Clinical Trials
Is the Bayesian Approach Ready for Prime Time? Yes!

Donald A. Berry, PhD

The defining characteristic of any statistical approach is how it deals with uncertainty. In the Bayesian approach, all uncertainty is measured by probability. Anything unknown has a probability, including future results in a clinical trial (based on current results). Frequentists also use probabilities, but in a restricted sense.

Bayesian conclusions depend on results actually observed. Because their use of probability is limited, frequentists go through contortions to draw conclusions. In particular, conclusions depend on more than just observed results. For example, frequentist p-values include probabilities of results more extreme than observed, in which probability calculations depend on the trial’s design. Both aspects are scientifically questionable. “More extreme results” were not observed and should not matter at all. Small p-values are taken to be evidence against the null hypothesis, so one may reject a hypothesis because it assigns little probability to unobserved results, and—for the same data—accept one because it assigns greater probability to unobserved results. The other questionable aspect is the strong dependence of conclusions on design. Because the same data but different intentions of the investigator had something happened that did not happen, the p-value may be any number from 0 to 1. And calculating a p-value is not possible unless the design is followed exactly. One implication is that keeping an experiment running to collect additional data is not an option. That seems unscientific.

In contrast, the Bayesian paradigm is tailored to the learning process. As information becomes available, one updates what one knows. This gives the Bayesian approach its flexibility and makes it ideal for clinical research. Bayesian probabilities can be calculated at any time and on the basis of whatever information is available. One consequence is the ability to calculate probabilities of future results. At any time, one can assess where a trial is going, the probability that it will be a success, etc.

An example is the Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN) trial.2,3 The dose assigned to the next patient was chosen to maximize information about the dose–response relationship. Doses proving to be uninteresting are little used. Informative doses garner more observations. And predictive probabilities allow for addressing whether one wants to make the next observation at all.

The Bayesian approach is also tailored to making decisions. Designing a clinical trial is a decision problem. The ASTIN trial was a major research advance because it used Bayesian updating, but even it did not fully exploit the Bayesian potential. For example, it carried artificial constraints. One was the timing of decisions to stop accrual. The algorithm would have recommended stopping well before it was given its first opportunity to do so. Another aspect of the design was dictated apart from Bayesian considerations: accrual rate. This is an important part of a trial’s design. A Bayesian decision analysis can address whether tempering accrual is wise (on the basis of a drug’s patent life, potential profits, costs associated with the trial, etc). It can also address changing accrual rate in an adaptive fashion, depending on the accumulating data. Had an accrual governor been available in the ASTIN trial, it would have been applied when the data were suggesting that no dose of the agent was effective.

A design’s false-positive rate is a traditional frequentist notion that is also important in a Bayesian approach. False-positive rate and statistical power can be calculated for any prospective Bayesian design and serve to relate the new with the old.

“Is the Bayesian approach ready for prime time?” suggests that the Bayesian approach has not been ready until now. Although I do not like this implication, there is an aspect of today’s world that makes the Bayesian approach more “ready” now: computational facility and speed. Adaptive Bayesian designs tend to be complex. For example, they look forward frequently during the trial and they steer the trial based on updated predictive probabilities. Predicting future results can be computationally intensive. Performing the necessary calculations during an actual trial is usually possible with 25-year-old computers. But calculating false-positive rate and power requires simulating the trial many thousands of times. The computer programs have loops within loops within loops. Computations are onerous and may require a supercomputer or a high-level cluster that were not available 25 years ago.

Received January 21, 2005; accepted January 21, 2005.

From the Department of Biostatistics & Applied Mathematics, University of Texas M. D. Anderson Cancer Center, Houston, Tex.

Correspondence to Donald A. Berry, Department of Biostatistics & Applied Mathematics, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Box 447, Houston, TX 77030-4009. E-mail dberry@mdanderson.org

(Stroke. 2005;36:1621-1622.)
© 2005 American Heart Association, Inc.

Stroke is available at http://www.strokeaha.org
DOI: 10.1161/01.STR.0000170637.02692.14
Why the question in the title? In view of the benefits described here, why is there not a long Bayesian tradition in clinical research? The answers lie in the 20th century history of statistics in which the Bayesian approach was usually dismissed as being subjective. Learning based on a trial’s results implies the existence of knowledge before the trial. Using Bayesian terminology, having a posterior probability implies the need for a prior probability. So the Bayesian approach requires specifying probabilities before the trial. This is a great virtue because it enables using historical and concurrent information. It also means there is an inevitably subjective aspect of prior probabilities and, therefore, of all probabilities.

Subjectivity in science is regarded by some as being bad. I think it is essential. For example, the choice of statistical model is subjective—there is no universal model. And the model chosen can easily have a greater impact on conclusions than does one’s prior probabilities. Moreover, if one wishes, one can use what is called noninformative or flat prior distributions in which the observed results dominate the final probability distribution. This emulates the frequentist attitude of ignoring historical information, but it preserves the Bayesian ability to learn and adapt.

The Bayesian approach provides tools for designing trials that treat participants more effectively and that identify better drugs and appropriate doses more efficiently and faster.

References

Key Words: Bayes theorem | clinical trials | probability

Is Bayesian Analysis Ready for Use in Phase III Randomized Clinical Trials? Beware the Sound of the Sirens

George Howard, DrPH; Christopher S. Coffey, PhD; Gary R. Cutter, PhD

Bayesian analysis is attractive (primarily) because it uses available data from other studies to potentially reduce study sample size. Bayesian approaches begin by mathematically assuming a “prior distribution” for the treatment effect, then adjusts this distribution using the results of the Phase III study to produce a “posterior distribution.” With even modest information available, this approach will result in substantial improvements in efficacy and reduced sample size. On the surface, ignoring information seems at first glance by mathematically assuming a “prior distribution” for the treatment effect, then adjusts this distribution using the results of the Phase III study to produce a “posterior distribution.” With even modest information available, this approach will result in substantial improvements in efficacy and reduced sample size. On the surface, ignoring information seems at first glance wrong, even unethical.

There are numerous technical aspects of trial design (such as “adaptive trial” methods) for study monitoring available using the Bayesian approach. However, alternatives to these approaches are available in the frequentist approach. We consider these issues to be tangential to the primary “reduction in sample size” argument for using the Bayesian approach.

Although Bayesian approaches have become more acceptable for statistical analyses in general over the past few decades, the majority of Phase III trials are conducted under the frequentist approach. We suggest that this is because of rather fundamental issues associated with Bayesian Phase III trials. So why wouldn’t we not use available information to reduce sample size?

We Have a Difficult Time Agreeing What We Know

The choice of studies guiding the “prior” and the relative value to be assigned to each study is largely subjective. The weight chosen is often a function of statistical precision, which provides more weight to a large poorly conducted trial than a smaller well-conducted study. It is not clear whether to include studies conducted in “almost” the same population, nor is it clear whether to include information from epidemiological studies.

Given the same Phase III trial, the use of different “priors” will result in different “posteriors.” Unless there is agreement on what is “known” before the study, this introduces the possibility that different interpretations could be drawn because of disagreements of what was “known.” This could be especially problematic when the study sponsor has a financial stake in the outcome of the study since critics could argue that the previous information was stacked to favor the desired finding. Although there are
many disagreements on the interpretation of frequentist trials, it is not inherent from the design.

**We Have To Be Careful To Know Exactly the Correct Amount**

With little or no previous information available, there is no meaningful reduction in sample size using Bayesian methods. The Bayesian approach becomes increasingly attractive with increasing previous information.

However, substantial a priori knowledge introduces potential ethical concerns in the conduct of the trial. Although investigators may have their individual priors, equipoise is reflected in a societal belief of equal likelihood that either treatment is superior. It is not clear that equipoise exists if the study “officially” assumes a more likely winner. Such an assumption may imply there is not true informed consent. However, substantial recruitment barriers are likely if true informed consent is provided (ie, consider the impact of a statement that “we are only 70% sure the new treatment is superior and therefore would like to randomize you to potentially receive either treatment”). This is not an issue in frequentist trials in which the trial position of equipoise is present.

**We May Know Something That Is (Systematically) Incorrect**

There are 2 reasons why previous studies may provide incorrect information—sampling variation and bias. Bayesian approaches appropriately adjust for sampling variation, but biases are “carried forward” to affect the posterior distributions. There is no study absolutely unbiased; hence, we are not discussing the presence but rather the magnitude of “carry-forward” bias.

We suggest that even under the theoretical assumption that previous studies are conducted bias-free, the use of Bayesian approaches may introduce bias through a process akin to the file drawer problem in meta-analysis. Because of their cost and complexity, Phase III trials are not proposed without strong supportive evidence from earlier studies. Suppose a number of studies using potential drugs are evaluated. Among those studies, some will suggest efficacy and will be “encouraged” for Phase III trials, whereas others not showing efficacy are “discouraged.” Because of this differential selection process, the estimated effect of studies leading to Phase III trials is positively biased. These biased estimates serve as the foundation for the “prior” for the Phase III trial and subsequently result in biased estimates of treatment effect. This bias is introduced by the transition from Phase II to Phase III studies using Bayesian approaches but is avoided by the independent replication of the frequentist approach.

**Conclusions**

It is a difficult decision to ignore available information when making a decision, and as such the underlying goal of Bayesian analysis is laudable (seductive?). However, should trials not be an independent confirmation of our previous work? We suggest that answering the siren’s song of Bayesian analysis for Phase III trials introduces additional issues in study interpretation, ethical issues, and bias (some researchers seem to have better hearing than others).

**References**


**Key Words:** Bayes theorem | clinical trials | probability

---

**The Bayesian Principle Can We Adapt?**

Geoffrey A. Donnan, MD, FRACP; Stephen M. Davis, MD, FRACP; John Ludbrook, MD, DSc, ASTAT

Most clinicians are more familiar with the traditional trial design, for trials of therapy in acute stroke, consisting of fixed sample sizes based on power calculations using preexisting data (the frequentist approach referred to by our protagonists).

The Bayesian approach is unique in that unlike the traditional design, an initial dependence on preexisting data are constantly adjusted using new data as they are accrued from the current trial. Its main advantage is that this allows potential reduction of sample size by continually updating the probability of success or futility. As argued by Berry, this flexibility of Bayesian design makes it ideal for clinical research.

Why the need for new trial designs such as this? There is no doubt that the large number of negative stroke trials using neuroprotection is the main driver. A reason for this may be trial design and suggests the potential for more sensitive and efficient testing of hypotheses using such approaches as adaptive designs (as used in the Acute Stroke Therapy by Inhibition of Neutrophils [ASTIN] trial) and responder analyses (in which success is calibrated against the initial neurological severity) used in the Abciximab in Emergency Stroke Treatment Trial (AbESTT) trial. Other drivers for novel trial design include the escalating cost of trial conduct, an increasingly wary pharmaceutical industry, and the need to...
consider the ethics of use of experimental agents in large trials. The Bayesian approach allows the use of online determination of futility, an approach that is ethically sound.

In essence, Howard argues that Bayesian analysis is seductive and appealing but introduces unacceptable biases because of the difficulty in accruing high-quality, unbiased previous data. He argues that the frequentist approach avoids these problems by providing an independent confirmation of previous work. Conversely, Berry is of the view that this problem is outweighed by the flexibility and almost “online” prediction of the likelihood of success. We would also submit that although the use of previous information occurs early in the trial process when using the Bayesian approach, as the trial progresses, this is overwhelmed by the new data accrued. For example, in ASTIN, the natural history data from the Copenhagen study were used as the previous distribution as a baseline for the treatment effect. The probability of efficacy becomes progressively more dependent on new data being continually accrued.

A third approach is to adopt frequentist design with a minimal sample size planned at the outset of the trial but then to perform prespecified interim analyses. Minimal sample sizes are often mandated by institutional ethics committees. Interim analyses necessitate adjustment of $P$ values, with re-estimation of minimal group size, because of the potentially increased risk of type 1 error.

Regardless of the pros and cons, we regard this as an extremely healthy debate because it is a manifestation of the drive to develop new approaches to trial design and analysis. Indeed, we may need an adaptive design based on success or failure of future stroke trials to make continual assessment of the benefits of these alternative approaches.

**Definitions**

Bayesian analysis is the use of previous distributions (past data) to estimate posterior distributions (what is happening now).

Frequentist design is the conventional approach of estimating minimal sample size in advance of the trial, completing the trial, and then analyzing the outcome in terms of $P$ values.

Adaptive design/sampling is the adjustment of estimate of minimal sample (group) size as the trial progresses.

**References**


**Keywords:** Bayes theorem | clinical trials | probability
The Bayesian Principle: Can We Adapt?
Geoffrey A. Donnan, Stephen M. Davis and John Ludbrook

Stroke. 2005;36:1623-1624; originally published online June 9, 2005;
doi: 10.1161/01.STR.0000170640.09580.fb

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/36/7/1623

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/