STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:469–482
Published online 15 June 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2928

Bayesians in clinical trials: Asleep at the switch

Lemuel A. Moyé*, †, ‡

School of Public Health, University of Texas, 1200 Herman Pressler – E815 Houston, TX 77025, U.S.A.

SUMMARY

The refreshing Bayes perspective has much to offer biostatistics. Yet, from its 225-year-old roots sprung difficulties that blocked its growth at the beginning of the 20th century. Computational obstacles in concert with an inability to identify the best indifferent prior revealed a weakness on which frequentists capitalized. It took Bayesians 40 years to recover, allowing the infant field of biostatistics to fall firmly in the hands of the frequentists. Recent disillusionment with the frequentist perspective, and its hegemony of p-values, has produced a second opportunity for the Bayesian philosophy to make solid contributions to clinical trials.

However, difficulty with the applicability of the likelihood principle, problems with prevalent prior 'disinformation' in clinical medicine, in concert with the complexity of truly representative loss functions threaten again to thwart the Bayesian march into biostatistics. Seven suggestions are offered to the Bayesians to help them adapt to the rigors of clinical research. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: Bayes; methodology; prior distributions; loss functions

INTRODUCTION

Even its critics must admit that the Bayesian point of view with its new and refreshing perspective for biostatistics can, like the first spring breeze blowing through a musty room, be energizing. Allowing the scientist to focus on the hypothesis that he or she actually believes (rather than being forced to disprove what they do not believe) can be uplifting, and the facility to formally incorporate relevant preliminary information is invigorating.

When first exposed to the Bayesian perspective, physician-scientists commonly react like weary travelers who, after a long and perilous journey through alien lands, have finally come home. The

*Correspondence to: Lemuel A. Moyé, School of Public Health, University of Texas, 1200 Herman Pressler – E815 Houston, TX 77025, U.S.A.
†E-mail: Lemuel.A.Moye@uth.tmc.edu
‡Professor.

Contract/grant sponsor: National Center for Research Resources; contract/grant number: US4-RR023417-01

Received 3 November 2006
Accepted 12 April 2007

notion that prior impressions and information, almost always in full supply, can now be formally incorporated in an analysis is not just intuitive—its order-shattering thought process is liberating. In fact, physician-researchers wonder aloud how the field of statistics ever allowed itself to stagger away from this intuitive Bayesian point of view, available for hundreds of years.

Bayesians are now breaking down traditional barriers in clinical trials. The possibility of including them was tentatively raised 30 years ago [1, 2]; now they boldly stake a claim on the clinical trial landscape. Flexing their computational muscles, Bayesians demonstrate undeniably impressive abilities in sample size computations [3–10], while others assess the causal nature of the exposure–disease relationship when clinical preference drives the exposure [11]. The ethics of the use of Bayes procedures have been discussed [12], and monitoring procedures from a Bayes perspective have been formulated and adopted [13–16]. Bayes procedures are also being applied to observational studies [17, 18], and a particular series of useful analysis on missing data is now available [19]. The role of adaptive randomization in clinical trials, i.e. permitting the trends in treatment effect within the trial itself to influence the proportion of patients who are allocated to a particular therapy, is steadily expanding into Phase II clinical trial methodology [20, 21], aiding drug discovery [22]. And, while several textbooks on Bayes procedures have been available [23–25], there is now a textbook devoted solely to the application of Bayes procedures in clinical trials [26].

In addition, the Bayesian approach has spilled into the pharmaceutical review and approval process, attracting new attention from the U.S. Food and Drug Administration (FDA). The Center for Devices and Radiological Health (CDRH) accepts Bayesian procedures as a standard statistical approach. Approximately 10% of FDA approvals for medical and radiological devices (e.g. spinal implants and cardiovascular stents) are based on Bayesian analysis. If required to go through traditionally designed Phase III trials, these medical implements may have taken longer to become available.

In fact, the use of Bayesian procedures is rising so rapidly that the uninitiated may be forgiven for reacting like the newcomer to a beach during high tide. The onlooker concludes that, since the water level is high now, it must always be so. Yet, the experienced observer appreciates tidal ebb and flow, and the established clinical trial methodologist, disciplined by a long, bitter experience with methodologic and clinical 'breakthroughs' that promised much but delivered little of practical consequence, understands that the most useful perspective when assessing a new tool's utility is time and practice, followed by calm assessment.

While relatively new to biostatistics, the Bayesian perspective has existed for over 225 years, almost three times longer than the Fisherian hypothesis-testing perspective, and four times longer than the clinical trial mindset. It therefore would be useful to stop for a moment to see where the Bayesian perspective has been, before we in biostatistics take the path it has chosen for our future.

BAYES' QUIET BOMBSHELL

The Reverend Thomas Bayes, having published only three manuscripts during his entire life, died in 1761. Upon his death, his family asked the Reverend Richard Price to review the decedent's papers. Although the two reverends befriended each other in years past, they could not have been more different. While Bayes was quiet, complacent, and relatively non-productive mathematically, the contentious Reverend Price fought continuously with both his secular and religious superiors,

DOI: 10.1002/sim

going so far as to argue against the deity of His Majesty the King. Price gravitated to the topic of his friend's incomplete paper that he discovered, revised it [27], and presented the manuscript before the Royal Society on 23 September 1763, giving Rev. Bayes full credit.

The paper addressed the vexing problem of inversion. Given two events A and B, probabilists of the time struggled with how to compute $\mathbf{P}[A \mid B]$ from $\mathbf{P}[B \mid A]$. The Bayes–Price paper generated the notation and concept that we now know as Bayes Theorem, which states

$$\mathbf{P}[\text{Hypothesis} \mid \text{Evidence}] = K\,\mathbf{P}[\text{Evidence} \mid \text{Hypothesis}] * \mathbf{P}[\text{Hypothesis}] \tag{1}$$

where K is a proportionality constant.

The idea was revolutionary. One starts with a collection of hypotheses, each of which is assigned a probability. Evidence is then gathered (e.g. the results of a research effort), producing a new, updated set of probabilities. This approach aligned nicely with the developing ideas of the scientific method, in which one updates one's scientific hypothesis through the collection of data. However, the Bayes–Price work introduced the revolutionary revelation that an event's probability need not be fixed. Rather than have it anchored to a fixed proportion (or relative frequency) as 'frequentists' believed, it could be subjective, with its own metric which may or may not be quantifiable.

Simon Laplace threw his considerable intellect and engaging personality into the mix,[§] justifying the concept of subjective probability. For the next 120 years, well into the second decade of the 20th century, the Bayesian approach attracted the attention of some of the best minds in mathematics and probability. During the fecund 19th century, applied statisticians commonly constructed prior distributions, obtained data, and updated the prior distribution to the posterior distribution [28, 29]. Then, as now, the Bayes approach embedded itself into many areas of applied statistics.

However, forward progress slowed in the early 20th century. While one difficulty revolved around the computational complexity that rapidly arose in Bayesian mathematics, Bayes procedures staggered under the weight of their own implications. Laplace had convincingly shown that any probability distribution could function as a prior distribution, but this effort begged the question, 'Which probability distribution serves as the best prior?' Enunciation of the 'indifference principle' i.e. the idea that the prior distribution should express no bias toward one set of values, provided some guidance to these early Bayesians. However, while this principle set the goal, it did not direct how the goal could be achieved, and the hunt for this perfect non-preferential prior took on the sense of an Arthurian search for a Bayesian Holy Grail. Bayesian forward momentum stagnated at the beginning of the 20th century, stuck between the un-scalable mountain of the indifference principle on the one hand, and the un-fordable river of statistical inference on the other.

Trapped there, the Bayesian philosophy was all but routed by Ronald Fisher and statistical hypothesis testing. Fisher's approach stunned many in science [30], catching the Bayesians by surprise. Fisher, no friend of the Bayes point of view, claimed that he had broken the back of the inference problem without having to rely on the concepts of subjective probability or prior distributions [31].[¶] The relative frequency juggernaut rolled forward through the

[§] The teenaged Simon Laplace, understanding that his poor family could not finance his education, went from house to house, alternately asking, cajoling, and demanding that the different households, themselves poor, provide money for his training. The astonished neighbors acceded to his request, committing their hard-earned money to his education. Few communities in history have had their investment so amply rewarded!

[¶] It was Fisher who suggested that even Reverend Bayes discarded his inversion approach, refusing to publish it during his lifetime.

DOI: 10.1002/sim

1930s on the foundational work of Egon Pearson, Jerzy Neyman [32, 33] and others [34–38].

It was Harold Jeffreys who revived the Bayes perspective. Jeffreys' seminal textbook, *Theory of Probability* [39], opened the field of physical sciences to the modern Bayesian approach. Jeffrey's focused attention on the utility of Bayes procedures in the 'hard sciences' led to his discovery of the 'vague' or 'non-informative' prior distribution [40, 41], a cornerstone of modern 'objective' Bayesian analysis. Vehemently defending the Bayes perspective, Jeffreys, [42–48] and Savage [49, 50] stood for the application of Bayes philosophy to geology, engineering, and other hard sciences [51]. The Bayesian–frequentist struggle was its most volatile during this tempestuous time.

However, while the Bayesian fought fiercely for a central role in the hard sciences, biostatistics moved quietly forward without them. The huge growth of medical research during World War II, in combination with limited resources to fund and promulgate that work [52], produced the need to develop an objective assessment of the 'worthiness' of a grant application for funding or a manuscript for publication. In addition, there was wide spread dissatisfaction with the degree to which unsubstantiated opinion insinuated itself in the medical research writing of the period.

Therefore, despite clear limitations [53], the Fisherian p-value of 0.05 was selected as the new standard of medical research. The embrace of p-values by granting agents and academia was followed by their wholesale acceptance by the regulatory industry and, consequently, the pharmaceutical industry. Uniformity was sought, and (for better and for worse) was achieved, effectively shutting the Bayesians out of mid-20th century modern medical research. Nevertheless, Bayesians purposely moved forward again, now making solid inroads into fields where frequentist-driven hypothesis testing had not taken such strong root, e.g. economics and astronomy.

Over the next 30 years, physician-scientists heard little from the Bayesians. However, the medical research community's wholesale embrace of the frequentist approach to inference in clinical trials began to exact a toll, with the most withering attacks reserved for their strict enforcement of the Fisherian p-value threshold [54–58]. A series of spectacular failures in clinical trial inference (covered in the second half of this commentary) made many questions whether the frequentist approach to clinical trials had outlived its usefulness.

By the 1990s, medical researchers were open to a new philosophy. The Bayesians, armed with (1) a wealth of experience gained from their fiercely contested role in the physical sciences, and (2) new computational tools, were standing by, waiting for their chance.

VALIDITY OF THE KEY INGREDIENTS

As it was at the beginning of the 19th century, so Bayesians would now have us believe at the beginning of the 21st century that they are poised to capture the heart and mind of applied statistical science. This time, the focus is health-care research in general and clinical trial methodology in particular. However, they cannot decisively take this ground until they demonstrate how to *constructively* apply the three innovations their philosophy generates: (1) the likelihood principle, (2) the prior distribution, and (3) the loss function, to a field that poses unique and sometimes treacherous dilemmas. Unfortunately, the challenges clinical investigation holds for the application of these precepts may be just as daunting as the obstructions that stunted the growth of the applied Bayesian biostatistics 100 years ago.

DOI: 10.1002/sim

The likelihood principle

The likelihood principle is one of the cornerstones on which the modern Bayes edifice now rests. It stands for the elegant proposition that the only relevant part of an experiment is, in the end, what has actually occurred [23]. While this simple statement may appear self-evident, Bayesians, quite perceptively, fault frequentists in their standard violation of this foundational precept. The frequentist argues that unobserved data can be just as influential as the observations that have occurred. An illustration of the conundrum frequentists find themselves in when caught in a likelihood principle violation was first developed by Pratt as an engineering problem [59], and later adopted for the health-care research field [60] as follows:

> 'I was just interested in producing a confidence interval for diastolic blood pressures in my clinic', one physician exclaimed to another at a clinical trial meeting. 'That's all I wanted to do! I had an automated device that would read blood pressure very accurately and produce data that appeared to be normally distributed with a mean of 87 and a standard deviation of 5. I turned the data over to a statistician for a simple analysis'.
>
> 'When the statistician came to my clinic to share his findings, he had a conversation with the staff nurse who told him that the automated blood-pressure-measuring device did not read above 100 mmHg. When I confirmed this, the statistician said that he now had to do a new analysis, removing the underlying assumption of a normal distribution, since the blood pressure cuff would read any diastolic blood pressures greater than 100 mmHg as only 100 mmHg. I understood his point, but assured him that in my data, no patient had a diastolic blood pressure greater than 97 mmHg. Also, if there had been one, I would have called another clinic for a backup automated device, equally sensitive to BP readings greater than 100 mmHg'.
>
> 'The statistician was relieved, thanked me, and returned to his office. However, after he left, I noticed that my backup unit was broken, and e-mailed this circumstance to him. He e-mailed me back, saying that he would have to redo the analysis after all! I was astonished and called him immediately. Why should the analysis be redone? No blood pressure was greater than 100 mmHg, so the broken backup device would not have had to be used. My measurements were just as precise and accurate as they would have been if all instruments were working fine. The results would have been no different. Soon he would be asking me about my stethoscope!'

To the Bayesian, the frequentist's fixation on the availability of a confirmatory sphygmomanometer is nonsensical. Since the data never required it, they ask, why be concerned about its presence? At first glance, one can only be amused (or angered) by the feckless frequentist as portrayed.

However, the statistician depicted in the example is quite right in his demand to learn of the availability of the backup meter, if his concern is generalizing to the population at large.[‖] The raison d'être of clinical trials is precisely this type of generalization.

The likelihood principle also suggests the inadvisability of the two-tailed hypothesis test. Two-tailed testing ensures that there is adequate protection for the finding that a therapeutic intervention

[‖]Clearly, no backup unit is required in the sample, since every subject's blood pressure was less than 100 mmHg. However, if the statistician is no longer concerned about what occurs in the sample, but in the population at large, his concern about the backup meter is central. In order to generalize the findings to the population at large, specifically in patients who have DBP$>$100 mmHg they must have had the opportunity to be admitted to the sample and had their blood pressure measured. If their blood pressure was $>$100 mmHg, their BP would remain unknown and hence the BP distribution would have to be censored. Generalization from the sample to the population requires a violation of the likelihood principle.

DOI: 10.1002/sim

may produce harm. The investigator creates his or her critical region so that the null hypothesis is rejected for hazard as well as for benefit. However, at the study's conclusion, the test statistic can only fall on one side of the distribution—it cannot be in both. Physician-scientists are commonly confused about this, asking 'Why, when we now know what side of the distribution the test statistic is on, must we bother with the type I error allocated in the other tail?', but frequentists have successfully argued that, since they said prospectively that they would divide the type I error into two, they should then follow through on their original plan. Bayesians argue that, since the data do not identify harm, regardless of what they stated prospectively, the investigator can focus all their concern on the benefit side. This is a central, clarifying difference. The likelihood principle permits the Bayesian to divorce him- or herself from the prospectively declared methodology. The frequentist would not.

In all fairness, one cannot fault this conclusion of Bayesian deductions without faulting the logic of the frequentists as well. The theoretical optimality of the one-tailed test to the two-tailed test has been established for many years. However, frequentists have learned to disregard such artificial optimality in clinical trials where two-tailed testing appropriately dominates.

Will Bayesians use the likelihood principle to dismantle a foundation tenet of clinical trial methodology—the importance of a prospective plan separate and apart from the data? Frequentists insist that if a sample's results are to be extended to the population, there must be in place a population-derived methodology (i.e. a methodology that will apply to all members of the population) for the evaluation of the sample results. However, the likelihood principle asserts that the only methodology that should be used is a sample-based one. This is a tremendous paradigm shift.

Prior information

In medicine, 'prior belief' is an amorphous mixture of fact-based knowledge with subjective impressions. Acting on prior belief and its important subjective component is unavoidable in clinical medicine, simply because we as physicians are required to treat diseases that we only incompletely understand. Both the second-year medical student who must obtain a history from a deaf–mute and the third year student vomited on by his patient during a physical examination quickly learn the need to make clinical appraisals in the face of partial knowledge. Years later, working as an expert surgeon in the operating suite and as an accomplished emergency room physician, each would have mastered the ability to make decisions in the presence of imperfect and sometimes incorrect information.

The challenge facing Bayesians in this arena are nothing short of monumental. Although we desire a degree of belief of '1' in all true propositions, and '0' in all false propositions, this lofty standard rises above what is achievable in clinical science [61]. Bayesians must therefore fall back to the best, most accurate prior information available, but here lies the trap—in medicine, the best, most accurate prior information is (1) frequently difficult to quantify and (2) commonly inaccurate, if not outright wrong.

Clinical trial investigators are among the world's greatest and saddest authorities in being led astray by false leads and unreliable prior information. The Cardiac Arrhythmia Suppression Trial (CAST), a one-sided (benefit-only) clinical trial designed to detect a beneficial effect of ecanide, flecanide, and moritzacine, was prematurely ended in the face of overwhelming evidence that two of these drugs (and eventually the third) generated excess mortality [62]. This study revealed the difficulties that occur when research design is based on strong but incorrect physician belief

DOI: 10.1002/sim

[63].** Yet, the clinical trial permitted the true nature of the effect size to be revealed precisely because the methodology utilized to measure the effect size was segregated from the prior belief. This is not the case in Bayesian analysis, and a Bayesian approach to CAST in which the prior belief influences the effect size estimator would have deepened the interpretational conundrum. Other examples, e.g. the United Kingdom Prospective Diabetes Trial (UKPDS) [64–68], abound.

One useful device that we researchers use in our quest for truth is the measuring rod of consistency. The observation that a set of different experiments, carried out by different investigators, in different patients, using different study designs produce similar results bolsters our belief in the truth of the findings. In this circumstance, our prior belief is smoothly updated and refined through the cascade of consistency. This circumstance is tailored for the Bayesian statistician, who requires reliable prior information.

Yet, even consistency can be a difficult track record to sustain in clinical research. The suggestion that the anticholinesterase inhibitor vesnarione was first helpful [69] and then harmful [70] in patients with congestive heart failure (CHF) flummoxed the cardiology community, wreaking havoc with its 'priors'. The collection of Evaluation of Losartan in the Elderly Study (ELITE) [71–73] trials and Prospective Randomized Amlodipine Survival Evaluation (PRAISE) [74, 75], each conducted to identify a beneficial effect of a medication demonstrate the wild ride that prior information can take. In each study, the investigators were surprised by a finding of benefit, leading them to conduct a second study to confirm the promising findings. Both second studies reversed the findings of the first study, bringing the investigators back to their original prior. One can be forgiven a sense of vertigo as one ponders the reaction of Bayesian trialists who must update their prior repeatedly with information that suggests not just different magnitudes of effects but also different directional effects.

It is difficult to create a consensus on prior information when different experiments purporting to study the same problem produce conclusions that are mirror images of each other. While frequentists are also troubled by this conflicting information, they can be comforted by the fact that the estimate of effect size is not affected by the prior information. This is inefficient when the prior information is accurate, but a useful buffer in the face of inaccurate prior 'knowledge'.

There may be little solace in the use of meta-analyses in developing prior belief. The literature contains useful discussions of the strengths and weaknesses of meta-analyses [76]. The controversies they can produce are commonly focused on the studies included in the analyses, e.g. when an analysis of corticosteroid use for brain injury [77] was believed to be flawed based on the exclusion of an influential study that demonstrated a benefit [78]. Thus, while a small oasis in this desert of prior information may appear as meta-analysis, the Bayesian must be especially cautious that it is not a mirage.

Both Bayesians and frequentists alike lament the unreliability of prior information in many clinical trials. However, only Bayesians are required to formally incorporate it into their measure of efficacy. Even though frequentists must also build an effect size into the study (for sample size computations), it does not directly affect the research data's estimate of that effect. Yet, prior information for Bayesians directly affects the estimate. There are many tools at their disposal permitting them to distance themselves from faulty prior information—for example, the use of vague, uninformative priors. However, the degree to which Bayesians distance themselves from

**Much of the discussion is taken from Thomas Moore's book, entitled *Deadly Medicine*.

DOI: 10.1002/sim

prior information is the degree to which they enter the land of the frequentist, where the strict rule is to completely separate *a priori* belief from the evidence-based product of the research effort.

Loss function

The third contribution of Bayesian procedures to clinical trials is the loss function. This function guides the Bayesian trialist in interpreting the posterior distribution, essentially telling Bayesians how they should use the posterior distribution to come to a decision.

Bayesians have great flexibility in choosing the loss function for the circumstance. However, in clinical medicine, real life loss functions are rarely accurately represented by dichotomous loss, absolute error loss, or squared error loss; these useful didactic tools have little relevance in clinical trials. For example, a patient's blood sugar assessment must be accurate, but the loss function varies depending on whether the patient is hypoglycemic (and overestimations may be lethal) or hyperglycemic (where underestimates are intolerable). The overall loss function will be complicated. Yet, it is incumbent upon the Bayesian who insists that the clinical research world is ready for Bayesian methods to construct realist loss functions. We require that they depict the true risk–benefit calculus the physician and their patients must work through as they ponder whether a therapy's salubrious effects are worth its adverse ones.

Asleep at the switch

Clearly, Bayesians have much good work before them as their reputation as clinical trialists moves through its nascent period. Yet, few doubt that the technical obstacles in computing and assembling helpful loss functions will block Bayesian progress for much longer. Bayesians will continue to overcome computing obstacles as they have in the past, and their insight into the intricacies of the complicated cost–benefit loss functions will only deepen over time.

Nevertheless, Bayesian biostatisticians labor under a dark cloud, i.e. the nagging suspicion that, while the Bayesian philosophy may be a giant leap forward for other applied sciences, it represents a step backward in clinical research. The embrace of prior, untested and subjective information, in combination with the likelihood principle's implications, creates a chaotic research environment in which, once again, it will be difficult to disentangle fact-based result from expert but misleading opinion.

Clinical medicine is overwhelmed with prior belief and opinion, which is many times both honest and inaccurate. Without specific safeguards, use of Bayesian procedures will set the stage for the entry of non-fact-based information that, unable to make it through the 'evidence-based' front door, will sneak in through the back door of 'prior distributions'. There, it will wield its influence, perhaps wreaking havoc on the research's interpretation.

In addition, the likelihood principle, with its insistence on making decisions based only on observations that have occurred and not on those that have not, reduces the moderating influence of established methodology that was put in place to curb the inappropriate generation of sampling error-generated results.

In essence, Bayesians willfully violate a cardinal rule in clinical trials—the isolation of prior belief from effect size estimates. This predictable dilemma still awaits a coherent response from the Bayesian community.

Frequentists have argued for years that clinical trial researchers must begin with not what they believe, but what has been established as the standard. They then build a fact-based case against this null hypothesis, and the research community requires the clinical trial to affirmatively and

DOI: 10.1002/sim

decisively reject the current community standard of care. This time-tested frequentist approach makes clinical investigators uncomfortable, precisely because it so completely protects us from our own natural weakness—our need to believe that the therapy that we are advocating will work!

We physicians are devoted, intelligent, and driven specialists; we are also quite human. We care about the patients we see with a compassion that is uniquely ours, personified in the oath that we take. As first responders to the Hurricane Katrina debacle, my colleagues and I did not function as dispassionate observers of a human tragedy, we continuously consoled, actively treated, and were deeply affected by the patients so desperate for attention and good medical care [79]. Physicians just do not look for cures, we ache for them! So, when we learn of a potential new treatment, our natural tendency is to give it the benefit of the doubt. It is difficult for physicians to keep in mind how bad things may be with an untested intervention, in the face of the reality of how bad things are without it. The frequentist approach plays the critical role of protecting us from our own strong convictions about the effects of untested therapy.

This edifice separating the physician-scientist's research from his or her own beliefs was not easily built. It was assembled brick by brick through uncountably many debates, discussions and arguments over the 60-year history of clinical trials. Yet, Bayesians, although supported by a doctrine that has been in existence for over 225 years, did not participate in these pivotal discussions during the formative years of clinical trial development. Bayesians were relatively unavailable as frequentists argued repeatedly and heatedly with physicians that clinical research methodology should be structured, that research questions should be stated objectively, that null findings should be as easily interpretable as positive ones, and that we physicians must be concerned about the possibility of therapy-induced harm even when we 'believe' that no harm will occur. Bayesians were curiously absent for the five decades that these arguments raged furiously.

Yet now, when the investigator community finally understands the problems that stem from relying on their own untested assumptions, and appreciates that the research benefits of a disciplined methodology outweigh its risks, Bayesians burst onto the scene with a mathematical justification for the use of prior information. The havoc from this confusion is something that the iconoclast Reverend Price may have delighted in, but neither Bayesians nor their frequentist counterparts can afford this now.

Frequentists have shed their intellectual and emotional blood building a wall that separates the clinical investigators' research execution and analysis from their prior belief. General George S. Patton may have been right when he said that 'fixed fortifications are monuments to the stupidity of man', but, nevertheless, disaster strikes in clinical trials when the wall is breeched. If Bayesian trialists are not vigilant, the work that they have started to break this wall down will be used to their own undoing, and we may very well have to wait to the beginning of the 22nd century for the next Bayesian wave.

CONCLUSION

The following represent seven steps that Bayesians can take to improve their standing in the community as rigorous clinical trialists.

Step 1: *Take a strong stand for disciplined research methodology*. Frequentists in clinical trials have a well-enunciated perspective on research discipline. The perspective of the Bayesian community has not been articulated, and their omnipresent likelihood principle opens the door to the possibility that Bayesians believe in 'letting the data decide' in clinical trials.

DOI: 10.1002/sim

Bayesians must send an unmistakably clear message through the clinical trial research community about the importance of methodologic rigor in clinical research. They cannot afford to stand by idly while loosely trained clinical investigators convert the impassioned Bayesian argument for inclusion of prior information into a lever to pry open the entire case of disciplined research methodology. Bayesians stand to lose all they have worked so hard for by being asleep at the switch during this important junction. Bayesian methodologists must act at once to ensure the scientific community that Bayesian clinical trials must be expertly designed and concordantly executed.

Step 2: *Fix the mistranslation*. While Bayesians have made an elegant and mathematically persuasive argument for the incorporation of prior information into clinical research, they have not sufficiently differentiated between prior information and prior belief. Thus, clinical investigators commonly think that their untested and sometimes wild opinions can be formally incorporated into research as 'prior knowledge'. This is certainly not the message Bayesians mean to transmit, yet it is the message many clinical investigators manage to receive. In this regard, Bayesians stand to make the same mistake that frequentists made with the p-value. It was never the intent of the field of frequentist biostatistics to allow a small p-value to 'cover a host of methodologic sins', e.g. small sample size and data-driven protocol deviation. Yet because frequentists were asleep at the switch during that critical junction, the clinical research community must still pay the price for a research climate contaminated by the '0.05 culture'. By remaining mute, Bayesians run the risk of making an equally catastrophic error.

Step 3: *Show us something new.* The research paradigm should offer an illumination of biologic processes and therapy–human interaction. For example, it was the clinical trial paradigm that revealed the 'placebo effect'. What such revelation will come from the application of Bayesian processes? If there are none, then the Bayesian approach will simply be a tepid alternative to the frequentist perspective, much of whose own luster has been rubbed away by years of difficult clinical trial experience.

Step 4: *Develop realistic prior distributions and loss functions.* Many didactic tools begin with simple priors and loss functions which illuminate Bayesian principles. However, the real clinical world is a messy place. In order to be truly illuminating, Bayesians need to move away from the safe haven of uniform prior probabilities, conjugate distributions, and squared error loss. Very few results can be as impressive as a new research insight that comes from a Bayesian clinical trial with an illuminating prior distribution and a sensitively constructed loss function. For example many scientists would rush to embrace Bayesian procedures if Bayesians could construct, however complex, realistic prior distributions and loss functions that produced real-time Bayes procedures informing public health officials when to evacuate populations in the face of a hurricane. Similarly, we need prior distributions and loss functions that do not just support, but also illuminate clinical research.

Step 5: *Actively incorporate counter-intuitive prior information.* Experienced clinical trial biostatisticians know to follow the peroration of a clinical opinion leader with a healthy dollop of skepticism. These opinions of clinical thought leaders must not be allowed to imbalance the prior distributions Bayesians incorporate in clinical trials. To many ill-trained clinical scientists, prior information is an opportunity to infuse non-fact-based 'information' into a trial to influence the results. A useful rule of thumb might be, the more passionate the investigator, the greater the protection the priors require from their strongly held opinion. This role would be filled by not simply a non-informative prior, but a counter-intuitive prior, to actively counterbalance the strong, frequently wrong opinions of clinical experts.

DOI: 10.1002/sim

Step 6: *Banish all discussion of the so-called 'improper priors' from clinical research.* Clinical scientists have worked hard to master the concepts of probability sufficient to understand the testing multiplicity dilemma, proportional hazard regression analysis, and the statistical monitoring of clinical trials. Probability that does not sum to one injects tremendous confusion into an environment in which doctors are already struggling with the proper incorporation of prior belief into their research. Improper priors are a mathematical nicety but an unnecessary clinical conundrum, distraction, and nuisance.

Step 7: *Develop good Bayesian primers for clinical investigators.* Bayesian mathematics, while elegant, are simply too dense at the level they are commonly taught for many physicians with only a remote exposure to calculus and a one-semester course in statistics. Bayesians need to write clear simple expositions on what they stand for, and how clinical trials conducted under the Bayesian paradigm are different from the current state-of-the art frequentist clinical trial methodology. This can begin during the early basic science years of medical school, when students can learn to segregate belief from knowledge (the first frequentist course in statistics is commonly offered during the first two years of medical school). In addition, the Bayesian paradigm can be demonstrated during a one-semester course in biostatistics that research assistants and clinical trial project managers commonly complete. Designing such a curriculum will be daunting, but there is much at stake. An early introduction can provide uninitiated students with a smooth learning curve easing their way to becoming seasoned Bayesian clinical investigators.

ACKNOWLEDGEMENT

This work was supported in part by National Center for Research Resources grant US4-RR023417-01.

REFERENCES

1. Berry DA. A case for Bayesianism in clinical trials (with discussion). *Statistics in Medicine* 1993; **12**:1377–1404.
2. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society Series A* 1994; **157**:357–416.
3. Diamond GA, Kaul S. Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. *Journal of the American College of Cardiology* 2004; **43**:1929–1939.
4. Spiegelhalter DJ, Freedma LS, Parmar MK. Applying Bayesian ideas in drug development and clinical trials. *Statistics in Medicine* 1993; **12**:1501–1511.
5. Cui L, Hung HMJ, Wang S. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; **55**:853–857.
6. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics* 1999; **55**: 1286–1290.
7. Muller H, Schafer H. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* 2001; **57**:886–891.
8. Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics* 1995; **51**:1315–1324.
9. Cheng Y, Su F, Berry DA. Choosing sample size for a clinical trial using decision analysis. *Biometrika* 2003; **90**:923–936.
10. Moyé LA. Sizing clinical trials with variable endpoint event rates. *Statistics in Medicine* 1997; **16**:2267–2282.
11. Korn EL, Baumrind S. Clinician preferences and the estimation of causal treatment differences. *Statistical Science* 1998; **13**:209–235.
12. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science* 2004; **19**(1): 175–187.
13. Speigelhalter DJ, Freedman LS, Blackburn PR. Monitoring clinical trials: conditional or predictive power. *Controlled Clinical Trials* 1986; **7**:8–17.

DOI: 10.1002/sim

14. Nicholl J, Goodacre S. Bayesian stopping rules for trials. *Lancet* 2002; **359**:76–77.
15. Freedman LS, Spiegenhalter DJ. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clinical Trials* 1989; **10**:357–367.
16. Freeman LS, Spoiegelhbalter DJ, Permar MK. The what, why, and how of Bayesian clinical trial monitoring. *Statistics in Medicine* 1994; **13**:1371–1383.
17. Ashby D, Hutton J, McGee M. Simple Bayesian analysis for case–control studies in cancer epidemiology. *The Statistician* 1993; **42**:385–397.
18. Dunson DB. Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology* 2001; **153**:1222–1226.
19. Kmetic A, Joseph L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. *Epidemiology* 2002; **13**:437–444.
20. Berrry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials—a decision analysis. *Statistics in Medicine* 1995; **14**:231–246.
21. Berry DA, Müller P, Grieve AP, Smith M, Parke T, Blazek R, Mitchard N, Krams M. Adaptive Bayesian designs for dose-ranging drug trials. *Case Studies in Bayesian Statistics V*. Lecture Notes in Statistics, vol. 162. Springer: New York, 2002; 99–181.
22. Berry D. A guide to drug discovery; Bayesian clinical trials. *Nature Reviews Drug Discovery* 2006; **5**:27–36.
23. Berger JO. *Statistical Decision Theory. Foundations*, *Concepts*, *and Methods*. Springer: New York, 1980.
24. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods of Data Analysis* (2nd edn). Chapman & Hall, CRC Press: London, Boca Raton, 2000.
25. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayes Data Analysis* (2nd edn). Chapman & Hall, CRC Press: London, Boca Raton, 2003.
26. Speigelhalter DJ, Abrams, KR, Myles JP. *Bayesian Approach to Clinical Trials and Health-Care Evaluation*. Wiley: New York, 2004.
27. Stigler SM. *The History of Statistics*: *The Measurement of Uncertainty before 1990*. The Belknap Press of Harvard University Press: Cambridge, London, 1986; P98.
28. Zabell S. R.A. Fisher on the history of inverse probability. *Statistical Science* 1989; **4**:247–256.
29. Zabell S. R.A. Fisher on the history of inverse probability. Rejoinder. *Statistical Science* 1989; **4**:261–263.
30. Fisher RA. *Statistical Methods For Research Workers*. Oliver and Boyd: Edinburg, 1925.
31. Fisher RA. Two new properties of mathematical likelihood. *Proceedings of the Royal Society Series A* 1934; **144**:285–307.
32. Neyman J, Peason ES. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society* (*London*) *Series A* 1933; **231**:289–337.
33. Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inferences: parts I and II. *Biometrika* 1933; **20**:175–240, 263–294.
34. Pytkowsi W. *The Dependence of the Income in Small Farms upon their Area*, *the Outlay and the Capital Invested in Cows* (*Polish*, *English summaries*). Monograph Series Bioblioteka Pulawska, No. 31. Publ. Agricultural Research Institute: Pulawy, Poland.
35. Wald A. *Statistical Decision Functions*. Wiley: New York, 1950.
36. Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society* (*London*) *Series A* 1937; **236**:333–380.
37. Neyman J. L'estimation statistique traitée comme un problème classique de probabilité. *Actualités Scientifiques et Industriels* 1938; **739**:25–57.
38. Neyman J, Peason ES. On the problems of most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, *Series A* 1933; **231**:289–337.
39. Jeffreys H. *Scientific Inference*, reprinted with additions in '37 and with new editions in '57 and '73. Cambridge University Press: Cambridge, 1931.
40. Hartigan J. Invariant prior distributions. *Annals of Mathematical Statistics* 1964; **35**:836–845.
41. Dawid P, Stone M, Zidek JM. Marginalization paradoxes in Bayesian and structural inference (with discussion). *Journal of the Royal Statistical Society Series B* 1973; **35**:189–233.
42. Fisher RA. Inverse probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society* 1932; **28**:257–261.
43. Jeffreys H. On the prior probability in the theory of sampling. *Proceedings of the Cambridge Philosophical Society* 1933; **29**:83–87.
44. Jeffreys H. On the theory of errors and least squares. *Proceedings of the Royal Society of London*, *Series A* 1932; **138**:48–55.

DOI: 10.1002/sim

45. Fisher RA. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London*, *Series A* 1933; **139**:343–348.
46. Jeffreys H. Probability, statistics, and the theory of errors. *Proceedings of the Royal Society of London*, *Series A* 1933; **140**:523–535.
47. Fisher RA. Probability, likelihood and the quantity of information in the logic of uncertain inference. *Proceedings of the Royal Society of London*, *Series A* 1934; **146**:1–8.
48. Jeffreys H. Probability and scientific method. *Proceedings of the Royal Society of London*, *Series A* 1934; **146**:9–16.
49. Savage LJ. The theory of statistical decision. *Journal of the American Statistical Association* 1951; **46**:55–67.
50. Savage LJ. *The Foundations of Statistics* (1972 edn). Wiley: New York, 1954.
51. Tales of Statisticians. LJ Savage (1917–1971). From http://www.umass.edu/wsp/statistics/tales/savage.html
52. Moyé LA. *Statistical Monitoring of Clinical Trials*: *Fundamentals for Investigators*. Springer: New York, 2005.
53. Goodman SN. Toward evidence-based medical statistics. 1: the *p*-value fallacy. *Annals of Internal Medicine* 1999; **130**:995–1004.
54. Walker AM. Significance tests represent consensus a and standard practice (Letter). *American Journal of Public Health* 1986; **76**:1033 (see also Journal erratum;**76**:1087).
55. Fleiss JL. Significance tests have a role in epidemiologic research; reactions to A.M. Walker. (Different views) *American Journal of Public Health* 1986; **76**:559–560.
56. Fleiss JL. Confidence intervals versus significance tests: quantitative interpretation (Letter). *American Journal of Public Health* 1986; **76**:587.
57. Fleiss JL. Dr. Fleiss response (Letter). *American Journal of Public Health* 1986; **76**:1033–1034.
58. Walker AM. Reporting the results of epidemiologic studies. *American Journal of Public Health* 1986; **76**:556–558.
59. Pratt JW. Discussion of A. Birnbaum's On the foundations of statistical inference. *Journal of the American Statistical Association* 1962; **57**:269–326.
60. Moyé LA. *Statistical Reasoning in Medicine*: *The Intuitive P-value Primer* (2nd edn). Springer: New York, 2006.
61. Ramsey F. Truth and Probability in Ramsey 1931. In *The Foundations of Mathematics and other Logical Essays*, Chapter VII, Braithwaite RB (ed.), Degan, Paul, Trench, Trubner & Co.: London, Harcourt, Brace, and Company: New York, 1926; 156–198 (1999 electronic edition).
62. The CAST Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *New England Journal of Medicine* 1989; **3212**:406–412.
63. Moore T. *Deadly Medicine*. Simon and Schuster: New York, 1995.
64. UK Prospective Diabetes Study Group. UK Prospective Diabetes Study (UKPDS) VIII. Study, design, progress and performance. *Diabetologia* 1991; **34**:877–890.
65. Moyé LA. *Multiple Analyses in Clinical Trials*: *Fundamentals for Investigators*, Chapter 8. Springer: New York, 2003.
66. UKPDS Study Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *The Lancet* 1998; **352**:837–853.
67. Turner RC, Holman RR on behalf of the UK Prospective Diabetes Study Group. The UK Prospective Diabetes Study. Finnish Medical Society DUOCECIM. *Annals of Medicine* 1998; **28**:439–444.
68. UKPDS Study Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *The Lancet* 1998; **352**:837–853.
69. Feldman AM, Bristow MR, Parmley WW, Carson PE, Pepine CJ, Gilbert EM, Strobeck JE, Hendrix GH, Powers ER, Bain RP, White BG for the Vesnarinone Study. Effects of vesnarinone on morbidity and mortality in patients with heart failure. *New England Journal of Medicine* 1993; **329**:149–155.
70. Cohn JN, Goldstein SO, Greenberg BH, Lorell BH, Bourge RC, Jaski BE, Gottlieb SO, McGrew F 3rd, DeMets DL, White BG. A dose dependent increase in mortality seen with vesnarinone among patients with severe heart failure. *New England Journal of Medicine* 1998; **339**:1810–1816.
71. Pitt B, Segal R, Martinez FA *et al*. on behalf of the ELITE Study Investigators. Randomized trial of losartan versus captopril in patients over 65 with heart failure. *The Lancet* 1997; **349**:747–752.
72. Jensen BV, Nielsen SL. Correspondence: Losartan versus captopril in elderly patients with heart failure. *The Lancet* 1997; **349**:1473.
73. Fournier A, Achard JM, Fernandez LA. Correspondence: Losartan versus captopril in elderly patients with heart failure. *The Lancet* 1997; **349**:1473.
74. Packer M, O'Connor CM, Ghali JK, Pressler ML, Carson PE, Belkin RN, Miller AB, Neuberg GW, Frid D, Wertheimer JH, Cropp AB, DeMets DL for the Prospective Randomized Amlodipine Survival Evaluation Study

DOI: 10.1002/sim

Group. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. *New England Journal of Medicine* 1996; **335**:1107–1114.
75. Packer M. Presentation of the *Results of the Prospective Randomized Amlodipine Survival Evaluation-2 Trial* (*PRAISE-2*) at the American College of Cardiology Scientific Sessions, Anaheim, CA, March 15, 2000.
76. Egger M, Smith GD, Phillips AN. Meta-analyses: principles and procedures. *British Medical Journal* 1997; **315**:1533–1537.
77. Alderson P, Roberts I. Corticosteroids in acute traumatic brain injury: systematic review of randomised controlled trials. *British Medical Journal* 1997; **314**:1855–1859.
78. Gregson B, Todd NV, Crawford D, Gerber CJ, Fulton B, Tacconi L, Crawford PJ, Sengupta RP. CRASH trial is based on problematic meta-analysis. *British Medical Journal* 1999; **319**:578.
79. Moyé L. *Face to Face with Katrina Survivors*: *A First Responder's Account*. Open Hand Publishing: Greensboro, 2006.

DOI: 10.1002/sim