International Journal of Technology Assessment in Health Care, **17:1** (2001), 114–122.

CAN BAYESIAN METHODS MAKE DATA AND ANALYSES MORE RELEVANT TO DECISION MAKERS?

A Perspective from Medicare

Steven H. Sheingold
Health Care Financing Administration

Abstract

Decision making in health care has become increasingly reliant on information technology, evidence-based processes, and performance measurement. It is therefore a time at which it is of critical importance to make data and analyses more relevant to decision makers. Those who support Bayesian approaches contend that their analyses provide more relevant information for decision making than do classical or "frequentist" methods, and that a paradigm shift to the former is long overdue. While formal Bayesian analyses may eventually play an important role in decision making, there are several obstacles to overcome if these methods are to gain acceptance in an environment dominated by frequentist approaches. Supporters of Bayesian statistics must find more accommodating approaches to making their case, especially in finding ways to make these methods more transparent and accessible. Moreover, they must better understand the decision-making environment they hope to influence. This paper discusses these issues and provides some suggestions for overcoming some of these barriers to greater acceptance.

Keywords: Bayesian statistics, Reimbursement, Healthcare policy

The breaking of a wave cannot explain the whole sea.

—*Vladimir Nabokov*

"I couldn't afford to learn it," said the Mock Turtle with a sigh, "I only took the regular course." "What was that?" inquired Alice.
"Reeling and writhing, of course, to begin with," the Mock Turtle replied, "and then the different branches of Arithmetic—Ambition, Distraction, Uglification, and Derision."

—From *Alice in Wonderland*

We are clearly at a time at which making data and analyses more relevant to healthcare decision makers is critical. The most important trends in managing care and developing purchasing strategies involve the collection, analysis, and dissemination of information. Indeed, the practice of medicine is likely to be revolutionized by the convergence of two developments: information technology and evidence-based medicine (4). It is possible, therefore, that those who believe that the application of formal Bayesian analyses will lead to improved decision making may have an opportunity to advance their cause. Their approach to make Bayesian analyses more accepted and accessible to mainstream decision making must be carefully considered, however.

It is quite apparent that classical, or frequentist, statistical methods are well accepted for designing clinical research and other analyses relevant to decision makers at all levels of healthcare financing and delivery. Bayesians, who believe their analyses provide more relevant information for decision making than do frequentist methods, seem in disbelief that the latter methods are so well accepted. There is a parallel here to the practitioners of cost-effectiveness analyses, who often are surprised that their results are not formally incorporated into reimbursement decision making. In both cases, these groups may not fully understand the internal and external constraints on decision-making processes that may limit their ability to utilize all information that is deemed to be relevant.

In *The Art of War*, Sun Tzu wrote, “If you know your enemy and know yourself, you need not fear the results of one hundred battles.” Many supporters of Bayesian statistics seem to know themselves and their analyses well, but may be somewhat less knowledgeable about healthcare decision making and the processes that might potentially use these data. In this paper, I discuss reimbursement decision making, its current environment, and the implications for greater acceptance of Bayesian methods. I also make several suggestions about strategies that might be employed to enhance the process of acceptance. The paper is written from the perspective of a decision maker who uses statistics but who is not an outstanding scholar in either Bayesian or frequentist methods.

BAYESIAN VERSUS FREQUENTIST ANALYSES FOR DECISION MAKING: IS IT TIME FOR A PARADIGM SHIFT?

In his book *Against the Gods: The Remarkable Story of Risk*, Bernstein credits advances in risk management—the ability to define what may happen in the future and to choose among the alternatives—with being the missing ingredient that has propelled science and enterprise into the world of speed, power, instant communication, and sophisticated finance (1). Risk management techniques guide modern societies over the vast range of decision making, including all areas of health care. Bernstein attributes these advances to a handful of innovators who made discoveries about the nature of risk and the art and science of choice. He identifies Reverend Thomas Bayes as being prominent among this small group of innovators due to his striking advance in statistics, which demonstrated how to make better decisions by blending new information into old information. Nearly 250 years later, it is difficult to escape the fact that much of the statistical training and information provided to decision makers seems biased against the formal use of Bayes’ striking discovery. Thus, perhaps it is also understandable, although still surprising, that the arguments made by modern Bayesians seem to contain a high degree of animosity toward their classical counterparts.

In all fairness, however, most decision makers are not pure frequentists. We all have a healthy skepticism of total reliance on p values and confidence intervals, and evaluate new data carefully in light of our past knowledge and experience. New technologies are not approved or denied for reimbursement solely based on the acceptance or rejection of a null hypothesis. Moreover, beneficial or harmful effects of services are not dismissed only because a p value exceeds .05. Indeed, new and different results are carefully scrutinized, although usually not with formal Bayesian methods, when they seem to contradict our prior knowledge.

Nonetheless, those who believe in formal Bayesian analyses tend to cast this issue as nothing less than a battle for the heart and soul of healthcare decision making. They seem sure that their analyses provide a richer array of information that is more relevant and useful to decision makers, and call for a paradigm shift away from frequentist methods (7).

Thus, it is worth considering from the point of view of decision makers whether such a shift would be useful and how it might occur. We should remember that outcomes research represented a fundamental shift in the emphasis of health services research over the past decade. In explaining this phenomena, Wennberg (10) referenced Kuhn's *The Structure of Scientific Revolutions* (7). That is, when current scientific procedures are no longer adequate to explain anomalies in theory or experimental evidence, a new scientific approach emerges and new techniques are applied. It should be noted, however, that outcomes research was a well-accepted refocusing of health services research rather than a substantial paradigm shift and did not meet with much resistance. A shift in emphasis from frequentist to Bayesian methods for research and decision making will face more significant obstacles. In this regard, it is worth briefly revisiting Kuhn's work in order to anticipate these obstacles.

Kuhn suggested that paradigms are accepted examples of actual scientific practice—including law, theory, application, and instrumentation together—which provide models from which spring particular coherent traditions of scientific research. Those whose research is based on a shared paradigm are committed to the same rules and standards for scientific practice. These traditions have been passed on through education and subsequent exposure to the literature. Accompanying such paradigms is the strong belief that they account quite successfully for most observation and experiments. Therefore, there is natural resistance to change due to new discoveries. One source of such resistance is particularly relevant to Bayesian methods; that is, that further scientific development from a comfortable paradigm requires construction of more elaborate equipment, development of esoteric vocabulary and skills, and a refinement of concepts that lessens resemblance to common-sense prototypes. Bayesian methods do require different applications of methods, can result in a more extensive set of parameter estimates, require different interpretations of results, and necessitate the purchase of new computer software.

On the other hand, rigidity with regard to the methods associated with a paradigm may also be an agent of change—it is against those methods and their expected results that an anomaly can be demonstrated. An anomaly is an outcome that could not be predicted by the current paradigm. The recognition of the anomaly eventually results in a new paradigm replacing an old one. The preceding discussion does not at all do justice to Kuhn's work, which should be required reading for those involved in the Bayesian initiative. It does, however, provide a background with which to consider resistance to accepting Bayesian methods and how such resistance might be overcome.

REIMBURSEMENT DECISIONS FOR MEDICARE

The trends toward managed care and value-based purchasing in health care have made information systems, data, and analyses all the more critical to decision making throughout the health industry. Health plans and providers are increasingly held accountable for their performance in terms of quality or value. Patients have faced increasing financial accountability for choosing their plans and providers based on cost and quality information. Thus, the ability to conceptualize measures of performance, collect and analyze data to implement these measures, and compare them to standards based on the best available scientific evidence has become essential.

For Medicare and private health plans, the analysis of medical evidence has become particularly important to reimbursement decisions for new technologies. Reimbursement decisions include the coverage decision—whether a service can be included in the benefit package and under what conditions—and at what rate the service can be paid (the payment decision). For the Medicare program, the coverage decision is most critical since payment methodology for many services is predetermined by law. The methods of making coverage determinations have changed significantly in recent years. Once driven by medical

https://doi.org/10.1017/S0266462301104101 Published online by Cambridge University Press

opinion and expert consensus, coverage processes have now adopted evidence-based decision making; that is, decision makers rely on systematic analysis of medical evidence to determine the strength of the conclusions that can be drawn regarding risks and benefits of new technologies (8). Careful scrutiny of study design and analytic methods for providing medical evidence is now the rule for most major decisions. Indeed, high impact and potentially controversial decisions usually require formal analyses, such as health technology assessments, from independent entities.

While the systematic review of data regarding the potential risks and benefits of a technology to patients provides the analytic core of the decision-making process, a number of other considerations are also critical. For almost any service, the available data will have strength and weaknesses that leave gaps in our knowledge about particular risks and benefits. Thus, other factors enter the decision such as patient prognosis, availability of therapeutic or diagnostic alternatives, likelihood of better data becoming available if a given decision is made, cost of obtaining better data, benefits and harms to patients, and potential program costs. Essentially, coverage of a service is based on a judgment of the risks and benefits to the Medicare program and its beneficiaries of making that decision given the available data. Therefore, at least on an informal basis, the process is consistent with Bayesian analyses.

OBSTACLES TO INCREASED USE OF BAYESIAN METHODS FOR REIMBURSEMENT DECISIONS

Bayesians should be somewhat encouraged that reimbursement decisions are not simply knee-jerk responses to classical hypothesis testing. Indeed, basic Bayesian principles are already incorporated into the process. Nonetheless, for reasons described below, there are likely to be obstacles to increased use of formal Bayesian analyses.

Many Who Participate in Reimbursement Decision Making Are Not Trained in Formal Bayesian Method

It is difficult to avoid the fact that basic statistics texts and courses are dominated by classical methods of inference. Although Bayesian methods typically appear somewhere in the standard statistics course, it is often as a passing thought. It is not surprising, therefore, that many senior analysts that support decision making are not well acquainted or comfortable with formal Bayesian analyses. In preparing my thoughts on this subject, I spent a considerable amount of time familiarizing myself with the methods and analyses. I am still quite uncomfortable with interpreting these results to inform the decision-making processes in which I participate. I conducted an informal and unscientific poll of colleagues with considerable skill and experience in research and statistics. Of the 15 I questioned, 14 said they would not be knowledgeable enough to use or interpret Bayesian analyses. Thus, regardless of the potential value of such analyses, it may be difficult to immediately convince decision makers they should demand that policy-related research include Bayesian analyses.

Current Presentations of Bayesian Methods Do Not Make Them Accessible or Understandable

> "It seems very pretty," she said when she finished it, "but its rather hard to understand!" (You see she didn't like to confess even to herself, that she couldn't make it out at all.) "Somehow it seems to fill my head with ideas—only I don't know exactly what they are!
>
> —Alice in *Through the Looking Glass*

Alice's thoughts after her first reading of Jabberwocky are much the same as my own after listening to many of the Bayesian presentations to which I have attended recently.

If I correctly understand the objective of the Bayesian initiatives under way, it is to make statistics and analyses more relevant to decision makers. Unfortunately, these presentations, and available publications, do little to make these methods transparent, accessible, and understandable to the uninitiated. This may be because the knowledgeable presenters assume too much about their audience's ability to comprehend the material easily, or that Bayesian analyses are more technically difficult to apply there are classical methods, or both. In either case, the current methods of delivering the message will not make decision makers, or researchers who provide them with analyses, comfortable with the Bayesian approach. The computational complexities and lack of easily usable software for these analyses compound these problems.

In addition, the trend toward evidence-based methods is already requiring retraining and new staffing for reimbursement decision making. The previous approach of basing coverage decisions on expert opinion and consensus required staff experienced in the details of the Medicare program and claims administration, along with some clinical input. The evidence-based methods now employed have required staff and advisers with greater sophistication in systematic reviews, clinical trial methods, research methodologies, cost-effectiveness analyses, and meta-analyses. At a time when all health plans face pressure to reduce costs and improve quality of care, the value of the evidence-based methods, and the need for accompanying changes in the staffing and decision processes, have been readily accepted. The ability to incorporate Bayesian analyses in this process would require further training. Unless public and private health plans can be convinced of significant added value to Bayesian analyses, committing additional training resources is not likely in the near future.

The Environment for Reimbursement Decision Making May Reduce the Value of Additional Information

The trend toward evidence-based medicine has not been free of controversy (8). Medicare and private health plans have adopted these methods as one way of bridging the cost/quality trade-off, as a tool for better directing expenditures to effective services and away from harmful or ineffective services. The result has been higher evidence thresholds for new technologies, a greater number of negative or limited reimbursement decisions, and requirements for better research designs. Providers, manufacturers, and patients, for a variety of reasons, desire more rapid diffusion and reimbursement of new services than might be consistent with the evidence-based approach. The result is often a very confrontational environment for decision making in which stakeholders resort to legislative and legal remedies, as well as the media, when they disagree with a policy. Indeed, researchers as well as policy makers are subject to attack when stakeholders disagree with their results (5). There is tremendous pressure to make decisions differently than might be consistent with a rational, data driven-approach. Thus, constraints exist that limit health plans' ability to use all information that might be deemed valuable for optimal decision making.

Within this environment, Medicare and many private health plans have developed criteria, processes, and policies to balance the need to make decisions based on high-quality evidence with the external pressures to make rapid, positive decisions. Strategies have included using public advisory committees, implementing open and interactive processes, and developing alternatives to the simple yes-or-no decision. Most importantly, developing the means to explain science-based decisions to a variety of audiences has become critical. Thus, with any methods used to inform decisions, there needs to be a comfort level with not only the value of the information, but also how easily they can be explained and justified in a volatile environment.

There are at least three reasons, therefore, that decision makers might not immediately see the value of Bayesian methods for reimbursement decisions. First, while the results may

https://doi.org/10.1017/S0266462301104101 Published online by Cambridge University Press

be of interest, explaining the methods that produced them to the various stakeholders would be seen as very difficult. While there are frequent disagreements over the quality of data and study designs among stakeholders, for better or worse, the *p* value provides a familiar and comfortable concept among all the parties (although the *p* value is often misinterpreted). Second, given the compromises and constraints imposed by the current environment, not all interesting information may be relevant or usable for making a decision. If one were to describe the pressure on decision makers in terms of classical statistics, it would be to allow for a large alpha error (accept a greater risk of covering a technology as effective when it is not) and eliminate the beta error (not reimburse a technology when it is effective). Accordingly, if a recognized study finds a service to produce better outcomes within the bounds of standard tests of significance (e.g., $p < .05$), it is questionable whether the array of actual probability estimates that could be produced by Bayesian analyses would modify the decision that the service was effective. For example, the results of the GUSTO trial seemingly erased doubt that Medicare's decision to allow certain diagnosis-related group (DRG) rates to increase as a result of tissue plasminogen activator (TPA) use was correct. A recent Bayesian reanalysis of GUSTO data suggested a more cautious conclusion, finding that the probability of TPA being more effective than streptokinase ranged from .05 to .999, depending on prior beliefs (2). While interesting, these data would not likely result in new consideration of Medicare's reimbursement decision unless a very persuasive case was presented that the priors associated with the very low end of the probability range were superior to the alternatives. The third and related reason that the potential value of Bayesian analyses is not yet apparent to decision makers follows from Kuhn's work. Bayesians have yet to demonstrate an anomaly related to the use of frequentist methods in policy analysis. The adoption of evidence-based methods for coverage decisions was possible, in part, because anomalies could be demonstrated. Specifically, there is a growing list of services that were accepted into medical practice based on inadequate data that were later shown to be ineffective or harmful by better study designs (e.g. randomized trials).

Prior Probabilities and the Quality of Evidence

One of the issues that arises in the Bayesian/frequentist debate is whether the use of prior probabilities adds too much subjectivity to analyses. Depending on which side is making the argument, the formal inclusion of prior beliefs can be a significant benefit or a significant liability. Without addressing the relative merits to each side of the argument, it is worth noting that this debate does have a particular relevance to reimbursement decisions. Over time, health plans have been able to move the debates over coverage of new technologies toward the adequacy of data for answering questions about the risks and benefits to patients. Responsibility for providing better data has been shifted to the proponents of technologies, and data from inadequate sources greatly discounted—even when backed by the opinion of national and international experts. Thus, Bayesian methods may be looked at suspiciously by decision makers as a way of undoing progress—that is, providing a method of more formally introducing prior data whose value they have sought to minimize.

The reanalysis of clinical trial results with prior data may also raise questions among decision makers. As the randomized clinical trial has grown in importance as a source of data for practice and reimbursement decisions, so have the ethical and patient choice issues involved. To the extent there is a prior belief that a procedure of interest is either superior or inferior to a comparator, the randomized design becomes controversial. For example, the Health Care Financing Administration's decision to reimburse lung volume reduction surgery only when provided under the protocol of a randomized clinical trial sponsored by the National Heart, Lung, and Blood Institute has been criticized as unethical because some believe the surgery has already been demonstrated as effective (9). Initiating and

https://doi.org/10.1017/S0266462301104101 Published online by Cambridge University Press

maintaining such a trial requires the strong judgment by researchers and policy makers that existing data is inadequate to draw conclusions about the effectiveness of the procedures. Thus, any method of reanalysis of the trial data that didn't assign zero weight to the pretrial data would also be viewed with skepticism. Moreover, the protocols for most clinical trials are developed based on analysis of existing data and to some degree already incorporate current knowledge. It is likely that all of these issues can be handled explicitly with Bayesian analyses, but again, the methods are not transparent enough yet for a solid understanding of how this might occur.

OVERCOMING RESISTANCE: SOME SUGGESTIONS FOR THE BAYESIAN INITIATIVE

Despite the obstacles described above, I believe there is value in, and perhaps eventual success for, the initiative to promote the use of Bayesian methods. The following are strategies that might be followed to make this initiative more acceptable and understandable to decision makers.

Improve Methods for Communicating Bayesian Methods

If the Bayesian initiative is to meet with any success, those involved must improve their ability to teach these methods and convey their value to a variety of audiences. The presentations I have attended seem directed at those already adept at the methods and convinced of their usefulness. The content of these presentations must be modified to capture the interest of those less technically sophisticated and those who may be skeptical. For example, one approach that has been used is to compare *p* values and power levels to medical diagnostic tests and demonstrate why they provide necessary but not sufficient information (3). Considerable resources should be devoted to developing strategies for communicating these methods to policy makers, the clinical and health services research communities, and other stakeholders. Moreover, accessible hands-on workshops to provide training to decision makers and researchers would likely be essential. If Bayesian analyses can be properly packaged and communicated, decision makers will realize that their processes already employ the basic concepts in terms of evaluating new data in light of past knowledge and considering risks and benefits of making decisions conditional on existing data. In that case, rather than needing to promote the introduction of their conceptual framework for decision making, Bayesians can focus on formalizing concepts that are already used informally.

Develop More Constructive Ways to Contrast Bayesian and Frequentist Methods

He who digs a hole for another may fall in himself.

—*A Russian proverb*

A great truth is a truth whose opposite is also a great truth.

—*Neils Bohr*

The approach taken by the Bayesian proponents seems to cast this issue as a choice between right or wrong—that is, Bayesian or frequentist methods. They assert that the former provides the right data for decision makers, the latter the wrong information. I would suggest that this approach is counterproductive in a number of ways. For all the reasons discussed, decision makers and many researchers have an established comfort level with frequentist methods. Attacking those methods as inappropriate for decision making will likely cause a defensive reaction, to the extent it raises any awareness at all. Decision makers may divert more energy to defending the use of existing methods rather than to considering the potential value to them of the Bayesian methods. Moreover, since the decision-making environment is already characterized by controversy, new sources of disagreement would not be welcomed. Indeed, what might be seen as a dispute among academics would likely be seen as less relevant and be discounted or ignored.

My suggestion, therefore, is to advocate the value to decision making of enhanced information; that is, recognize that *p* values and related measures are accepted input for decision making that will not easily be replaced and emphasize the additional value that Bayesian methods can add to the process. One way of doing this would be to present a continuum of results for a given research question. The array of results would first include the familiar information from standard hypothesis testing. Next, relevant probabilities derived from Bayesian analysis using noninformative or least imposing priors should be presented. Finally, the same probabilities using stronger priors could be added and contrasted. A comprehensive discussion of the meaning and potential implications of these results for decision making would also be critical. Most importantly, the clinical and methodologic rationales for choosing the various prior distributions should be described in detail to assist policy makers in assessing the results. Sensitivity analysis is useful when clear meaning can be associated with the range of assumptions and results. Simply using a range of priors that do not have solid, well-understood underpinnings reduces the value of sensitivity analysis.

Demonstrate Anomalies in Past Decisions

For the Bayesian initiative to be successful, it will at some point need to demonstrate, rather than simply assert, that the analyses provide more useful information for decision making. One way to do this would be to reanalyze past decisions, particularly when it has become acknowledged that they were incorrect. For example, there are now numerous examples of services that have diffused into medical practice that were later shown to be ineffective or harmful to patients. Could a Bayesian analysis of the data available at those times have resulted in different reimbursement or practice decisions? While this question cannot be answered with certainty, a current reanalysis of those data might be very suggestive of whether a different decision would have been made. Such analyses may capture the attention of decision makers enough to create interest in how Bayesian methods could affect current policy issues.

CONCLUSION

As healthcare decision making becomes increasingly information based, there are likely to be greater opportunities for new methodologic paradigms to emerge. Initiatives to promote the use of Bayesian methods will face several obstacles in the short run, but with well-developed education strategies these efforts may realize some success over time. Some lessons can be taken from the evolution of cost-effectiveness analyses in recent years. Ten years ago, the methods and applications of cost-effectiveness were thought to be so underdeveloped and nontransparent that they could not be considered in the policy arena. During this decade, several public and private efforts have been initiated to gain greater consensus and understanding of the methods, how results should be reported, and how results might be interpreted and applied. While we are not yet ready to make reimbursement decisions based on cost-effectiveness ratios, the debate has now shifted to how and when these methods can be used appropriately, rather than if they can be used at all. The Bayesian initiative will need to launch similar efforts among all interested parties to gain greater understanding and acceptance. A less careful and engaging approach is likely to relegate these methods to academic enclaves for some time to come.

REFERENCES

1. Bernstein P. *Against the gods*: *The amazing story of risk*. New York: John Wiley and Sons Inc; 1996.
2. Brophy JM, Lawrence J. Placing trials in context using Bayesian analysis: GUSTO revisited by Reverend Bayes. *JAMA*. 1995;273:871-875.
3. Browner WS, Newman TB. The analogy between diagnostic tests and clinical research. *JAMA*. 1987;257:2459-2463.
4. Cohen JJ. Higher quality at lower cost: Maybe there is a way. *Acad Med*. 1995;73:414
5. Deyo RA, Psaty BM, Simon G, et.al. The messenger under attack: Intimidation of researchers by special interest groups. *N Engl J Med*. 1997;336:1176-1179.
6. Kuhn TS. *The structure of scientific revolutions*. Chicago:University of Chicago Press; 1962.
7. Lilford RJ, Braunholtz D. The statistical basis of public policy: A paradigm shift is overdue. *BMJ*. 1996;313:603-607.
8. Sheingold SH. Technology assessment, coverage decisions and conflict: The role of guidelines. *Am J Managed Care*. 1995;4:SP117-125.
9. U.S. Department of Health and Human Services, Report to Congress. *Lung volume reduction surgery and medicare coverage policy: Implications of recently published evidence*. Washington, DC: Government Printing Office; 1998.
10. Wennberg J. What is outcomes research? In Gelijns AC, ed. *Modern methods of clinical investigation*. Washington, DC: National Academy Press; 1990:33-46.

https://doi.org/10.1017/S0266462301104101 Published online by Cambridge University Press