection and model averaging. He demonstrates these methods in an analysis of a clinical trial for treatment of cirrhosis of the liver and in an analysis to determine whether fat and alcohol consumption are associated with breast cancer. The use of Bayes factors in model determination is also demonstrated by Greenhouse (1992) and Greenhouse and Wasserman (1995) in the analysis of clinical trials. While the use of Bayes factors is attractive, they must be used and interpreted with caution. Kadane and Dickey (1980) argue that Bayes factors are not useful as a complete data summary for decision making, because the implicit 0-1 loss function rarely applies. Similarly, Lavine and Schervish (1997) argue that Bayes factors should not be used for quantifying the degree to which observed data support or conflict with particular models because of logical incoherencies. They investigate the source of the problem and offer more appropriate interpretations of Bayes factors.

Prediction

One of the greatest advantages of the Bayesian paradigm is the ease with which one can focus on predictive distributions of variables of interest. For much of medical research it is the future observables rather than model parameters that are of primary focus. However, the structure of hypothesis testing has molded our thinking in terms of questions about parameters instead of predictions about future observations. Exceptions to this mode of thinking are presented in Berry (1991,1995), Berry et al. (1992, 1994), Grieve (1987, 1988a, 1988b), Gamerman (1991), Stangl (1995), and Qian et al(1996). Geisser and Grieve provide some of the early work on using predictive distributions in a medical context. Geisser showed that posterior interval estimates for probabilities for success can be interpreted as asymptotic prediction intervals for the proportion of successes from a large number of future trials. Grieve describes uses of predictive distributions in pharmaceutical research. Much of Berry's work on treatment allocation and sequential sampling relies on the use of predictive distributions. His work with Wolff and Sack demonstrates the use of predictive distributions to estimate the number of future cases of Haemophilus influenzae type b among Navajo children over a 20 year horizon.

Showing that treatment differences between chemotherapy alone and chemotherapy plus radiation therapy change across time, Gamerman uses predictive survival distributions to show that there is no easy choice between these two treatments for gastric cancer. In modeling the time-to-recurrence of major depressive disorder, Stangl demonstrates the use of calculating predictive distributions to compare treatments and easily incorporate results into a decision theoretic framework to ascertain which treatment gives the higher expected utility. Qian et al.(1996) presents the use of predictive distributions for interim monitoring of clinical trials and determining when it is optimal to stop a trial.

Decision Making

The Bayesian paradigm is decision-oriented. Some statisticians argue that learning, not decision making, is the goal of science. In our view, learning is decision making. It's deciding that their is enough evidence to believe one theory or hypothesis over another. We make decisions about what to believe, and we usually do so under great uncertainty. Rarely are studies replicated under identical conditions to the point where conclusions can be stated conclusively. Even more rarely are studies replicated under altered conditions to ensure previous results generalize. Yet, we choose to generalize and use our new beliefs to make all sorts of decisions such as whether to fund further studies, what research questions to fund next, what applications are worthwhile, etc.

In the Bayesian paradigm there is little room for the separatist notion that reaching statistical conclusions and deciding what to do with conclusions should be distinct enterprises. Indeed how results will be used should guide our choice of analysis and presentation methods. Wider use of Bayesian methods will likely diminish the prevalence of the separatist attitude many statisticians hold, and we will see more integration of statistics into the decision making process. References that cover the topic of decision making from an applied perspective include Sox et al.(1988), Clemen (1991), Berry (1995), Lewis and Berry (1998). Coverage from a more theoretical perspective can be found in DeGroot (1970) and Berger (1986).

While the role of decision making is clear in the design of medical studies where decisions on sampling, trreatment allocation, dosage, follow-up times, etc. must be made, decision making is also a big part of other aspects of studies as well. One example that demonstrates the centrality of decision making in every aspect of a study is presented in Parmigiani and Kamlet (1993). Assuming a 4-state stochastic process for the natural history of breast cancer, they derive analytic expressions for the expected number of examinations, treatment costs, and quality-adjusted life years, which are then used to evaluate and compare different screening strategies. In another example, Parmigiani et al.(1996) describe the role of Bayesian methods in developing clinical recommendations and illustrate their use in context of the

Stroke Prevention Policy Model developed by a large interdisciplinary group called the Stroke Prevention Patient Outcome Research Team. Berry et al.(1992, 1994) use a decision-theoretic framework to determine when to stop a trial to determine the effectiveness of a vaccine for a particular type of influenza for which American Indian and Eskimo children are at very high risk. The authors' goal is to minimize the number of influenza cases among Navajo children over a horizon of 20 years. Stangl (1995) shows how predictive distributions are easily incorporated into a decision theoretic framework when analyzing clinical trial data. Wolfson and Kadane (1996) use a decision theoretic framework to determine who should remediate a geographic area where contamination presents a health risk to those living there.

Of course decision making implies quantification of utilities. While it is often convenient to think about financial utilities, other consequences of decisions may be as or more important. The joint assessment of side effects, quality of life, associated risks, and other factors make elicitation and quantification of utilities the most challenging and least developed component of Bayesian methods.

Summary and Conclusions

Thanks to the advancement of numerical and computational algorithms, Bayesian methods are now being applied in diverse and complex medical applications. The areas of design and estimation have benefited most from these advances, while elicitation, prediction, and decision making are advancing, but at a slower rate. While the use of Bayesian methods is clearly increasing in quantity and quality, the authors would like to caution users in the areas of reference priors, sensitivity analysis, Bayes factors, multiplicities and inference.

We argued that in applying the Bayesian analysis to medical studies reference priors are frequently being used to the exclusion of modeling what is apriori known. Indeed one of the greatest advantages to a Bayesian approach is the ability to model what is known going into a clinical trial. Yet researchers are reluctant to do so for reasons including time, reviewer critique, and lack of know-how. We would like to encourage more use of experts and more use of previously collected data to formulate prior distributions.

Second, sensitivity analyses can not serve as a substitute for careful thought and quantification of what is know apriori, and these analyses must be carried out with more attention to all assumptions, not just those easy to check. Most Bayesians analyses present sensitivity checks to assess the impact of priors; however, most are inadequate, altering only a few assumptions, one at a time. More attention is needed to challenging assumptions pertaining to central location, exchangeability, symmetry, and the dependence between parameters. These are typically ignored, but authors conclude that results are robust to the prior specification.

Related to sensitivity analysis for the priors we choose, more efforts must be taken to examine the robustness of results to the often implicit utility/loss functions we choose. These functions will have as much or more impact on our answers as the prior, but yet rarely are papers rejected for lack of a sensitivity analysis on the utility function. In most cases we do not even ask that the utility function be made explicit. Whether the analysis is frequentist or Bayesian, we should demand that utility functions be made explicit, and for Bayesian analyses, the sensitivity analysis should focus on utility functions as well as the prior distributions.

The use of Bayes Factors has gained considerable popularity in the last several years. This use of Bayes factors must be viewed with caution for several reasons. First, in general the Bayes Factor approach provides a focus that leads us astray from a more important focus. In using Bayes factors the primary concern becomes the model instead of trying to make good predictions. Models themselves are not of primary interest, but of secondary interest in the sense that they are helpful in coming up with a distribution of the outcome that reflects the available information. This is not to say that Bayes factors cannot be used to improve predictions. Raftery (1996a) shows that model averaging can improve predictions, and Bayes factors are used for averaging. Second, for Bayes factors to give coherent answers, the structure of the problem must be quite special. Statisticians must be very careful to ensure that their problems fit this structure.

Directly related to design, estimation, and inference discussed in this paper is the problem of multiplicities. The types of multiplicities that are problematic for Bayesian as well as frequentists include multiple comparisons, variable selection in regression, selecting transformations in regression, subset analyses, data dredging, and publication biases. Bayesians disagree on whether adjustments should be made in estimation and inference to guard against errors that may arise due to these multiplicities. Some Bayesian argue that adjustments are legitimate and appropriate in many applications (Berry, 1988, DuMouchel, 1988, Gopalan and Berry, 1997, Berry and Hochbert, 1998). There are no standard adjustments for multiplicities, somethimes they are appropriate and other times they are not. When they are appropriate the amount of adjustment depends on the available evidence from both within and outside the study in question. The absence of automatic procedures for adjustments, like the absence of automatic procedures for elicitation and analysis, is inherent in Bayesian analysis. No two problems are identical, no two researchers will have exactly the same knowledge base or interpet evidence in exactly the same way. This is just the nature of statistical inference. This is not wrong, it just requires more thought and communication between parties involved.

Our last suggestion is an encouragement for the use of decision theoretic approaches in statistical analysis. The following quote, from Dennis Lindley's discussion of Spiegelhalter et al. (1994) states the need quite eloquently:

"The authors dismiss the use of expected utility as being unrealistic....It must be recognized that clinical trials are not there for inference but to reach a decision, and the omission of their raison d'être is serious. In the long term, expected utility is realistic, and indeed, necessary... We should use Bayesian concepts firstly because they work and secondly because only with them is it possible to be coherent. To appreciate this, contemplate the situation when you have just been given some data. Why not just look at them and express an opinion about the parameter? Why go through elaborate calculations, whether frequentist or Bayesian? If we can assess a prior directly, why not a posterior? The answer is: to achieve coherence, to make all our beliefs cohere: in particular, to make our final beliefs cohere with those that went into our likelihood. This is why frequentist p-values, interpreted as beliefs in the null hypothesis, are unsound; they do not cohere. The same reasoning applies to decisions as to beliefs. It is only by using expected utility that we can be sure that our actions will fit together sensibly. I suspect that the procedure of continuing with the trial until a tail area probability in the posterior is small, is just as incoherent as a belief based on the tail area, p-value. Or if it is coherent, it implies an inept utility, such as one taking only values 0 and 1."

References

1. Abrams, K., Ashby, D., Houghton, J. and Riley, D. (1996). Assessing drug interactions: Tamoxifen and cyclophosphamide. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 531-554.

2. Albert, J.H. (1988). Bayesian estimation of Poisson means using a hierarchical log-linear model. In *Bayesian Statistics III*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), Oxford University Press: 519-531.

3. Berger J.O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, **25**: 303-328.

4. Berger J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag.

5. Berger J.O., and Berry D.A. (1988). The relevance of stopping rules in statistical inference. *Statistical Decision Theory and Related Topics IV*. New York: Springer-Verlag: 29-72.

6. Berger J.O., and Berry D.A. (1988). Statistical analysis and the illusion of objectivity. *Am Sci* **76**:159-165.

7. Belin, T.R., Elashoff, R.M., Leung, K, Nisenbaum, R, Bastani, R, Nasseri, K., and Maxwell, A. (1995). Combining information from multiple sources in the analysis of non-equivalent control group design. In *Case Studies in Bayesian Statistics II*, (eds., C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 241-260.

8. Berry D.A. (1985). Interim analysis in clinical trials: classical versus Bayesian approaches. *Statistics in Medicine* **4**:521-526.

9. Berry, D.A. (1987). Interim analysis in clinical trials: The role of the likelihood principle. *The American Statistician* **41**:117-122.

10. Berry, D.A. (1988a). Interim analysis in clinical research. *Cancer Invest.* **5**:469-477.

11. Berry, C.A. (1988b). Multiple comparisons, multiple tests, and data dredging: A Baysian perspective with discussion). In *Bayesian Statistics 3*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), Oxford University Press: 79-94.

12. Berry, D.A. (1989a). Monitoring accumulating data in a clinical trial. *Biometrics* **45**:1197-1211.

13. Berry, D.A. (1989c). Inferential aspects of adaptive allocation rules. *Proceedings of the Pharmaceutical Section of the American Statistical Association*, ASA, Washington DC:1-8.

14. Berry D.A. (1991). Bayesian methodology in phase III trials. *Drug Information Association Journal* **25**:345-368.

15. Berry DA (1991). Experimental design for drug development: A Bayesian approach. *Journal of Biopharmaceutical Statistics* **1**:81-101.

16. Berry, D.A. (1993). A case for Bayesianism in clinical trials (with discussion). *Statistics in Medicine* **12**:1377-1404.

17. Berry, D.A. (1995). Decision analyis and Bayesian methods in clinical trials. In *Recent Advances in Clinical Trial Design and Analysis*. (ed. P. Thall), Kluwer Press, New York: 125-154.

18. Berry D.A., and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments.* London: Chapman-Hall.

19. Berry, D.A., and Hardwick, J. (1993). Using historical controls in clinical trials: Application to ECMO. *Statistical Decision Theory and Related Topics* **V,** New York: Springer-Verlag. (eds. Berger JO, Gupta S): 141-156.

20. Berry, D.A., and Eick, S.G. (1994). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Statistics in Medicine* **14**:231-246.

21. Berry, D.A., and Ho, C.H. (1988). One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics* **44**:219-227.

22. Berry, D.A., Wolff, M.C., and Sack, D. (1992). Public health decision making: A sequential vaccine trial (with discussion). In *Bayesian Statistics*, (eds. Bernardo JM, Berger JO, Dawid AP, Smith AFM)**,** Oxford, England: Oxford University Press: 79-96.

23. Berry, D.A., Wolff, M.C., and Sack, D. (1994). Decision making during a phase III randomized controlled trial. *Controlled Clinical Trials* **15**:360-379.

24. Berry, D.A., and Hochberg, Y. (1988). On Bayesian and quasi-Bayesian approaches to multiple comparison problems. *Journal of Statistical Planning and Inference* (To appear).

25. Best, N., Cowles, M.K., and Vines, K. (1995). CODA: Convergence diagnostics and output software for Gibbs sampling output. Version 0.30. MRC Biostatistics Unit, Cambridge, UK.

26. Brophy J.M., and Joseph, L. (1995). Placing trials in context using Bayesian analysis: Gusto revisited by Reverend Bayes. *Journal of the American Medical Association* **273**(11): 871-875.

27. Carlin, B.P., Kadane, J.B., and Gelfand, A. (1996). Approaches for Optimal Sequential Decisions in Clinical Trials, *Research Report 96-004*, Division of Biostatistics, Univ. of Minnesota.

28. Carlin, B.P., and Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall.

29. Carlin, B.P., Chaloner, K. M., Louis, T.A., and Rhame, F.S.(1995). Elicitation, monitoring, and analysis for an AIDS clinical trial. In *Case Studies in Bayesian Statistics: Volume II*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer: 48-84.

30. Chaloner, K. (1996). Elicitation of Prior Distributions. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker: 141-156.

31. Chaloner, K and Verdinelli, I. (1994). Bayesian design: A review. *Statistical Science*, **10**: 237-304.

32. Chaloner, K., Church, T., Matts, J.P., and Louis, T.A. (1993). Graphical elicitation of a prior distribution for an AIDS clinical trial. *The Statistician* **42**:341-353.

33. Christiansen, C.L., and Morris, C.N. (1996). Fitting and checking a two-level poisson model: modeling patient mortality rates in heart transplant patients. In In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 467-501.

34. Churchill, G.A. (1995). Accurate restoration of DNA sequences. In *Case Studies in Bayesian Statistics*, ed, C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla, Springer-Verlag: 90-148.

35. Clemen, R.T. (1991). *Making Hard Decisions*. PWS-KEnt, Boston.

36. Clyde, M., Mueller, P., and Parmigiani, G. (1995). Optimal design for heart defibrillators. *In Bayesian Statistics in Science and Engineering: Cases Studies II,* (eds. C. Gatsonis, J.S. Hodges, R.E. Kass, and N.D. Singpurwalla), Springer Verlag, New York.

37. Clyde, M., Mueller, P., and Parmigiani, G. (1995). Optimal design for heart defibrillators. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 297-320.

38. Cowles, M.K, Carlin, B.P., and Connett, J.E. (1996). Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association*, **91**: 86-98.

39. Craig, P.S., Goldstein, M., Seheult, A.H., and Smith, J.A. (1997). Constructing partial prior specifications for models of complex physical systems. *The Statistician* **46**(4):xx-xx.

40. Craig, B.A., and Newton, M.A. (1997). Modeling the history of diabetic retinopathy. In *Case Studies in Bayesian Statistics*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 305-324.

41. Crawford, S., Tennstedt, S.L., and McKinlay, J.B. (1995). Longitudinal care patterns for disabled elders: A Bayesian analysis of missing data. In *Case Studies in Bayesian Statistics*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 293-308.

42. Crawford, S.L., Johnson, W.G., and Laird, N.M. (1993). Bayes analysis of model-based methods for nonignorable nonresponse in the Harvard Medical Practice Survey (with discussion). In *Case Studies in Bayesian Statistics*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 78-117.

43. DeGroot, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

44. Delaportas, P., and Smith, A.F.M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl. Statist*., **42**: 443-459.

45. DuMouchel W. (1988). A Bayesian model and graphical elicitation procedure for multiple comparisons. In *Bayesian Statistics III*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), Oxford University Press.

46. DuMouchel W. (1989). Bayesian metaanalysis. *Statistical Methodology in the Pharmaceutical Sciences*, (ed. Berry, D.A.), New York: Marcel Dekker, 509-529.

47. Erkanli, A., Soyer, R., and Stangl, D. (1997). Hierarchical Bayesian analysis of prevalence estimation. In *Case Studies in Bayesian Statistics III*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 325-346.

48. Etzioni, R. and Carlin, B. (1993). Bayesian Analysis of the Ames Salmonella/Microsome Assay. In *Case Studies in Bayesian Statistics I*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 311-323.

49. Flournoy, N. (1993). A clinical experiment in bone marrow transplantation: Estimating a percentage point of a quantal response curve. In *Case Studies in Bayesian Statistics I*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla, Springer-Verlag: 324-336.

50. Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Appl. Statist.*, **40**(1): 63-79.

51. Gatsonis, C., and Greenhouse, J.B. (1992). Bayesian methods for phase I clinical trials. *Statistics in Medicine*, **11**:1377-1389.

52. Geisser, S. (1982), Aspects of the predictive and estimative approaches to the determination of probabilities. *Biometrics Supplement: Current Topics in Biostatistics and Epidemiology*, **38**:75-93.

53. Gelfand, A.E., and Vlachos, P.K. (1995). Bayesian clinical trial design for categorical endpoint models. *Paper presented at the American Statistical Association Meetings*, Orlando, FL.

54. Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**:972-985.

55. Gelman, A., and Rubin, D. (1996). Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research* **5**: 339-355.

56. Gelman, A., Bois, F., and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91:1400-1412.

57. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**: 721-741.

58. George, S.L., Li, C.C., Berry, D.A., and Green, M.R. (1994). Stopping a clinical trial early: Frequentist and Bayesian approaches applied to a CALGB trial in non-small cell lung cancer. *Statistics in Medicine* **13**:1313-1327.

59. Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall.

60. Gilks, W.R., Clayton, D.G., Spiegelhalter, D.J., Best, N.G., McNeil, A.J., Sharples, L.D., and Kirby, A.J. (1993). Modeling complexity: Applications of Gibbs sampling in medicine. *J. R. Statist. Soc. B.* **55**(1): 39-52.

61. Gopalan, R., and Berry, D.A. (1997). Bayesian multiple comparisons using Dirichlet process priors. (Under revision for *Journal of the American Statistical Association*).

62. Gray, R. (1994). A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, **50**:244-253.

63. Greenhouse J.B., and Wasserman, L. (1995). A practical, robust method for Bayesian model selection: A case study in the analysis of clinical trials. *Paper presented at the American Statistical Association Meetings*, Carnegie-Mellon Technical Report #626.

64. Greenhouse, J.B., and Wasserman, L. (1995). Robust Bayesian methods for monitoring clinical trials. *Statistics in Medicine* **14**:1379-1391.

65. Greenhouse, J.B. (1992). On some applications of Bayesian methods in cancer clinical trials. *Statistics in Medicine*, **11**: 37-53.

66. Grieve, A.P. (1987). Applications of Bayesian software: two examples. *Statistician*, **36**: 283-288.

67. Grieve, A.P. (1988a). Some uses of predictive distributions in pharmaceutical research. *In Biometry - Clinical Trials and Related Topics*, (ed. Okuno, T.): 83-99.

68. Grieve, A.P. (1988b). A Bayesian approach to the analysis of LD50 experiments. In *Bayesian Statistics III*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), Oxford University Press: 617-630.

69. Gustafson, P. (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics*, March **53**: 230-242.

70. Gustafson, P. (1995). A Bayesian Analysis of Bivariate Survival Data from a Multicentre Cancer Clinical Trial, *Statistics in Medicine, **14***:2523-2535**.**

71. Hasselblad, V., and Jarabek, A.M. (1996). Dose-response analysis of toxic chemicals. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 235-259.

72. Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97-109.

73. Johnson, V., Bowsher, J., Jaszczak, R., and Turkington, T. (1995). Analysis and reconstruction of medical images using prior information. In *Case Studies in Bayesian Statistics II*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 149-240.

74. Joseph, L., Wolfson, D.B., and du Berger, R (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Statistician* 44:167-171.

75. Kadane, J.B., and Wolfson, L. J. (1996). Experiences in elicitation. *The Statistician* **46**(4): xx-xx.

76. Kadane, J.B. (ed.) (1996). *Bayesian methods and ethics in a clinical trial design*. Wiley.

77. Kadane, J.B., and Dickey, J.M. (1980). Bayesian decision theory and the simplification of models. In *Evaluation of Econometric Models*, (eds. J. Kmenta and J. Ramsey), Academic Press: 245-268.

78. Kahn, M.J. and Raftery, A.E. (1996). Discharge rates of Medicare stroke patients to skilled nursing facilities: Bayesian logistic regression with unobserved heterogeneity. *Journal of the American Statistical Association*, **91**: 29-41.

79. Kass, R.E., and Greenhouse, J.B. (1989). Comments on "Investigating therapies of potentially great benefit: ECMO" (by J.H. Ware). *Statistical Science* **4**:310-317.

80. Lavine, M., and Schervish, M. (1997). Bayes factors: What they are and what they are not. *ISDS Discussion* #97-02.

81. Lewis, R. J. (1996). Bayesian hypothesis testing: Interim analysis of a clinical trial evaluating phenytoin for the prophylaxis of early post-traumatic seizures in children. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 279-296.

82. Lewis R, and Berry DA (1994). Group sequential clinical trials: A classical evaluation of Bayesian decision-theoretic designs. *Journal of the American Statistical Association*, **89**:1528-1534.

83. Lewis R,and Berry DA (1998). Decision theory. In *Encyclopedia of Biostatistics*. (eds. P. Armitage and T. Colton) John Wiley & Sons, New York.

84. Liu, J.S. Neuwald, A.F., and Lawrence, C.E. (1995). Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *Journal of the American Statistical Association*, 90: 1156-1170.

85. Malec, D., Sedransk, J., and Tompkins, L. (1993). Bayesian predictive inference for small areas for binary variables in the national health interview survey. In *Case Studies in Bayesian Statistics I*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 377-389.

86. Naylor, J.C. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist*., **31**:214-255.

87. Normand, S.T., Glickman, M.E., and Ryan, T.J. (1997). Modeling mortality rats for elderly heart attack patients: profiling hospitals in the cooperative cardiovascular project. In *Case Studies in Bayesian Statistics III*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 435-456.

88. O'Hagan, A. (1997). Eliciting expert beliefs in substantial practical applications. *The Statistician* **46**(4):xx-xx.

89. Parmigiani, G., Ancukiewicz, M. and Matchar, D. (1996). Decision models in clinical recommendations development: The stroke prevention policy model. In *Bayesian Biostatistics*, ed. D.A. Berry and D.K. Stangl, Marcel Dekker, 207-233.

90. Parmigiani, G. and Kamlet, M.S. (1993). A cost-utility analysis of alternative strategies in screening for breast cancer. In *Case Studies in Bayesian Statistics*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla, Springer-Verlag: 390-402.

91. Racine-Poon, A. and Wakefield, J. (1996). Bayesian analysis of population pharmacokinetic and instantaneous pharmacodynamic relationships. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker: 355-388.

92. Racine, A., Grieve, A.P., Fluhler, H., and Smith, A.F.M. (1986). Bayesian methods in practice: experiences in the pharmaceutical industry (with discussion). *Appl. Statist.*, **35**:93-150.

93. Raftery, A.E., Madigan, D., and Hoeting, J.A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**:179-191.

94. Raftery, A.E., Madigan, D., and Volinsky, C.T. (1996a). Accounting for model uncertainty in survival analysis improves predictive performance. In *Bayesian Statistics V*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith), Oxford Press: 323-349.

95. Raftery, A.E., and Richardson, S. (1996b). Model selection for generalized linear models via GLIB: Application to nutrition and breast cancer. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 321-353.

96. Sargent, D. and Carlin, B.P. (1996) Robust Bayesian design and analysis of clinical trials via prior partitioning (with discussion). *IMS Lecture Note Series: Second International Workshop on Bayesian Robustness*.

97. Simon, R., Dixon, D., and Freidlin, B. (1996). Bayesian subset analysis of a clinical trial for the treatment of HIV infections. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker,555-576.

98. Slate, E.H. and Cronin, K.A. (1997). Changepoint modeling of longitudinal PSA as a biomarker for prostate cancer. In *Case Studies in Bayesian Statistics III*, (eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla), Springer-Verlag: 435-456.

99. Smith, T. C., Spiegelhalter, G.J., and Parmar, M.K.B. (1996). Bayesian meta-analysis of randomized trials using graphical models and BUGS. In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 411-427.

100. Sox, H.C., Blatt, M.A., Higgins, M.S., and Marton, K.I. (1988). *Medical Decision Making*. Butterworth and Heinemann, Boston.

101. Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. (1994). *BUGS: Bayesian Inference Using Gibbs Sampling*, Version 0.30. Cambridge: Medical Research Council Biostatistics Unit.

102. Spiegelhalter , D.J., Freedman, L.S., and Parmar, M.K. (1994). Bayesian approaches to randomized trials. *J. Royal Statistical Soc. Ser. A* **157**:357-416.

103. Stangl, D. (1995). Prediction and decision making using Bayesian hierarchical models, *Statistics in Medicine* **14**:2173-2190.

104. Stangl, D.K. and Huerta, G. (1996). Using Bayesian hierarchical models to assess the impact of managed-care strategies. *Discussion paper* 97-17, Institute of Statistics and Decision Sciences, Duke University.

105. Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**: 82-86.

106.Tierney, L., Kass, R., and Kadane, J.B. (1986). Fully exponential Laplace approximations to expectation and variances of nonpositive functions. *Journal of the American Statistical Association*, **84**: 710-716.

107.Verdinelli, K. Andrews, C., Detre, K. and P. Peduzzi (199x) The Bayesian approach to Meta-analysis: a case-study, *Carnegie Mellon Technical Report* 641.

108.Wakefield, J., and Bennett, J. (1996). The Bayesian modeling of covariates for population pharmacokinetic models. *Journal of the American Statistical Association*, **91**: 917-927.

109. Waller, L.A., Carlin, B.P., Xia, H., and Gelfand, A.E., (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92** , 607-617.

110.Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference. In *Bayesian Statistics IV*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), Oxford University Press:, 483-502.

111.Weir, I.S. (1997). Fully Bayesian reconstructions from single-photon emission computed tomography data. *Journal of the American Statistical Association*, **92**:49-60.

112.Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86**:79-86.

Verdinelli, K. Andrews, K. Detre and P. Peduzzi (1996) The Bayesian approach to Meta-analysis: a Case-study, Carnegie Mellon Technical Report 641. We consider the meta-analysis study of Yusuf et al. (1994) that used data collected by the Coronary Artery Bypass Graft Surgery Trialist Collaboration group from seven randomized clinical trials to compare coronary artery bypass graft surgery versus medical treatment. Among other characteristics of the studies involved, Yusuf et al. examined a particular aspect, regarding the relation among mortality rates and gravity of the baseline risk characteristics of the patients. As an alternative, we propose here the use of Bayesian meta-analysis for the same classification of patients that they considered. The method of stratification, denoted ``Step-wise Risk Score", includes clinical and angiographic baseline characteristics of patients [gif] and was developed in Yusuf et al. (1994).In this paper, we are mainly concerned in a detailed revision of the Bayesian methodology for meta-analysis studies. We will show the flexibility of the Bayesian method, examine the basic assumptions on which the method relies, and illustrate on the availability of the relevant software while examining this special case study.

Weir, I.S. (1997). Fully Bayesian reconstructions from single-photon emission computed tomography data. *Journal of the American Statistical Association*, 92:49-60. MCMC methods are used to estimate posterior mean and pixelwise credibility bounds for reconstructions from single-photon emission computerized tomography data.

Gelman, A., Bois, F., Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91:1400-1412. Describes a general approach to using Bayesian analysis of the estimation of parameters in physiological pharmacokinetic models. Complexity in modeling arises from having nearly as many parameters as observations. They discus how to check model fit and sensitivity to the prior distribution using posterior predictive simulation. They illustrate with an application to the toxicokinetics of tetrachloroethylene.

Liu, J.S. Neuwald, A.F., and Lawrence, C.E. (1995). Bayesian Models for Multiple Local Sequence Alignment and Gibbs Sampling Strategies. *Journal of the American Statistical Association*, 90: 1156-1170. Develop a Bayesian foundation for comparing alignment of multiple homologous protein or nucleic acid sequences obtained from different organisms.

Kahn, M.J. and Raftery, A.E. (1996). Discharge rates of Medicare stroke patients to skilled nursing facilities: Bayesian logistic regression with unobserved heterogeneity. *Journal of the American Statistical Association*, 91: 29-41. Uses 2-stage betabinomial hierarchical model to determine factors associated with hospitals' propensities for discharging Medicare stroke patients to skilled nursing facilities. Covariate information is included in the hyperparameters of the second-stage beta distribution.

Cowles, M.K, Carlin, B.P., and Connett,J.E. (1996). Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association*, 91: 86-98. Implemented a hierarchical model for longitudinal medication compliance.

Wakefield, J, and Bennett, J. (1996). The Bayesian modeling of covariates for population pharmacokinetic models. *Journal of the American Statistical Association*, 91: 917-927. Discuss modeling covariates that determine drug concentration in pharmocokinetic models. Illustrate with a study on the antibiotic vancomycin in babies.

Waller, L.A., Carlin, B.P., Xia, H., and Gelfand, A.E., (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92 , 607-617. Extend existing hierarchical spatial models to account for temporal effects and spatio-temporal interactions. They model county-specific lung cancer rates within the larger context of environmental justice.

Kadane, J.B. (ed.) (1996). *Bayesian methods and ethics in a clinical trial design*. Wiley. Centers around a trial to test the effectiveness of a treatment for hypertension after open heart surgery.

Christiansen, C.L., Morris, C.N. (1996). Fitting and checking a two-level poisson model: modeling patient mortality rates in heart transplant patients. In In *Bayesian Biostatistics*, (eds. D.A. Berry and D.K. Stangl), Marcel Dekker, 467-501. Provides a tutorial on the modeling process for hierarchical models. It also provides a structured process for specifying and checking future medical profiling models for transplant programs.