International Journal of Technology Assessment in Health Care, **17:1** (2001), 56–66.

WHY BAYESIAN ANALYSIS HASN'T CAUGHT ON IN HEALTHCARE DECISION MAKING

Robert L. Winkler
Duke University

Abstract
The objective of this paper is to discuss why Bayesian statistics is not used more in healthcare decision making and what might be done to increase the use of Bayesian methods. First, a case is made for why Bayesian analysis should be used more widely. Serious weaknesses of commonly used frequentist methods are discussed and contrasted with advantages of Bayesian methods. Next, the question of why Bayesian methods are not used more widely is addressed, considering both philosophical differences and practical issues. Contrary to what some might think, the practical issues are more important in this regard. Finally, some steps to encourage increased use of Bayesian methods in healthcare decision making are presented and discussed. These ideas are straightforward but are by no means trivial to implement, largely because it is difficult to fight tradition and make major paradigm shifts quickly. The primary needs are improved Bayesian training at the basic level (which means textbooks and other materials as well as training of those who teach at the basic level), procedures to make Bayesian analysis easier to understand and use (better software and standard methods for displaying and communicating Bayesian outputs will help here), and the education of decision makers about the advantages of Bayesian methods in important healthcare decision-making problems.

Keywords: Bayesian statistics, Decision theory, Healthcare decision making

Bayesian methods have received an increasing amount of attention in the statistics literature in recent years, and their proponents argue that they have important advantages over frequentist methods. Yet if we look at applications of statistics in healthcare decision making, the use of Bayesian analysis is relatively limited. The purpose of this paper is to discuss why this is the case and what might be done to change the state of affairs.

The situation discussed here is by no means unique to healthcare decision making. Although there are areas where Bayesian modeling has made inroads in applied studies, in many fields the prevalence of frequentist methods in statistical applications is similar to that in the health arena. However, aspects of healthcare decision making that are not present in some other arenas may complicate matters. For example, regulatory matters add another layer of complexity to many healthcare decisions, as do issues related to who ultimately pays for healthcare services (government, insurance, managed care organizations, the consumer, etc.). A key factor is that healthcare decision makers are usually not the direct consumers of healthcare services.

The author is grateful to Dennis Fryback, Tina Shih, and a reviewer for helpful comments. This work was supported in part by the National Science Foundation under grant SES 98-18855.

WHY SHOULD BAYESIAN ANALYSIS BE USED?

Much of the statistical analysis encountered in healthcare applications focuses on hypothesis testing, pitting a null hypothesis against an alternative. The results are summarized in terms of a *p* value, or often just an indication of whether the *p* value is less than certain values such as .05 or .01. Where estimation is considered instead of hypothesis testing, point estimates and confidence intervals are encountered. All of this is consistent with the standard frequentist approach to statistics.

Since this is widely used and widely taught, what can be wrong with it? The frequentist approach, while it does provide some information about the substantive questions being studied, has serious weaknesses, including the following:

1. *Frequentist methods tend to answer the wrong questions*. For example, instead of seeking the probability that a treatment effect is positive or is above a certain level, they give probabilities about the data given values of the treatment effect. Probabilities are conditioned on the treatment effect, when what we really need are probabilities *about* the treatment effects, conditioned on the data. In estimation, anyone who has taught elementary frequentist statistics realizes that most students find the correct frequentist interpretation of a confidence interval to be counterintuitive and cumbersome. The students want to interpret the interval as a probability statement about the parameter of interest, which directly relates to the question of interest. In hypothesis testing, the error probabilities of interest are the probability that a hypothesis is true (if we reject the hypothesis) or false (if we do not reject it), given the data. Frequentist error probabilities condition the other way, giving the probability that the test statistic is in a particular region (the rejection region), given that a hypothesis is true or false.
2. *Frequentist methods often violate important principles, primarily the likelihood principle*. The likelihood function gives the likelihood of the observed outcome given different values of the parameters, and the likelihood principle means that the likelihood of any outcome other than the observed outcome is irrelevant for inferential purposes (6). If we look at a *p* value, however, it involves likelihoods for outcomes other than the observed outcome and thereby violates the likelihood principle. For example, in a one-tailed test to the right, suppose the observed statistic is $z = 1.96$ and the sampling distribution of z under the null hypothesis H_0 follows a normal distribution. Then the *p* value is $P(z \geq 1.96|H_0) = .025$. This probability, however, involves not just the observed value 1.96, but also $z = 2.4$, $z = 5$, and all other values of z greater than 1.96. Since these other values of z were not observed, they should not be relevant.

   For an example in which violation of the likelihood principle is blatantly nonintuitive, consider an experiment in which patients' temperatures are recorded in degrees Fahrenheit using a thermometer that is extremely accurate. All of the recorded temperatures are between 96 and 104. If the experimenter then learns that the thermometer was inoperative at temperatures of 105 or greater but still accurate below 105, should this change the inferences that are made? Intuition would tell us no, and Bayesian methods would agree with intuition. Under some frequentist methods, however, the inferences would be modified even though it is clear that no actual temperatures in the experiment were as high as 105.
3. *The output of frequentist analyses is often too simplistic and can be misleading.* An example is the widespread use of arbitrary values such as .05 for alpha, the probability of a type I error in hypothesis testing. Such values are often used without any serious thought about the relative seriousness of the two types of error and without any consideration of the probability of a type II error, failing to reject a false null hypothesis. Also, *p* values may not be a good measure of the strength of the evidence against the null hypothesis. Lindley's paradox (24) shows that the distinction between statistical significance (as measured by a low *p* value) and real significance can be quite substantial. Berger and Sellke (5) demonstrate that *p* values can be much smaller than lower bounds on posterior probabilities of the null hypothesis, indicating that the evidence against the null hypothesis is not nearly as strong as the *p* value might imply.
4. *Much of the above discussion focuses on hypothesis testing, and part of the problem is, in my opinion, an overuse of the hypothesis-testing framework*. Too many problems are arbitrarily forced

into a dichotomous accept-or-reject framework that may not be appropriate. For example, statistically we look at null hypotheses, such as the hypothesis that a treatment effect is zero, with the alternative including all nonzero values. A treatment effect of *exactly* zero may not be reasonable, and the real interest may be in the size of the treatment effect. Is it small or large, or better yet, how small or large is it? This leads to viewing the situation more in terms of estimation than hypothesis testing. Of course, Bayesians as well as frequentists can fall into the trap of framing the problem in a black-and-white hypothesis-testing manner, but it seems more natural to be led to an estimation framework in the Bayesian approach.

5. *The output of frequentist analyses is not very useful for decision making*. What we typically need for decision-making purposes are probabilities, either probabilities for future outcomes (predictive probabilities) or probabilities for parameter values (posterior probabilities). The frequentist approach does not admit such probabilities, and therein lies the difficulty.
6. *Frequentist methods are too easy to apply mindlessly*. Plug-in formulas and easy-to-use software can spit out p values and other frequentist measures without forcing the user to really understand what is going on with the data.
7. *Contrary to claims, frequentist methods are not objective*. The way the problem is set up, the way data are collected, and various modeling assumptions (e.g., the choice of a linear model, a particular set of independent variables, and a normal distribution for the errors) are subjective choices. Sampling distributions and likelihood functions do not just sit there for the taking; they are created by modeling choices made by the statistical analyst. See Berger and Berry (4) for a discussion of the illusion of objectivity in statistical analysis.

Do these weaknesses mean that the frequentist approach should be discarded completely? It is certainly tempting to make that claim, as extreme as it sounds. The use of frequentist methods in practice could be improved greatly. Even when we limit our consideration to statisticians who are careful modelers, the Bayesian approach provides a more natural, sensible way to analyze data and to provide information that is useful for decision-making purposes.

Some advantages of Bayesian methods include:

1. *In contrast to frequentist methods, Bayesian methods answer the right questions and agree with natural common sense*. That is, they give explicit probability distributions for both parameters and future outcomes and revise these probabilities as new evidence becomes available. All probabilities are appropriately conditioned on the observed data, and users can find any probabilities of interest. Examples are the probability that one treatment effect is greater than another, the probabilities of various side effects from a given treatment, or the probability that a particular patient will survive a surgical procedure.
2. *Important basic principles are consistently followed by Bayesian procedures*. Most notable among these is the likelihood principle. In conditioning on the observed data, Bayesian methods ignore the likelihoods of any possible past outcomes that might have but did not occur. Only the likelihoods associated with the outcomes that actually occurred are used. Any probability manipulations, such as the determination of posterior and predictive probabilities, follow the usual rules of probability theory.
3. *The output of Bayesian methods is ideal for decision making and therefore for healthcare decision making*. Posterior and predictive probabilities represent the uncertainties of interest to decision makers and can be used in calculating any expected values of interest such as expected payoffs, expected losses, or expected utilities. Prior probabilities play an important role in preposterior decisions, which are decisions about whether to gather information, how to gather it, and how much information to gather. Such decisions are very important in settings such as the testing of new drugs or medical procedures. Work in the 1950s and early 1960s leading to the recent increase in interest in Bayesian methods was motivated in large part by decision-making considerations (28;29;30).
4. *Bayesian methods force careful thought*. They require more inputs than are needed in frequentist procedures, and this generally encourages more careful thought about the model. This is not to say

that many frequentist analyses are not done with careful modeling and thought, but as noted above, frequentist procedures are easier to apply mindlessly. It is certainly also possible to do "quick and dirty" Bayesian analysis, but in general the nature of the beast is that learning how to do Bayesian analyses and doing Bayesian analyses often heighten awareness of some modeling issues.

5. *In general, Bayesian analyses are more thorough and more transparent*. There are fewer formal inputs to a frequentist analysis, which leaves greater leeway for modeling choices that are not always explicitly discussed. The need to specify explicitly the inputs to a Bayesian analysis makes the analysis and any assumptions more transparent to observers and to decision makers for whom the analysis is relevant.
6. *Bayesian methods allow for the formal incorporation of relevant information other than the data immediately at hand*. Some actually view this as a weakness of the Bayesian approach, but in important real-world problems it is important to draw on any information that may be available pertaining to the question of interest. Excluding available information as the frequentist approach does is just as much a subjective judgment as including it explicitly and is less defensible. Lilford and Braunholtz (23, 607) state, "Health issues are now much more complex and the amount of disparate evidence that impacts on belief has increased. Only the Bayesian approach can do justice to all this information and provide the probabilistic basis for action."
7. *Bayesian techniques lend themselves better to situations with messy data sets and with multiple data sets*. For example, in complicated multiparameter models the likelihood function may fall prey to identification problems, where different combinations of parameter values lead to identical likelihoods. Prior information, as expressed through the prior distribution, can serve to identify such models and enable us to differentiate between these different combinations (7;26). With regard to multiple data sets, Bayesian procedures provide a natural framework for meta-analysis (11;12;14;15).

Of course, the above discussion has highlighted weaknesses of the frequentist approach and advantages of the Bayesian approach. Ironically, some of these characteristics contribute to the greater use of frequentist methods and the more limited use of Bayesian methods in practice. Nonetheless, it is better to have a basic understanding and appreciation of how to think about and model real-world problems under uncertainty and to be able to answer the questions of interest directly and usefully than to have a more superficial exposure to a large, simple-to-use tool kit that only gets at those questions of interest in a more indirect manner. An analogy, biased and less than perfect to be sure, might be to view frequentist analyses in the same vein as fast food and Bayesian methods as gourmet cuisine. In an ideal world, most people would prefer gourmet cuisine, but it is costly and difficult to prepare. Similarly, Bayesian methods are more suitable for healthcare decision making than frequentist techniques but are more difficult to apply, as shown in the next section.

WHY AREN'T BAYESIAN METHODS USED MORE WIDELY?

Now we get to the main issue of this paper. If Bayesian methods have so many advantages, why are they not used more widely in practice? Real estate agents often say that the three most important aspects of a property are location, location, and location. Similarly, we might say that the three most important reasons why Bayesian methods are not more widely used are tradition, tradition, and tradition. However, we should look a little deeper into this question. After all, Bayesian methods have won a considerable amount of acceptance among statistical researchers, as indicated by the space devoted to Bayesian work in leading statistics journals and the many conferences on Bayesian modeling. This acceptance has not translated into an equivalent "boom" in the use of Bayesian methods in actual applications, although such applications are on the increase in healthcare decision making (e.g., 10;17;33).

Impediments to the use of Bayesian methods in practice can be categorized roughly as philosophical differences or practical issues. Therefore, this discussion will be organized along those lines. Of the two, philosophical differences have received greater attention than practical issues over the years. I believe, however, that philosophical differences have receded into the background and that practical issues provide the main stumbling block to increased use of Bayesian modeling in practice.

First, let's consider philosophical differences. In the heady days when the groundwork was laid for modern Bayesian statistics in the 1950s and 1960s, philosophical differences were at the forefront of discussions between frequentists and Bayesians. At the heart of these differences was the need for a prior distribution in Bayesian analysis and the notion that probability statements could be made about parameter values. Such probability statements do not have relative-frequency underpinnings, and debate focused on the subjective interpretation of probability. One side argued that parameters are not random variables; they have fixed values, and we just don't know what those values are. The other side rebutted that this is exactly the point; if we are uncertain about the value of a parameter, we should express that uncertainty in the mathematical language of uncertainty: probability.

The debate about prior probabilities and the interpretation of probability led to claims that Bayesian analysis is inherently subjective while frequentist statistics is objective. The argument of subjectivity for Bayesian analysis is that different analysts can be expected to have different prior distributions, which implies that the results of the analysis will differ depending on the analyst. What is often forgotten is that different analysts using frequentist techniques can also wind up with different results, for reasons such as different models for the sampling process or different modes of analysis. Particularly in complex analyses, five different frequentist statisticians working independently could come up with five different sets of results. Imagine, for instance, a forecasting situation with a large number of potential independent variables. How likely is it that different analysts will choose exactly the same independent variables and the same form for the model? Even after seeing data, different people often have different opinions, with or without formal analysis.

This all calls to mind de Finetti's claim (16) that "probability does not exist." My interpretation of this claim is that the quest for "true," or objective, probabilities is futile. In any event, a smokescreen of alleged "objectivity" was created for non-Bayesian methods, and emphasis was given to the claimed subjective/objective dichotomy.

But how important are these philosophical differences? Does anyone stay up at night worrying about this sort of thing? Among applied statisticians, how many give serious thought to these issues? My sense is that the heat of the philosophical debate has been reduced considerably over the years. Frequentist results are typically misinterpreted along Bayesian lines anyway, with confidence intervals being interpreted as probability statements about the parameters and p values being interpreted as probabilities that the null hypothesis is true. Often, of course, the data pass the "interocular trauma test," which means that the results hit you between the eyes and the specific form of analysis or interpretation becomes irrelevant.

In some ways, the Bayesian spirit has made serious inroads into healthcare decision making. For example, there is widespread acceptance of the use of Bayesian analysis in diagnosis (31). The probability that a patient has a particular disease is revised after seeing the results of a test, and I understand that medical students are taught this revision procedure. Also, the existence of data safety monitoring boards to oversee data collection in experiments with new drugs or new procedures and to intervene to stop or modify the experiment if it seems appropriate is very much in the Bayesian spirit. Finally, some developments in expert systems and artifical intelligence for healthcare decision making rely on "Bayes nets" (20;21). When decision making is at the forefront, Bayesian techniques are more readily accepted and more often encountered than is the case in data analysis. This suggests to me that philosophical issues are not the primary reason that Bayesian procedures are not used more often.

This leads us to the consideration of practical issues. Under this heading, there are three aspects that deserve careful consideration: a) training; b) ease of use; and c) ready acceptance. These are the primary reasons that Bayesian practice has not shared the rapid growth of Bayesian theory.

First, let's consider training. This relates not just to university training in statistics, but to the way in which uncertainty is dealt with (or not dealt with) in our educational system. In general, students are taught from early years to learn facts, and on tests they are right or wrong in answering questions about those facts. There is precious little attention paid to dealing with uncertainty, despite the fact that everyone must deal with uncertainty continually in life, both for major decisions and minor everyday activities. In terms of a formal treatment of uncertainty, few students are exposed to probability except in a very limited sense.

With this background, many university students find probability and statistics to be a foreign way of thinking for them. This is not helped by the fact that with very few exceptions, basic statistics courses are taught with a frequentist orientation. Because the frequentist orientation is only able to address the questions of primary interest indirectly rather than directly (e.g., with probabilities that are not conditioned on the data), it is harder for students to understand and appreciate statistical analysis. This is probably an important reason why, to the average student, statistics courses are viewed as very difficult and, indeed, something to be feared. The most common response when I tell someone I teach statistics is "That was the course that gave me the most trouble in school."

Basic statistics, of course, sets the stage for future courses, which are designed to build on the initial course. In most cases, only those who place a considerable amount of emphasis on statistics get to the point where they are exposed to a serious Bayesian course. This is a very small proportion of statistics students. Many of those who actually do statistical analysis have relatively minimal statistics training, perhaps a statistics course or two as a "tools" course related to their primary field of interest. Furthermore, training is related to the way clinicians read clinical journals and to the historical segregation of clinical efficacy from decision making.

Very little has been done to make Bayesian methods accessible to a beginning audience. A notable exception is Berry (8). The typical text for the beginning course, however, follows the frequentist tradition and pays at most a bit of lip service to Bayesian analysis. In a recent series of articles with discussion (2;9;25), the pros and cons of teaching an elementary statistics course from a Bayesian perspective are debated.

The limited use of Bayesian methods in elementary statistics courses is in part due to tradition. But another important factor is ease of use. In general, although the underlying ideas behind the frequentist theory may be hard to grasp, the theory does provide relatively simple rules of thumb that are easy to apply in many situations. Rigid, all-purpose rules such as "reject the null hypothesis if the p value is less than .05" lead to a simple, plug-in-the-formula approach that grinds out numbers that are widely accepted (if not really widely understood).

It also helps that easy-to-use software is widely available for frequentist methods. This means that the analyst does not have to actually plug numbers into the formula, but can let the computer do all the work. The software can take a data set and spit out a dizzying array of summary statistics, p values, and confidence intervals.

The Bayesian approach, on the other hand, is more difficult to use. It requires more inputs, and it tends not to use simple rules of thumb. Instead, there is a need for harder thinking, both in terms of the inputs and in terms of the implications of the outputs, which in Bayesian analysis consist of probability distributions for parameters and future outcomes. But is that necessarily bad? Berry (9, 242) notes that "Bayesian statistics is difficult to the extent that thinking is difficult." Hard decision-making problems, after all, deserve serious thought.

Another reason Bayesian analysis is not used more in practice is that not enough good "role models" of successful Bayesian analyses are readily available. In contrast, there are lots of frequentist applications to provide templates to help the practitioner in his or her own analysis. There are, to be sure, examples of Bayesian applications in important problems, but they are not sufficiently numerous and they often appear together in books of Bayesian applications instead of being spread throughout the literature (10;17) (see Stangl and Berry [33] for other references).

The final practical issue of concern relates even more to tradition than do training and ease of use. It is the ready acceptance of frequentist analyses. The frequency paradigm is the accepted paradigm, which means that there are lots of people who follow this paradigm and have a vested interest in the paradigm. This ready acceptance includes acceptance of work by fellow scientists, journal editors, regulators, and even our legal system. This means that it is not only easier to implement a frequentist analysis because of readily available software and general understanding of simple frequentist rules of thumb, but it also may require jumping through fewer hoops to convince others that your analysis is correct and appropriate.

Journal editors and referees not familiar with or supportive of the Bayesian approach may be disinclined to recommend acceptance of papers using this approach. Moreover, some editors, even those willing to consider Bayesian analyses, may view such analyses as supplementary to standard frequentist measures. *Annals of Internal Medicine*, for example, has information for authors on the World Wide Web. Here is a quote from such information, dated July 6, 1999 (3): "If Bayesian methods are used as an *adjunct* to frequentist approaches, . . ."(emphasis mine). The advice they go on to give authors is good advice, but the implication is that a Bayesian analysis without some frequentist analysis is not acceptable. And this is from a journal that recently published a strongly pro-Bayesian editorial (13) and two serious Bayesian articles (18;19).

Regulators may have explicit guidelines requiring frequentist measures, and even in the absence of such guidelines there may be a tendency to view the use of Bayesian analysis in studies submitted to regulatory agencies to be somewhat risky. Spiegelhalter et al. (32,412) state that "the pharmaceutical industry tends to follow the lead of the regulatory bodies and is less likely to change its practices unless the regulatory agencies actively encourage it." Perhaps things are changing and such encouragement is beginning to occur; Stangl and Berry (33) claim that "thinking in regulatory agencies is at the forefront of Bayesian innovation—some divisions in the FDA actually encourage taking a Bayesian perspective." Structural changes in healthcare systems are forcing decision makers to be more pragmatic, more aware of uncertainties, and more willing to embrace methods that speak directly to their decision problems.

In the litigious society in which we live, healthcare decisions always have the possibility of winding up in a courtroom. Bayesian methods ought to have an advantage in that setting because they answer the right questions and agree with common sense. However, courts have not always been willing to accept probabilities as evidence, and under detailed probing Bayesian methods are harder to explain to a lay audience. Kadane and Schum note that:

> Though the word *probability* occurs with great frequency in legal cases and treatises, both courts and legal scholars have not always been enthusiastic about incorporating into legal rules and procedures what probabilists and others have learned about probability calculations and judgments. (22,60)

The primary outputs from Bayesian analyses are, of course, probabilities. I have been involved in probabilistic risk analyses of environmental issues where attorneys for a regulatory agency expressed doubt that such analyses would be admissible in legal proceedings.

The issue of acceptance falls within a sociopolitical context in which objectivity is very appealing. Much of the presentation of data analysis is a matter of persuasion: persuading fellow scientists that the conclusions are valid, persuading journal editors and referees that the results are deserving of publication, persuading regulators to approve new procedures, persuading judges and juries that analyses are valid, and persuading the general public that it is desirable to follow certain regimens. In such persuasion, arguments are more convincing if those being persuaded feel that the arguments are backed up by so-called objective analysis. Here is where the smokescreen of alleged objectivity of frequentist methods comes into play. This harkens back to the philosophical arguments of a few decades ago but is probably more reflective of tradition and force of argumentation than of deep philosophical concerns.

While the desire for objectivity seems to be deeply ingrained in our culture, especially in science and in public policy and politics, practical issues provide a greater impediment to the use of Bayesian procedures. Most statistics training, especially at the basic and intermediate levels, has emphasized the frequentist approach for a long time. Changing this is not easy, since it means creating more options in terms of textbooks, materials, and software for use in teaching. It also means retraining many instructors. It is always easier to stick with the status quo than to make major paradigm shifts. Similarly, most statistical practitioners have invested a lot of time and effort in learning to perform frequentist analyses. These analyses are found acceptable, so what is the incentive to shift gears and move to a new, more demanding form of analysis that runs the risk of less acceptance?

WHAT'S A BAYESIAN TO DO?

The picture painted above suggests that there are serious obstacles to greatly increased use of Bayesian methods in healthcare decision making. Although the Bayesian approach appears to have substantial advantages, the road to greater use of this approach in day-to-day statistical practice appears to be a long and hard one. Is there hope? Will we see a “Bayesian 21st century” in healthcare decision making? I would like to think so, although things cannot be changed overnight.

Since the Bayesian approach really represents a different, and better, paradigm, it is important to start a more-or-less exclusive emphasis on this approach with early statistical training. To my mind, too much time has been spent by Bayesians on reconciliation; in other words, trying to show how Bayesian techniques are consistent with frequentist techniques under certain conditions. The Bayesian approach should not be viewed merely as a supplement to the frequentist approach or as a justification for frequentist procedures.

With that spirit in mind, what might be done to help Bayesian analysis catch on in healthcare decision making and other fields of application? Below are some ideas along these lines, which follow directly from the previous discussion. They sound relatively straightforward but are by no means trivial to implement.

1. *More materials should be developed for Bayesian training in basic statistics courses, including short courses as well as regular university courses*. There are many excellent texts for more advanced audiences, but most basic statistics texts follow the traditional paradigm. A notable exception is Berry (8), and we need more books of this type, along with software appropriate for an elementary Bayesian course and other accompanying materials. Two examples of software developed for teaching introductory Bayesian statistics are First Bayes (27) and a set of Minitab macros developed by Albert (1); however, further software development is needed. Also, the importance of “training the trainers” should not be forgotten; instructors of basic courses need to understand Bayesian methods and how to present them effectively to students.
2. *Efforts should be devoted to developing easier-to-use Bayesian software for practitioners as well as students*. Advances in computing power and computing procedures have greatly expanded the scope of problems that can be handled computationally within the Bayesian framework. Software such as BUGS and WinBUGS (17) is available for such computations, but it is not readily accessible to modelers other than those already quite sophisticated in Bayesian methods and related computational aspects. Training on such software would help, but what is also needed is other software that is appropriate for less sophisticated users. A good start would be easy-to-use software for commonly used models that do not require advanced computational procedures. This could

be stand-alone software, but it might be more readily accepted if it is added to existing statistical software (e.g., SAS) that is widely used. Extensions to more complicated models will be more difficult and can follow later.

3. *The primary additional input that is needed in Bayesian analysis is a prior distribution*. As a result, better procedures are needed for choosing prior distributions in practice. The main focus in this regard should be on careful assessment of the decision maker's own prior distribution. Spiegelhalter et al. (32) call this a clinical prior and note that sources of evidence for such a prior include information from previous studies and clinical opinion. The latter could be the decision maker's own judgments or those of experts consulted by the decision maker. Another type of prior distribution frequently encountered is a reference prior (7), or diffuse prior, intended to represent minimal prior information. Other possibilities are skeptical priors, which might be particularly appropriate for regulatory authorities (32), enthusiastic priors (32), or extreme priors that provide bounds (5). Standard procedures for sensitivity analysis should be part of the package so that users can investigate the sensitivity of results to the choice of the prior distribution or to other modeling choices. The bottom line here is that users should be able to "let the data speak for themselves" when appropriate but should also be able to include any other available information and to understand the sensitivity of the results to the various inputs. Making all of this as easy to do as possible is important.
4. *The development of standard methods for displaying and communicating the output of Bayesian analysis would be very helpful*. This would provide a template for the creation of teaching materials and software, and it would encourage greater acceptance by setting standards for Bayesian presentation. Useful output might include graphs of likelihood functions, posterior distributions, and predictive distributions, along with various summary measures. The focus of Bayesian analysis is on probabilities, and its output is thus well suited to graphical presentation, with options for calculating any specific probabilities or summary measures of interest. A good Bayesian analysis makes any modeling assumptions clear, might include results for more than one prior distribution, and considers the sensitivity of the results to possible variations in the model (where the model includes both the likelihood function and the prior distribution).
5. *Better connections between statistical analysis and decision making are needed*. In working on cases it is easy for students to get caught up in the details of statistical data analysis and inference to the point where they forget that such analysis was originally motivated by a decision-making problem. The development of modern Bayesian methods was inspired primarily by the suitability of such methods for decision making and was done in large part under the rubric of statistical decision theory. In the past two decades, more emphasis in the Bayesian community has been placed on the statistics and less on the decision. This is reflected by undue emphasis on a hypothesis-testing mentality. A much better job could be done of capitalizing on the suitability of Bayesian methods for *decision making*.
6. *One way to build connections to decision making is to demonstrate the advantages of Bayesian methods in important healthcare decision-making problems*. This would require not just statistical analysis, but also a detailed structuring of the underlying decision-making problem, including a careful assessment of consequences and utilities. Too many analyses use standard loss functions that often are not very suitable for the real problem at hand. With the decision problem and preferences carefully modeled, the value of Bayesian procedures to handle the uncertainty side of the problem becomes clear.
7. *More generally, it would be useful to develop a large set of test cases for important healthcare problems using Bayesian analysis*. This would provide examples for use in teaching Bayesian methods and would provide "role models" for analysts. I prefer to think that showing how effective a good Bayesian analysis can be would suffice, but given the ingrained tradition, comparisons with frequentist analyses should be helpful.
8. *There is so much focus on Bayesian versus frequentist comparisons that we lose sight of the value of debate among Bayesians*. Different Bayesians may approach a given problem from different perspectives. Therefore, evaluations of Bayesian analyses and Bayesian versus Bayesian comparisons can be very instructive in terms of Bayesian modeling. Is there a debate about the likelihood function? Is there a debate about the prior distribution? What are the implications for the posterior

distribution and, perhaps even more importantly, for predictive probabilities? This emphasizes the importance of providing output in terms of such probabilities and suggests that more stress should be placed on evaluating these probabilities (34;35).

9. *Finally, there is the issue of "selling" the advantages of Bayesian methods*. Much as we may feel that these methods will sell themselves because of their advantages, changes will not happen rapidly without persuasion. Bayesians need to convince consumers of statistics that they should want posterior and predictive probabilities, not p values and confidence intervals. This is a question of educating decision makers about the advantages of Bayesian methods, not just educating statistical analysts.

As noted above, these ideas will not be easy to implement. In moving forward to help Bayesian analysis catch on in healthcare decision making and other fields of application, we should carefully consider what it is that we would like to see in a Bayesian world. Do we want a simple Bayesian plug-in formula approach in which some modeling issues as well as computational issues are hidden in a black box so the user doesn't need to worry about them? That would surely hasten the adoption of Bayesian techniques. However, it would also encourage relatively mindless application of those techniques. Serious statistical modeling cannot be totally hidden in a black box except for very simple models, and I do not think we would want it so hidden. Part of the advantage of a shift to Bayesian procedures is that they make things more transparent (if more difficult) and encourage the user to think about modeling options. On the other hand, too much encouragement along these lines may dissuade users from shifting to the Bayesian approach, at least until improvements in Bayesian training lead to a population of users who understand and appreciate what the Bayesian approach can do for them.

REFERENCES

1. Albert J. *Bayesian computation with Minitab*. Belmont, CA: Duxbury Press; 1996.
2. Albert J. Teaching Bayes' rule: A data-oriented approach. *The American Statistician*. 1997;51: 247-253.
3. *Ann Intern Med* information for authors. Available at: www.acponline.org/journals/resource/ ino4aut.htm. Accessed July 6, 1999.
4. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *American Scientist*. 1988;76:159-165.
5. Berger JO, Sellke T. Testing a point null hypothesis: The irreconcilability of significance levels and evidence. *J Am Stat Assoc*. 1987;81:112-139.
6. Berger JO, Wolpert RL. *The likelihood principle*, 2nd ed. Haywood, CA: Institute of Mathematical Statistics; 1988.
7. Bernardo JM, Smith AFM. *Bayesian theory*. Chichester, England: Wiley; 1994.
8. Berry DA. *Statistics: A Bayesian perspective*. Belmost, CA: Duxbury Press; 1996.
9. Berry DA. Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*. 1997;51:241-246.
10. Berry DA, Stangl DK, eds. *Bayesian biostatistics*. New York: Marcel Dekker; 1996.
11. Berry DA, Stangl DK, eds. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker; 2000.
12. Carlin JB. Meta-analysis for 2 × 2 tables: A Bayesian approach. *Stat Med*. 1992;11:141-158.
13. Davidoff F. Standing statistics right side up. *Ann Intern Med*. 1999;130:1019-1021.
14. DuMouchel W, Harris JE. Bayes methods for combining the results of cancer studies in humans and other species. *J Am Stat Assoc*. 1983;78:293-315.
15. Eddy DM, Hasselblad V, Shachter R. *Meta-analysis by the confidence profile method: The statistical synthesis of evidence*. Boston: Academic Press; 1992.
16. de Finetti B. *Theory of probability*, vol. 1. trans. Machai A, Smith A. London: Wiley; 1974.
17. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov chain Monte Carlo in practice*. London: Chapman & Hall; 1996.

18. Goodman SN. Toward evidence-based medical statistics, I: The P-value fallacy. *Ann Intern Med*. 1999;130:995-1004.
19. Goodman SN. Toward evidence-based medical statistics, II: The Bayes factor. *Ann Intern Med*. 1999;130:1005-1013.
20. Heckerman D. *Probabilistic similarity networks*. Cambridge, MA: MIT Press; 1991.
21. Heckerman D, Horvitz E, Nathwani B. Toward normative expert systems, I: The Pathfinder project. *Methods Inf Med*. 1992;31:90-105.
22. Kadane JB, Schum DA. *A probabilistic analysis of the Sacco and Vanzetti evidence*. New York: Wiley; 1996.
23. Lilford RJ, Braunholtz D. The statistical basis of public policy: A paradigm shift is overdue. *BMJ*. 1996;313:603-607.
24. Lindley DV. A statistical paradox. *Biometrika*. 1957;44;187-192.
25. Moore DS. Bayes for beginners? Some reasons to hesitate. *The American Statistician*, 1997;51: 254-261.
26. Neath AA, Samaniego, FJ. On the efficacy of Bayesian inference for nonidentifiable models. *The American Statistician*. 1997;51:225-232.
27. O'Hagan A. *Instructions for using First Bayes*. Unpublished manuscript, University of Nottingham; 1996.
28. Raiffa H, Schlaifer R. *Applied statistical decision theory*. Boston: Harvard Business School; 1961.
29. Savage LJ. *The foundations of statistics*. New York: Wiley; 1954.
30. Schlaifer R. *Probability and statistics for business decisions*. New York: McGraw-Hill; 1959.
31. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical decision making. Boston: Butterworth-Heinemann; 1988.
32. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J R Stat Soc A*. 1994;157:357-416.
33. Stangl DK, Berry DA. Bayesian statistics in medicine: Where are we and where should we be going? *Sankhya, Series B*. 1998;60:176-195.
34. Winkler RL. Scoring rules and the evaluation of probabilities. *Test*. 1996;5:1-60.
35. Winkler RL, Poses RM. Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science*. 1993;39:1526-1543.