International Journal of Technology Assessment in Health Care, **17:1** (2001), 9–16.

INTRODUCTION TO BAYESIAN REASONING

John Hornberger
Roche Pharmaceuticals and Stanford University School of Medicine

Abstract

Interest in Bayesian analyses has increased recently, in part as a response to policy makers wanting sound scientific bases for health technology assessments, and associated healthcare funding decisions. This paper provides a brief and simplified description of Bayesian reasoning. Bayes is illustrated in a clinical setting of an expert helping a woman understand the potential risk of passing on an inheritable disease (hemophilia) to her next child, based on disease occurrence in two living children. The illustration describes fundamental concepts and derivations, such as Bayes theorem, likelihood functions, prior probability, and posterior probability. A second illustration shows the use of Bayes for interpreting clinical trial results. The uncertainty in the clinical effect before and after the trial analyses has been completed is characterized by the Bayes prior and posterior probabilities, respectively. Techniques are also shown for estimating the potential loss (e.g., in lives lost) for making the wrong decision with and without knowledge of the trial results, an estimation that cannot be carried out using techniques of hypotheses testing associated with the frequentist school of statistics. Information from Bayes analysis then may be used to help policy makers decide, or justify, whether the analyses provides a sufficient basis for making a treatment recommendation, or whether there remains a need to request more information. Subsequent papers in this volume offer additional examples and clarification of the use of Bayes in clinical practice and in interpretation of clinical studies.

Keywords: Bayes, Medical decision making, Economics, Statistics

Life requires making decisions. The basis for hundreds of decisions made each day often is little more than one's personal experience, informed by common sense. However, some decisions are so important—the consequences of making the wrong decision so grave—that people and institutions have sensibly required more systematic evidence before a decision is made. Such it is with deciding whether a new medical technology (e.g., drug, procedure, or diagnostic) is safe enough, and effective enough, to be approved for widespread public use.

This paper outlines the fundamentals of Bayesian methods for helping clinicians and policy makers draw conclusions and make recommendations on key clinical/policy issues (1;3;11;12). We first present a simple example of the use of Bayesian method to assess the probability that a woman carries an X-linked gene for hemophilia. The second section outlines the intuition behind extending Bayes from the case of a single individual to inferences on the effect of a treatment on a population. The last section concerns use of Bayesian reasoning applied to interpretation of clinical trials.

The illustrations have been deliberately simplified to highlight the key elements of the approach. Other papers in this special section will present increasingly more complex examples and contrast Bayesian with frequentist statistical methods.

The author is grateful to three anonymous referees for their very thoughtful comments and suggestions.

DOES THIS WOMAN CARRY THE HEMOPHILIA GENE?

A woman has an appointment with a genetic counselor to assess the risk she might be a carrier of the gene for hemophilia.[1] Hemophilia is a rare disorder that follows simple Mendelian characteristics, i.e., it is expressed only in males who have inherited a copy of the gene on the X chromosome from their mother, and is a serious disorder that can result in frequent and severe hemorrhages into joints or the brain. The woman seeking counseling believes that the risk of her carrying the gene might affect her decision to have more children.

Knowing nothing more about her, the probability that she is a carrier is extremely low, but it is not equal to zero. Early in the interview, the counselor learns that the patient has a biologic (not adopted) brother who has hemophilia. Based on simple Mendelian inheritance pattern of genetic inheritance (X-linked recessive gene), the brother had to have gotten the gene from his mother. Assuming the history is reliable, the patient now has a probability of $^1/_2$ that she is a carrier. Define $\theta = 1$ if she is a carrier and 0 if not. The probability that she is a carrier, $p(\theta = 1)$, is equal to $^1/_2$. In Bayesian terminology, this is referred to as the "prior" probability of her being a carrier.

On further discussion, the counselor learns that the patient has two nonidentical sons, neither of who have hemophilia. How does this extra information affect the probability that she is a carrier? Let $i = 1$ or 2 index the two sons, where $x_i = 1$ if the ith son has hemophilia and 0 if not.

The probability of these two sons not having hemophilia, with respect to whether their mother is or is not a carrier for the gene, can be derived mathematically, and is referred to as the likelihood function. This function is conditional on two different levels of θ (one for if the mother is a carrier and one for if she is not a carrier) shown below:

$$f(x_1 = 0, x_2 = 0|\theta = 0) = (1)(1) = 1$$

$$f(x_1 = 0, x_2 = 0|\theta = 1) = (1/2)(1/2) = {}^1/_4 \qquad \textbf{(1)}$$

where the function f represents the probability of the sons not having hemophilia (x_1 and x_2 both equal to 0) "given" (denoted by the | mark) the mother is or is not a carrier ($\theta = 0$ or 1, respectively). If the woman is not a carrier, then there is no chance of her passing the gene to her sons and thus the probability for both sons not having hemophilia is 1. By contrast, if the woman is a carrier, then she has a probability of one-half of passing the gene on to each son. The probability of both sons not having hemophilia if she is a carrier then is the square of one-half, or one-fourth, because the probabilities of the children's genetic composition are independent of each other.

The probability she is a carrier given that her sons do not have hemophilia—called the "posterior" probability—can be estimated using Bayes rule. Specifically, the posterior probability is proportional to the prior probability and the likelihood function. For this specific problem, the posterior is calculated as:

$$p(\theta = 1|x) = \frac{f(x|\theta = 1)p(\theta = 1)}{f(x|\theta = 1)p(\theta = 1) + f(x|\theta = 0)p(\theta = 0)}$$

$$= \frac{(1/4)(1/2)}{(1/4)(1/2) + (1)(1/2)} = \frac{1/4}{5/4} = 0.20 \qquad \textbf{(2)}$$

The uncertainty about whether she carries the hemophilia gene has dropped from 50% to 20%, taking into consideration she already has had two sons without hemophilia.

In this instance, the calculation takes into account all of what is known about the problem. It represents, therefore, an example where critics are less able to argue that relevant

information is missing, as they might argue in a more complex example. Webster and colleagues (15) provide another meaningful genetic example beyond that of simple Mendelian inheritance, where they use Bayes rule to estimate the probability of a child having Hippel-Lindau disease, given that the child is observed to have an isolated ocular hemangioma.

BAYESIAN UNCERTAINTY AT A POPULATION LEVEL

We now look at inferences made at the population level instead of the individual level. Specifically, Bayes is shown in the situation where the conclusion is derived from analyses of data of patients taking an experimental treatment. Assume in the simplest case that the decision maker is considering the merits of continuing to study an experimental treatment, and that the judgment will be based on the treatment's effect on one clinical endpoint, the probability that the patient survives beyond a certain time period. For example, consider the case of a drug company having just completed a placebo-controlled phase 2 trial to determine the optimal dose for an experimental treatment. However, additional pivotal phase 3 trials will need to be based on comparison with a standard treatment representing current practice. The decision maker might wish to estimate the effect of the drug, independent of its effect relative to a placebo, to help decide whether to invest further in the treatment's comparison with an active substance instead of placebo. The case of Bayes analysis and interpretation of a two-armed, active comparator clinical trial is the subject of the next section.

Assume that the outcome of interest is the survival probability at 3 years. An analyst would likely denote the 3-year survival probability mathematically; for example, representing it by a parameter, such as λ. Decisions are rarely, if ever, made with absolute certainty in the knowledge about the true value of λ. Instead, the expected value of λ must be inferred from the accumulated education and experience of the investigators and from the best quality evidence provided by medical science. For example, suppose a study had been done with 100 patients receiving the experimental treatment, and the results showed that 50 patients were still alive at 3 years. In the absence of other information, a reasonable estimate of λ for the experimental treatment is 50% ($=50/100$). Note, however, that there is still uncertainty about the true 3-year survival probability, λ.

Figure 1 shows probabilities for λ for two studies of the experimental treatment, differing only in the number of patients enrolled in the studies (25 or 200), where patients were followed for 3 years. Suppose also that both studies had the same fraction of patients survive to 3 years, equal to 50%. The curves show the Bayes posterior probabilities for each possible value of λ based on the data. Because the studies sampled from the population of all possible patients, there still is some uncertainty in the true survival rate, λ.[2] An important value of the trial with 200 subjects is that the 95% central range of possible values of λ is essentially limited to between 0.38 and 0.64. By contrast, after the trial with just 25 subjects, there still is some possibility that the true value of λ lies outside this range. These curves then express not only the likely value of λ, but also one's uncertainty in this estimate, given the data. This example illustrates how increasing the number of patients enrolled in a trial reduces the uncertainty associated with an estimate in the endpoint. The following section shows how Bayesian reasoning from early drug development trials might be used to estimate uncertainty about an experimental treatment's effect.

BAYES APPLIED TO CLINICAL TRIAL DATA

Suppose there is abundant basic science and epidemiologic evidence to support a new clinical hypothesis about how to prevent deaths. Moreover, a drug has undergone testing in clinical trials of humans. A pertinent question is whether the trial data are sufficiently compelling to recommend widespread public use of the drug.

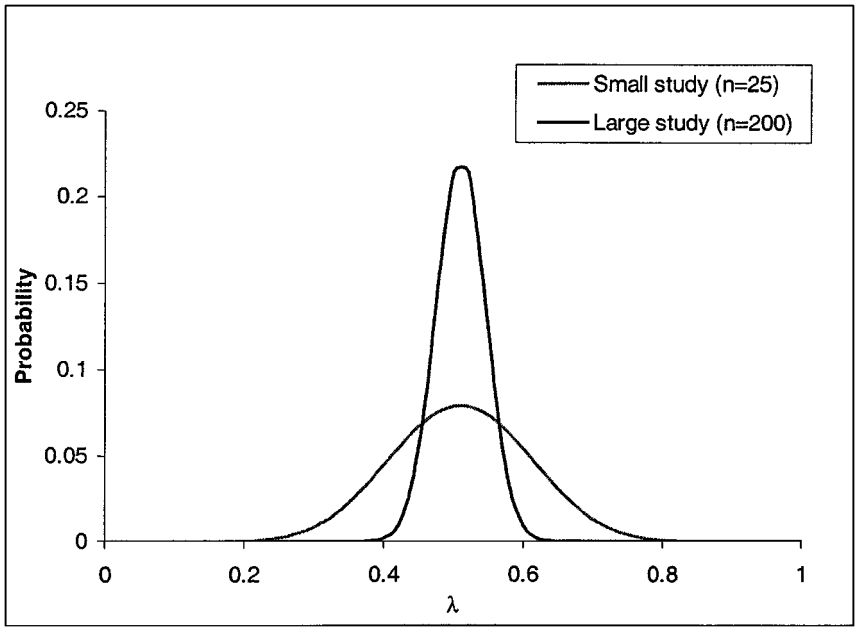

Figure 1. Posterior probabilities for the primary endpoint (3-year survival, λ) for a study of 25 and 200 patients, respectively. In the following section, these curves represent the prior probability functions for interpreting the next study between standard and experimental treatment.

There are two relevant considerations, or risks, when making this decision. The first risk is mistakenly recommending widespread use of the drug, when the drug is not truly effective. The second risk is mistakenly not recommending widespread use of the drug, even though the drug is truly effective. For this example, assume that safety and tolerability of the drug is not an issue; the only concern for the decision maker is clinical effectiveness. Several approaches have been developed to compute the average consequence to a patient of using a drug; however, it is rare to formally estimate the risks of making a wrong policy recommendation. What is presented here is the Bayesian method for computing these risks. (The third option of recommending additional testing in clinical trials before making a final recommendation is discussed elsewhere [8;9;10]).

Table 1 shows the steps necessary to estimate the magnitude of the risk (in this example, expected number of lives lost) of making the wrong recommendation. Step 1 involves determining the loss, called Bayes loss, if the wrong recommendation is made. This could be represented in clinical terms (e.g., lives lost, years of life lost, or quality-adjusted years of life lost), in economic terms (direct or total costs saved), or in a combination of both (cost-effectiveness or net health benefits [13]). Many trials directly measure the outcome of interest for deciding a policy. However, in some instances, the trial may have to rely on a surrogate measure. For example, for trials of drugs to treat hepatitis C, the main outcome measure is viral load, because it would be exceedingly difficult to keep patients in trials lasting many years to show an effect on liver cirrhosis, and thus mortality. In either case, we assume that, if the primary outcome measure of the trial is a surrogate, it can be transformed mathematically into a relevant measure of risk relevant to the decision maker.

For this example, assume that the primary outcome measure of the trial was whether the patient was alive 3 years after randomization, where $\theta = 1$ if alive and 0 if dead. As in the clinical example, let $i = 1$ if the patient received the experimental treatment and 0 if

Table 1. Summary of a Bayesian Analysis

Step	Bayesian term
1. Formulate the problem. Determine what outcomes are important to making a policy recommendation. Define the magnitude of the loss associated with a unit change in each outcome (e.g., excess deaths, lost life-years).	Bayes loss
2. Specify one's beliefs about the probability distribution for these outcomes, i.e., the prior probability function(s).	Prior probability
3. Estimate the probability of the observed data given the prior distribution on the outcomes.	Likelihood function
4. Apply Bayes rule to estimate the probability of the outcome, given the trial data.	Posterior probability
5. Integrate the loss function with the posterior probability to estimate the risk of making the wrong recommendation.	Bayes risk

the patient received standard treatment. The risk that the policy maker is assessing is how many lives are likely to be lost with each of the two recommendations.

In step 2, the policy maker specifies the prior distribution function before the trial (the results of a Bayes analyses are likely to be viewed more credibly if the prior is defined before trial data are analyzed). For both standard and experimental treatment, the functions, denoted as $p_i(\theta_i)$, give the probability for each possible value of θ_i. If there exists an abundance of prior information (e.g., biological plausibility, epidemiologic studies, prior randomized studies), then the prior distribution function is likely to appear narrower than if there exists little prior information. Assume that the prior distributions are as represented in Figure 1. Usually, experience with the standard treatment is more extensive than for the experimental treatment; as a consequence, there is greater uncertainty about the effects of the experimental treatment compared to standard treatment. It is also assumed that, on an ethical basis (2), the trial would not be conducted unless there was substantial uncertainty about the likelihood of one treatment being better than the other, so the mean of both curves are set equal to each other (50%).

At Step 3, the statistician estimates the probabilities of observing the outcomes seen in the clinical trial, for any given level of parameter, θ_i. These represent the likelihood functions, $f_i(x_i|\theta_i)$, discussed in the first section.

Bayes rule is used to estimate the posterior distribution of the θ_i's (step 4) based on the prior distribution and the likelihood function. For this illustration, assume that the clinical trial enrolled 200 patients in each arm of the study. Also, at the end of 3 years, the survival of patients randomized to the experimental treatment was 65%, but survival of patients randomized to standard treatment was only 45%. Figure 2 shows the posterior distribution of the θ_i's based on the prior and these data. This curve shows that given this size of trial, there is little overlap in the posterior distributions of the two treatments, so the decision maker can feel reasonably confident in making a recommendation in favor of the experimental treatment.

However, the Bayes method permits the policy maker to be more explicit about the risk of making the wrong recommendation. For either the prior or posterior distribution, the expected number of lives lost because of making a wrong decision due to uncertainty in the true survival rates can be computed, and is called the Bayes risk (step 5). The risk of deciding to recommend using the experimental treatment, when not using it is less harmful, occurs when $\theta_1 < \theta_0$, or R_0. Conversely, the risk of deciding not to recommend using the experimental treatment, when using it is less harmful, occurs when $\theta_1 > \theta_0$, or R_1. Because we decided to express the loss in terms of number of deaths, the risk is proportional to

Table 2. Because of the Uncertainty About Which Treatment is Superior, the Excess Number of Deaths Associated with Each Recommendation, Before Study and After Study, by Number of Subjects per Study Arm[a]

		Treatment recommended	
		Standard	Experimental
Before study		4.16%	4.16%
After study			
Subjects per arm			
	25	5.60%	1.43%
	50	6.88%	0.54%
	75	8.09%	0.25%
	100	9.09%	0.09%
	125	10.00%	0.03%
	150	10.73%	0.02%
	175	11.38%	0.01%
	200	11.94%	<0.01%

[a] Assume survival rates, regardless of sample size, of 45% and 65% for standard and experimental, respectively.

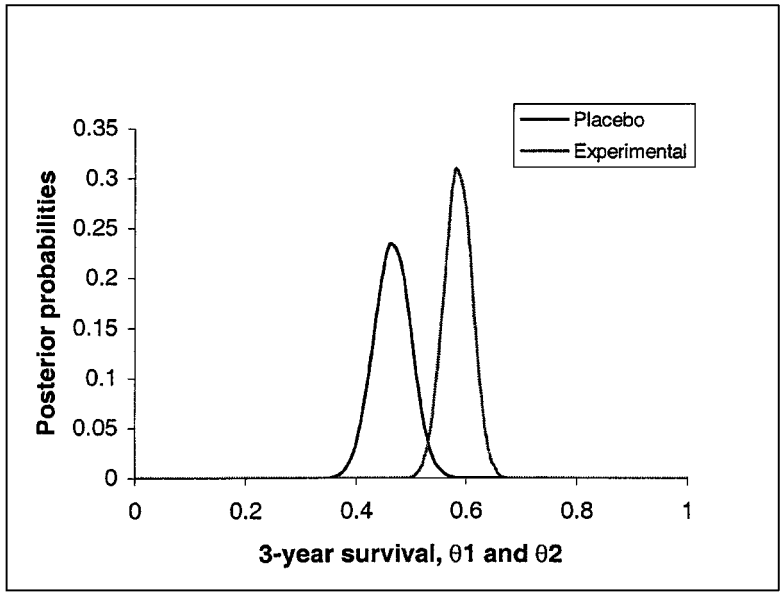

Figure 2. Posterior probabilities of the endpoint (3-year survival) for the standard and experimental treatments.

the differences between survival rates (e.g., the absolute value of $\theta_1 - \theta_0$), weighted by the likely values of these survival rates given by posterior distributions of the θ_i's. By contrast, if the decision had been made to express the loss in terms of years of life lost, the risk would be proportional to the difference in reciprocal hazard rates, weighted by posterior distributions of these hazard rates.

Table 2 shows the risks of a policy decision both before and after the trial. Before the study, regardless of the recommended treatment, the expected percentage of patients who would die needlessly because of uncertainty about the true survival rates exceeds 4%, and is the same regardless of which treatment is chosen. After the study, the rate of excess deaths depends on the number of patients in the study and the policy recommendation. Not

surprisingly, given the higher survival rate among patients taking active treatment, there is a lower risk for recommending the experimental drug. Also, even though the trial data helped to reduce the risk of making a wrong decision, it did not reduce the risk to zero. The trial analyses provide the policy maker with a method to assess whether there is enough data to warrant making any definitive recommendation at this time, or whether more information from other studies, such as another clinical trial or a registry, would be advisable.

To facilitate Bayesian reasoning and calculation, a number of computer programs, or Bayesian calculators, have been developed and are available on the World Wide Web, including:

- members.aol.com/johnp71/bayes.html;
- members.aol.com/johnp71/javastat.html;
- www-sci.lib.uci.edu/HSG?RefCalculators2A.html; and
- omie.med.jhmi.edu/bayes.

A calculator that estimates the Bayes risk and Bayes-optimal sample size for two-arm binary endpoint clinical trials also is available from the author upon request (6).

CONCLUDING REMARKS

My interest in Bayes began in clinical training. It was presented as a logical, qualitative approach to organizing the evidence and resolving diagnostic and treatment dilemmas for individual patients in clinical practice, with the genetic example above being a simple example. The extension of Bayes for use in clinical investigations was therefore a straightforward and logical step from the early underpinnings of applying Bayes in the clinic.

The potential value of Bayes became even more relevant when observing how results reported from standard frequentist statistical methods were used for making clinical and policy decisions. In some instances, the data clearly achieved the critical threshold of a p value of .05 in favor of the experimental treatment, but the decision was made not to adopt the treatment. Conversely, some treatments have been adopted when the data in no study achieved this critical threshold. Well-considered and timely discussions of these and other underappreciated anomalies of the frequentist approach are discussed elsewhere (4;5;7).

Explicit Bayesian analysis has not, nor is likely, to dominate other ways we use to make decisions. For example, the time constraints of a busy clinic require the use of heuristics ("rules of thumb") to make clinical decisions quickly. Experimental evidence in the cognitive psychology literature has described a number of situations where people have updated their opinions in light of new data (14). Moreover, some heuristics in medicine may even prove dangerous, such as the one that leads people to interpret test sensitivity as equal to the posterior probability of disease (example from an anonymous referee).

In selected instances, explicit computation of posterior probabilities with Bayes has proven a useful tool for rationally sorting out the complexities of a given problem. This is especially true when interpreting results of a pivotal new study, and then integrating these findings into new clinical policies. The papers that follow in this section contribute toward addressing the philosophical and methodologic issues of using Bayes and demonstrating its applicability in selected case studies.

NOTES

[1] This example based on that presented by Gelman and colleagues (2).

[2] Other sources of uncertainty exist beyond that of just sampling issues; for example, there may be correlation of the observed and unobserved characteristics of the next patient that might have been enrolled in the study, compared with the previously enrolled patients. While frequentist statistics deal

with sampling error, Bayesian reasoning allows for modeling of this other type of so-called secondary uncertainty, although this is not dealt with in this example.

REFERENCES

1. Berger JO. *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag; 1980.
2. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med.* 1987;317:141-145.
3. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*, London: Chapman & Hill; 1996:1.
4. Goodman SN. Toward evidence-based medical statistics, 1: The P value fallacy. *Ann Intern Med.* 1999;130:995-1004.
5. Goodman SN. Toward evidence-based medical statistics, 2: The Bayes factor. *Ann Intern Med.* 1999;130:1005-1013.
6. Halpern J, Brown WB, Hornberger J. The sample size of a clinical trial: A Bayesian-decision theoretical appraoch. *Stat Med.* In press.
7. Harrell FE Jr, Shih Y-C. Using full probability models to compute probabilities of actual interest to decision makers. *Int J Technol Assess Health Care.* 2001;17:17–26.
8. Hornberger J, Brown BW, Halpern J. Designing cost-effective clinical trials. *Stat Med.* 1995; 14:2249-2259.
9. Hornberger J, Wrone E. When to base clinical policies on observational versus randomized trial data. *Ann Intern Med.* 1997;127(part 2):697-703.
10. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med.* 1980;302:1109-1117.
11. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials. *J R Stat Soc Series A.* 1994;157:357-387.
12. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. *BMJ*. 1999;319:508-512.
13. Stinnett AA, Mullahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making.* 1998;18(suppl 2):S68-S80.
14. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974; 185:1124-1131.
15. Webster AR, Maher ER, Bird AC, Moore AT. Risk of multisystem disease in ocular angioma (haemangioblastoma). *J Med Genet.* 2000; 37:62-63.