Stroke

TOPICAL REVIEW

Section Editors: Seana Gall, PhD, Linxin Li, DPhil, and Mathew J. Reeves, PhD

An Introduction to Bayesian Approaches to Trial Design and Statistics for Stroke Researchers

Johanna M. Ospel, MD, PhD*; Scott Brown, PhD*; Jessalyn K. Holodinsky, PhD; Leon Rinkel, MD; Aravind Ganesh, MD, DPhil; Shelagh B. Coutts, MD; Bijoy Menon, MD; Benjamin R. Saville, PhD; Michael D. Hill, MD; Mayank Goyal, MD, PhD

ABSTRACT: While the majority of stroke researchers use frequentist statistics to analyze and present their data, Bayesian statistics are becoming more and more prevalent in stroke research. As opposed to frequentist approaches, which are based on the probability that data equal specific values given underlying unknown parameters, Bayesian approaches are based on the probability that parameters equal specific values given observed data and prior beliefs. The Bayesian paradigm allows researchers to update their beliefs with observed data to provide probabilistic interpretations of key parameters, for example, the probability that a treatment is effective. In this review, we outline the basic concepts of Bayesian statistics as they apply to stroke trials, compare them to the frequentist approach using exemplary data from a randomized trial, and explain how a Bayesian analysis is conducted and interpreted.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: Bayes theorem ■ observational studies as topic ■ sample size ■ statistics ■ stroke

Currently, the majority of stroke research results are reported using a frequentist approach. Frequentist statistics tell us how likely it is to see a result that is as extreme as, or more extreme than, the observed result given an underlying assumed distribution.[1] One key concept in frequentist studies is hypothesis testing, whereby the null hypothesis usually denotes that there is no difference between 2 groups: for example, the null hypothesis could be that there is no difference in clinical outcomes between the control and treatment arms of a clinical trial. One key measure in frequentist hypothesis testing is the *P* value; that is, the probability of obtaining a result as extreme or more extreme as the observed result given the null hypothesis is assumed to be true. If this probability is less than a prespecified threshold (conventionally $P<0.05$ is used as a threshold), we reject the null hypothesis.

See related articles, p 2726, p 2731

Frequentist analyses allow for the incorporation of existing evidence at the design stage but not at the analysis stage. For example, when the sample size of a randomized trial is determined during the trial planning phase, investigators must estimate the treatment effect and corresponding variability based on existing evidence from prior studies, observational data, or their own clinical experience. Once the trial protocol is finalized, prior information does not have any further impact.

From a technical standpoint, frequentist analyses are often misinterpreted. For example, failure to reject the null hypothesis simply means that we have failed to find evidence against it, but it is not proof of a lack of an effect (absence of evidence is not evidence of absence).[2,3] To add another example, the 95% CI often used in frequentist statistics does not indicate that the population parameter falls within the confines of that interval 95% of the time (this is in fact the definition of the 95% credible interval that is used in Bayesian statistics). Rather, the frequentist

This article is Part 3 in a 5-part series. Parts 1–3 appear in the November 2024 issue of *Stroke*. Parts 4 and 5 will appear in subsequent issues.

Correspondence to: Mayank Goyal, MD, PhD, Departments of Radiology and Clinical Neurosciences, Foothills Medical Centre, University of Calgary, 1403 29th St NW, Calgary, AB T2N2T9, Canada. Email mgoyal@ucalgary.ca
*J.M. Ospel and S. Brown contributed equally.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.123.044144.
For Sources of Funding and Disclosures, see page 2752.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

DAWN	Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo
ENRICH	Early Minimally Invasive Removal of Intracerebral Hemorrhage
EVT	endovascular treatment
FDA	US Food and Drug Administration
OR	odds ratio
PROPPR	Pragmatic Randomized Optimal Platelet and Plasma Ratio

95% CI is defined as follows: if the experiment were to be exactly replicated 100 times, with 100 distinct estimates and CIs, we expect 95 of the 100 intervals to capture the population parameter within them.[4] Thus, a 95% CI denotes a 95% probability that the interval captures the population parameter, rather than the probability that the population parameter falls within the interval.

The Importance of the *P* Value

The *P* value is a key measure of the hypothesis test. It describes how likely it is to see results as extreme as, or more extreme than, the observed results if the null hypothesis is true (eg, if there is no difference between 2 groups or no effect of the treatment). It does not provide information on the magnitude of the observed difference or treatment effect although it is sometimes mistaken as such. A common misunderstanding of the *P* value is that it represents a predictive value (ie, the probability that the null hypothesis is false). Limitations of using the *P* value as a means to draw inferences from trial data have been reviewed in an earlier paper in this series (Reeves et al: The Changing Landscape of Randomized Clinical Trials in Stroke: A Series of Commission Papers That Explain Contemporary Trial Designs and Methods).

Although not unique to frequentist statistics, the common approach of dichotomizing research results into statistically significant versus nonsignificant is flawed. On the one hand, only a small change is required to move a nonsignificant result to a significant result (eg, it is plausible that a change of outcome status in 1 single trial participant could change the *P* value from 0.051 to 0.049; the results are significant but fragile[5]). Thus, it is always preferable to report exact *P* values rather than simply stating that a result was significant or not significant. For example, a *P* value of 0.08 should be interpreted much differently than a *P* value of 0.87 although both are nonsignificant. Ironically, as a result of this dichotomization at the 0.05 level, the difference between significant and nonsignificant results is often itself not significant.[6] Dichotomization is not a problem of frequentist statistics themselves but rather a problem of how they are interpreted in practice, where a strong, universal emphasis is put on a frequentist *P* value cutoff of 0.05 to denote a statistically significant result.

Frequentist statistics are currently the most commonly used analytical approach in medical research, and most stroke researchers and clinicians are familiar with the way frequentist study results are presented and interpreted. In many instances, a frequentist *P* value is well-suited to answer the research question if the study is testing a specific hypothesis. However, often, the research question is actually to estimate the magnitude of the treatment effect with a certain precision. Analysis and reporting should then focus on estimation and 95% CIs (a common theme of modern journals). In addition, it is more appropriate to analyze and interpret study results in the context of preexisting or coexisting evidence and beliefs rather than in isolation.[7]

KEY PRINCIPLES OF BAYESIAN STATISTICS AND COMPARISON TO THE FREQUENTIST APPROACH

Research does not take place in a vacuum, and contextualizing study results by incorporating existing evidence can be helpful. A Bayesian approach to statistics enables researchers to do exactly this: evidence from prior studies can formally be incorporated into the analysis in the form of priors.[8] Although this is also true of frequentist adaptive study designs, the accumulating evidence from the current, ongoing study can also be used to modify the study in a prespecified manner (Bayesian adaptive study design). Key concepts of Bayesian and frequentist approaches are explained in Table 1.

A key advantage is that the Bayesian approach provides a probabilistic interpretation of model parameters. In contrast to *P* values, a Bayesian analysis directly informs us how likely an experimental treatment is beneficial in a study population. Furthermore, it allows researchers to calculate the Bayes factor, a likelihood ratio that compares how well the null hypothesis and alternative hypothesis predict that data[9] (see detailed explanation in the Supplemental Appendix).

At a high level, a Bayesian scientist starts with a belief about a parameter (eg, treatment benefit) based on preceding evidence or experience, collects new data, and subsequently updates that belief based on the observed data. Technically, this is done by formalizing prior knowledge (eg, from preexisting studies) in a prior probability distribution called, the prior. The new data observed in the current study are then formalized in a second probability distribution, the likelihood. These 2 distributions are mathematically combined to form a third, posterior probability distribution. The underlying assumption (prior) is updated by the observed data (likelihood), and the

 Stroke. 2024;55:2742–2753. DOI: 10.1161/STROKEAHA.123.044144

Table 1. Key Terms Used in Bayesian Analysis and Their Definitions

Key term	Definition
Frequentist methods	
Null hypothesis	Hypothesis intended to disprove, expressed as a population parameter equal to the null value (typically 0). For example, the hypothesis that the mean difference between treatment groups equals 0.
Alternative hypothesis	Hypothesis intended to prove, expressed as a population parameter, is different than the null value. For example, the hypothesis is that the mean difference in treatment groups is >0.
P value	Probability of obtaining a result as extreme or more extreme than the observed result given the null hypothesis is true.
95% CI	If an experiment were to be exactly replicated 100×, with 100 point and interval estimates generated, we expect 95 of the generated intervals to capture the (unknown) population parameter within them. Therefore, a 95% CI implies a 95% probability that the interval captures the population parameter, rather than a 95% probability that the population parameter falls within the interval.
Bayesian methods	
Prior probability distribution ("prior")	Assumed probability distribution of a certain parameter (eg, a proportion of patients or an effect size estimate) before the data from the current study/trial is taken into account.
Posterior probability distribution ("posterior")	Probability distribution for Bayesian inference that is obtained by updating the prior probability distribution with observed data from the current study/trial.
95% credible interval	Interval that contains the true parameter with 95% probability. Therefore, a 95% credible interval implies a 95% probability that the true (unknown) parameter falls within the interval.
Bayes factor	Posterior odds of 1 hypothesis versus another when the prior probabilities are the 2 hypotheses are equal, represented as ratio >0. The calculation involves a ratio of marginal likelihoods that integrate over the parameter prior distributions, which can be difficult to compute.
Bayesian hierarchical model	Bayesian models that incorporate multiple levels of clusters/groups in the data, thereby allowing for partitioning of sources of variability (eg, variability at the study, patient subgroup, individual patient level). Bayesian hierarchical models also allow for shrinkage estimation/partial pooling.
Shrinkage estimation/partial pooling	Statistical technique that provides subgroup-specific estimates that leverage information from other relevant subgroups, often increasing the precision of the subgroup estimate relative to stand-alone estimates.

updated assumption is called the posterior distribution. The mathematical formulation of this process is made possible via the Bayes theorem, given by the following formula in which *H* represents a hypothesis:

$$\text{Probability}\,(H|\text{Data}) = \frac{\text{Probability}\,(\text{Data} \mid H) * \text{Probability}\,(H)}{\text{Probability}\;\;(\text{Data})}.$$

The Bayes theorem states that the probability of a hypothesis of the given observed data is equal to the data likelihood times the prior probability of the hypothesis, divided by the unconditional probability of the data. The posterior distribution forms the foundation of Bayesian inference, including estimation and hypothesis testing. The Bayesian posterior allows probabilistic interpretation of hypotheses and parameters; for example, the posterior can provide the estimated probability that a treatment is effective.

Because of the ability to formally combine prior information with observed data to formulate a posterior belief, the Bayesian approach can be seen as the mathematical embodiment of clinical diagnostic or therapeutic reasoning: imagine a stroke neurologist on call who has a certain base assumption of the probability that a code stroke will result in a thrombectomy. Let us assume that based on many years of clinical experience, the stroke neurologist knows that on average, 15% of all code strokes will undergo thrombectomy. When the code stroke is triggered, she assumes that the probability of this particular code patient with a stroke undergoing a thrombectomy is roughly 15%. After examining the patient who has a severe deficit with a National Institutes of Health Stroke Scale score of 20, she may update her estimate: the additional information makes it much more likely that this patient will undergo a thrombectomy. After having obtained these additional data, the neurologist now estimates the probability of thrombectomy at 60%. The patient may then undergo a computed tomography angiography, which shows a large vessel occlusion: it has now become even more likely that the patient will undergo a thrombectomy, and the neurologist may update her previous estimate once more, from 60% to 90%. While this example pertains to individual decision-making in the clinical setting, this example illustrates how the principle of updating prior beliefs based on new data is a natural part of scientific reasoning and clinical practice.

Choosing the Right Prior: A Key Requirement for Bayesian Statistics

While the possibility to incorporate prior information into the analysis is viewed by many as a great advantage of Bayesian statistics (relative to frequentist statistics), the Bayesian reliance on priors can be alternatively viewed by others as a disadvantage. A common criticism is that the choice of the priors is subjective and may have a large influence on the interpretation of the Bayesian analysis.[10]

In general, the more informative the prior (eg, the larger the sample size, the smaller the variance, and the

larger the observed effect size in the prior study), the greater the influence on the analysis results. Generally speaking, a Bayesian analysis based on a noninformative prior will require a similar sample size as a frequentist analysis. However, Bayesian analyses with informative prior distributions often require smaller sample sizes than traditional frequentist analyses. In most instances, prior knowledge about treatment effects is provided in the form of absolute differences or relative effect size measures and corresponding uncertainty. Prior distributions are chosen by the researcher and take different forms: priors may be symmetrical, skewed, bimodal, or flat. Often for convenience, a normal or log-normal distribution is assumed on treatment effect parameters. For example, when conducting an analysis on a ratio, a prior for the log ratio can be specified with a mean (μ) and a standard deviation (SD; Figure 1).

Neutral, Optimistic, and Pessimistic Priors

A neutral prior is defined as a prior distribution that is centered at zero effect and considers a positive and negative treatment effect equally likely (in the example of a log odds ratio [OR], that number would be 0, corresponding to an OR of 1; Figure 1A). Optimistic priors assign a higher probability to a positive treatment effect (ie, they have centered at a log odds >0, corresponding to an OR >1; Figure 1B) but still allow for the possibility that there is a zero or negative treatment effect. Pessimistic priors are the opposite of optimistic priors: they assign a higher probability to a negative treatment effect (they are centered at a log odds <0, corresponding to an OR <1) while still allowing for the possibility that there is a zero or positive treatment effect (Figure 1B).

Weakly Informative and Strongly Informative Priors

While the terms neutral, optimistic, and pessimistic describe beliefs with regard to the direction of effect (ie whether there is no effect, benefit, or harm and where the mean of the prior distribution lies along this spectrum), they do not quantify the strengths of these beliefs. The latter is captured in the standard deviation of the prior distribution. Weakly informative prior describes a prior with a large standard deviation in relation to its mean, that is, a widespread distribution (Figure 1). They are usually centered at zero effect. While they are sometimes referred to as noninformative priors, they do convey some information about the range of possible values; therefore, weakly informative prior is the preferred terminology, with 1 exception: a flat prior distribution literally takes the shape of a flat line; that is, it has an infinite SD and assigns every treatment effect the same probability. Unsurprisingly, when using a weakly informative or uninformative flat prior, the results of a Bayesian analysis will resemble the results of a frequentist analysis although their interpretation is notably different. Strongly informative priors on the other hand have a narrower standard distribution; that is, the range in which the treatment effect is assumed to fall is smaller.

The weight of a prior conveys how much information is contained in a prior: the more the weight is assigned to the prior (ie, the more informative the prior is), the more

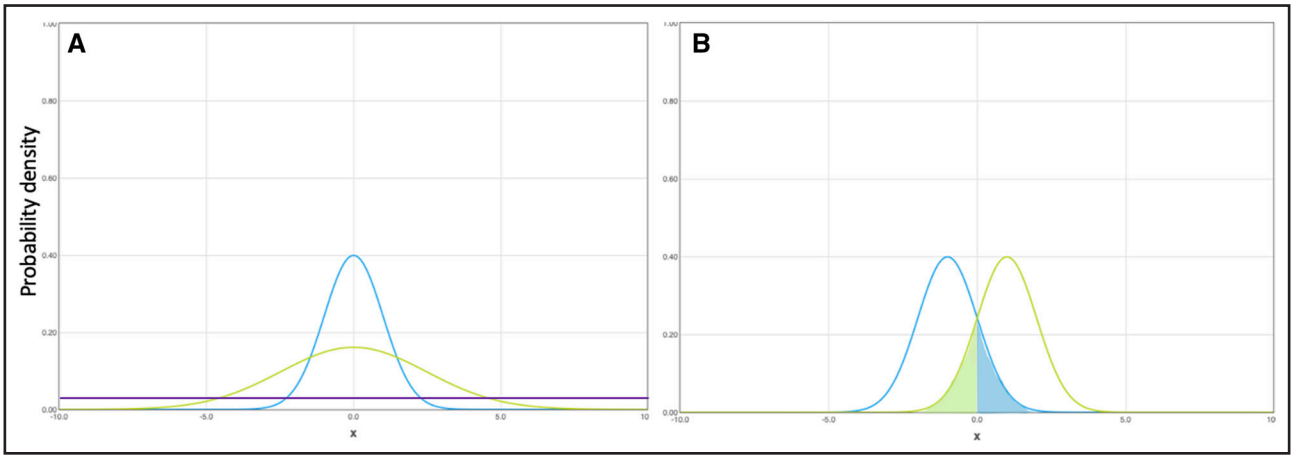

Figure 1. Overview of different prior distributions.
Examples of different prior distributions, namely, neutral priors (**A**) and optimistic and pessimistic priors (**B**). Treatment effect is plotted on the *x* axis, and the probability density is plotted on the *y* axis. The green line in (**A**) shows a weakly informative neutral prior distribution, with a mean of zero and a rather wide standard deviation. The blue line in (**A**) shows a strongly informative neutral prior with a mean of zero and a rather small standard distribution. Compare the shapes of the blue and red curves. The blue curve indicates much higher probabilities of a narrow interval centered around zero than the green curve. The purple line in (**A**) denotes an uninformative (flat) prior that literally assumes the shape of a flat line and assigns each treatment effect equal probability. The green line in (**B**) shows a strongly informative, optimistic prior. The prior is centered around a positive treatment effect (in this case, 1), but it does still account for a possible negative treatment effect (green-shaded area under the curve). The blue line in (**B**) shows a strongly informative, pessimistic prior. The prior is centered around a negative treatment effect (in this case, −1), but it does still account for a possible positive treatment effect (blue-shaded area under the curve).

influence the prior has on the posterior distribution. The weight of a prior is inversely related to the variance of the prior distribution; that is, multiplying the variance by 2 will halve the weight of the prior distribution. It is important to note that a flat prior distribution on a given scale (eg, a log OR scale) can translate into an informative prior on a transformed scale (eg, OR scale) or, alternatively, if the outcome is measured in a different metric. Hence, consideration should be given to understanding the scale and impact of a prior distribution. Of note, the definitions of weakly informative, optimistic, or pessimistic priors will depend greatly on the specific context of each trial and will be affected by both the center (eg, mean/median) and the variance of the prior distribution. In this regard, virtual trial simulations are often used in the design stage to help understand and evaluate the impact of prior distributions.

How to Define a Prior?

The need to define a prior probability distribution is unique to Bayesian statistical approaches, and as such, researchers have to face the question of what constitutes the best possible prior for their study. To avoid bias, the choice of the prior should be prespecified in a statistical analysis plan in detail before looking at the data. If priors are specified post hoc, that is, after the researchers have looked at their data, priors may be chosen in a biased way, and in the worst case, selective choice of prior results could be used to manipulate study results.

The most common approach to Bayesian analyses in modern research (and particularly clinical trials) is to apply weakly informative priors, which typically allows the data to dominate the posterior inference. In most clinical situations, it seems logical to rely on weakly informative priors. After all, the researchers chose to undertake a particular trial or study because they think that there is scientific equipoise; that is, the answer to the study question is not already known. The prior that best reflects this equipoise is a weakly informative prior that allows the data collected in the trial to dominate the posterior inference. If there is enough prior evidence to construct a strongly informative prior, probably there was no sufficient equipoise to conduct the trial/study in the first place. In some instances, it is reasonable to select more than 1 prior and present a set of analyses. For example, 1 weakly informative prior centered around zero effect, and 1 optimistic prior (ideally based on prior studies) and 1 pessimistic prior may be presented.

Prior data could stem from previous trials investigating the same treatment (eg, when a phase III trial is conducted, the results of a phase II trial could be used to inform the prior distribution), previous trials investigating different but mechanistically similar treatments (eg, similar but nonidentical thrombolytic drugs, different formulations, or application routes of the same drug), high-quality observational studies (particularly in rare diseases for which randomized trial data are available), or expert consensus.

Relative Weighing of Prior Information

Usually, there is more than 1 existing study that contains relevant prior information. Once researchers have identified prior research that contains relevant prior information (being careful not to select studies in a biased manner), these studies can be combined into a single prior distribution or at least boiled down into a small set of prior distributions. The challenge is that not all studies may be equally relevant: some may be more similar to the current study than others, have a larger sample size than others, have a lower or higher risk of bias, etc. The relative importance of prior studies has to be established, and different techniques exist to combine and weigh prior information.

A random-effects meta-analysis can be used, whereby the weight assigned to a particular study is inverse to the deviation of its result from the pooled effect size estimate—the more extreme the results are in comparison to the pooled estimate, the lower the study weight.[11]

Power priors are another method to weigh historical data[12]: first, the heterogeneity between current and historical data is quantified, and a power prior distribution is then constructed by raising the likelihood function of the historical data to the power a_0, whereby $0 \leq a_0 \leq 1$. $A_0=0$ denotes a weakly informative prior, and $a_0=1$ gives full weight to the historical data (full borrowing). The weight a_0 is chosen by the researchers, and usually, a set of analyses using different values of a_0 is presented.

In many situations, the appropriate weight of information incorporated, or borrowed (eg, a_0), from historical data may be unclear. For example, consider a randomized trial in pediatrics for a rare disease that incorporates a prior distribution of a treatment effect from an adult-based study. Because the primary analysis model needs to be completely prespecified, and it is unknown how similar the treatment effect will be in the pediatric population, one could specify a prior distribution that allows dynamic borrowing.[13] The basic idea of dynamic borrowing is to determine the appropriate amount of information to be borrowed based on the observed data. If observed data are similar to the historical data, more information is borrowed by giving more weight to the prior distribution. If observed data are different from historical data, less information is borrowed by giving less weight to the prior distribution.

The Delphi technique, an iterative forecasting technique in which a panel of experts works toward a consensus by repeatedly answering structured questionnaires,[11] inserts clinical judgment into weighting. One problem that arises with this technique is that experts may be biased and assign greater weight to those studies that influence the prior distribution in the direction that is more likely

to result in the desired/expected posterior distribution. Elicitation of study weights is one of the most underresearched problems in Bayesian statistics.[11]

Whatever methods that researchers choose to apply to synthesize prior evidence, we suggest to explicitly list the sources from which informative priors were generated, either in the article text, a table, or a supplement, and to describe the criteria by which they were selected and the rationale behind the selection process. Key parameters, including the range, of prior distributions should be clearly defined and justified based on existing evidence, logic, or reasonable assumptions that are explicitly stated.

INFLUENCE OF BAYESIAN PRIORS IN BAYESIAN POST HOC REANALYSES OF TRIALS

Post hoc Bayesian reanalyses of randomized trials with frequentist design are becoming increasingly common, particularly in scenarios in which a clinically meaningful effect size was observed that barely missed the frequentist 0.05 significance mark. As an example, the PROPPR trial (Pragmatic Randomized Optimal Platelet and Plasma Ratio) randomized patients with severe trauma to transfusion of plasma, platelets, and red blood cells in a 1:1:1 ratio versus a 1:1:2 ratio. Based on the prespecified frequentist analysis, the trial failed to demonstrate the benefit of 1 transfusion strategy over the other with regard to mortality, the primary outcome.[14] Lammers et al[15] reanalyzed the trial data post hoc using a Bayesian approach with weakly informative priors and found a high probability of the 1:1:1 transfusion scheme being superior with regard to mortality. They concluded that a post hoc Bayesian analysis of the PROPPR Trial found evidence in support of mortality reduction.[15] However, such a scenario should generally be avoided unless there are compelling arguments to perform such analyses because choosing priors post hoc is prone to generating biased posterior distributions. Examples of trials in stroke with an adequately prespecified Bayesian analysis include DAWN (Thrombectomy 6 to 24 Hours After Stroke With a Mismatch Between Deficit and Infarct Trial)[16] and ENRICH (Trial of Early Minimally Invasive Removal of Intracerebral Hemorrhage).

To illustrate this, we investigate the association of endovascular treatment (EVT) and good clinical outcome (modified Rankin Scale score, 0–2 at 90 days) in a subset of the ESCAPE trial (Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke), in which patients with acute ischemic stroke with large vessel occlusion were randomized to undergo EVT in addition to the best medical management versus the best medical management alone.[17] To this end, a binary logistic regression model with adjustment for patient age and baseline National Institutes of Health Stroke Scale is used to determine the effect size of EVT on 90-day good clinical outcomes. First, we reduce the ESCAPE trial sample size by randomly dropping observations (but maintaining the 1:1 ratio between the EVT arm and control arm) until a traditional frequentist logistic regression does not show a statistically significant EVT treatment effect anymore; this is the case at a sample size of 48 patients (24 in the EVT arm and 24 in the control arm). Importantly, although, at this sample size, the trial would have just failed to show a statistically significant result, there is still a strong nominally positive—albeit not statistically significant—treatment effect, again illustrating the limitation of the definition of statistically significant. It is in situations like the one we artificially created here—at the verge of statistical significance and in cases with small sample sizes—where the prior choice can often really make a difference, and this is what we will illustrate with this example. Then, we perform different sets of Bayesian logistic regression analyses with varying priors to illustrate how the results could differ based on the choice of the prior (the results are summarized in Table 2). The following analyses were conducted in Stata 17.0 (Stata LLC Corp) using the bayes and bayesmh commands although Bayesian analyses can also be conducted in other statistical software programs, for example, with the JAGS (https://mcmc-jags.sourceforge.io/) or STAN (https://mc-stan.org/) and relevant packages in R (http://www.rproject.org). An open-source

Table 2. Treatment Effect of Endovascular Thrombectomy on Good Clinical Outcome at 90 Days in a Random ESCAPE Trial Patient Sample (n=48 With 24 Patients From Each Arm)

Analysis	Prior	Adjusted OR*	95% confidence/ credibility interval†	Posterior probability of the adjusted OR being >1
Frequentist	…	1.95	0.47–8.13	…
Bayesian	Weakly informative (normal distribution with mean 0 and variance 2)	2.34	0.49–7.11	0.826
Bayesian	MR CLEAN full weight (actual standard deviation, 0.51)	2.57	0.90–5.89	0.956
Bayesian	EXTEND-IA full weight (actual standard deviation, 0.55)	3.42	1.20–7.98	0.989
Bayesian	EXTEND-IA half weight (2× standard deviation, 1.10)	3.20	0.91–8.18	0.967

ESCAPE indicates Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke; EXTEND-IA, Endovascular Therapy for Ischemic Stroke With Perfusion-Imaging Selection; MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Therapy for Acute Ischemic Stroke in the Netherlands; and OR, odds ratio.

*Adjusted for patient age and baseline National Institutes of Health Stroke Scale.

†95% CIs are provided for the frequentist regression, and 95% credibility intervals are provided for the Bayesian regression.

software purely dedicated to Bayesian analysis is also available (https://jasp-stats.org).

Frequentist Analysis

After dropping the sample size to 48 patients, the results of the frequentist logistic regression resulted in an adjusted OR for EVT of 1.95 (95% CI, 0.47–8.13; *P*=0.358; Table 2). The CI crosses 1, and the *P* value is 0.358, indicating that the effect of EVT on good clinical outcomes is not statistically significant when using a frequentist interpretation. The ESCAPE trial did in fact enroll 316 patients and did show a significant benefit of EVT,[17] but, for now, let us assume that the trial enrolled only 48 patients and this is the outcome of the primary trial analysis.

Bayesian Post Hoc Analysis With Weakly Informative Prior

We now use a weakly informative prior in the form of a normal distribution with a log odds mean of 0 and a variance of 2. Results of a Bayesian post hoc analysis with weakly informative priors are shown in Table 2. In this scenario, the influence of the prior is minimal. The adjusted OR of EVT is 2.34 (95% credibility interval, 0.49–7.11). The 95% credibility interval crosses 1.0, and the posterior probability of the OR being >1 is 0.826, which is less than the traditional 0.975 Bayesian threshold required for statistical significance. Remember that the Bayesian 95% credible interval contains the true effect size estimate with a 95% probability (which is often assumed to be true but is not actually true for the frequentist 95% CI), and the posterior probability is simply the updated probability of an event occurring (in this case of the OR being >1) after taking into account the new data, in this case, the ESCAPE patient data. The Bayesian significance threshold of 0.975 for the posterior probability using weakly informative prior distributions typically aligns with the traditional frequentist 1-sided significance level of 0.025 or equivalently a 2-sided significance level of 0.05. In other words, this particular Bayesian analysis is a statistically nonsignificant result using traditional decision criteria. Note that the effect size estimate is different from the frequentist analysis (most likely due to random variance introduced by the small sample size), and the 95% confidence and 95% credibility intervals are similar and relatively wide: both analyses suggest a similar plausible range for EVT effect size.

Bayesian Analysis With Informative Prior—MR CLEAN

Now, we conduct a second post hoc analysis with an informative prior (ie, a prior expressing specific information about the effect size estimate by assigning certain values a higher probability than others), for example, from the MR CLEAN trial (Multicenter Randomized Clinical Trial of Endovascular Therapy for Acute Ischemic Stroke in the Netherlands), another randomized trial that showed significant benefit of EVT, with an adjusted OR of 2.16 (95% CI, 1.39–3.38).[18] When incorporating the MR CLEAN EVT OR as a prior with full weight (full borrowing: the raw standard error of the trial effect size estimate is used) for the association of EVT and good outcome, we would expect the credibility interval to become smaller and perhaps, and it may not include 1.0 (after all, we are borrowing information from the MR CLEAN trial that showed a statistically significant EVT benefit). Incorporating the MR CLEAN prior results in an adjusted OR of 2.57 and a 95% credible interval of 0.90 to 5.89. The posterior probability of the OR being >1 is 0.956, which is less than the traditional 0.975 Bayesian threshold for statistical significance but may be thought of as being borderline significant using traditional decision criteria.

Bayesian Analysis With Informative Prior—EXTEND-IA

To further illustrate the influence of the prior choice, we now use an alternative prior, namely, the EXTEND-IA trial (Endovascular Therapy for Ischemic Stroke With Perfusion-Imaging Selection), in which EVT was also beneficial with regard to good clinical outcome, with an adjusted OR of 4.2 (95% CI, 1.4–12.0).[19] Note that the EXTEND-IA EVT effect size estimate is much larger than the MR CLEAN effect size estimate. Unsurprisingly, incorporating EXTEND-IA as a prior results in an adjusted OR of 3.42 (95% credible interval, 1.20–7.98) and a posterior probability of the OR being >1 of 0.989, which exceeds the traditional 0.975 Bayesian thresholds required for statistical significance. Of note, this analysis gives full weight to the prior (full borrowing), assuming that the EXTEND-IA design and patient samples were similar to/exchangeable with ESCAPE. There are, however, many reasons to think that this is not the case: for example, ESCAPE was mostly conducted in North America and used multiphase computed tomography angiography as baseline imaging, while EXTEND-IA was conducted in Australia and used computed tomography perfusion as baseline imaging. Thus, we may choose to give partial, rather than full weight to the prior (partial borrowing). As mentioned earlier, there are systematic approaches to elucidate prior study weights, but let us assume that we simply decide to assign half weight to the prior information from EXTEND-IA by multiplying the prior's standard deviation by 2. This results in an adjusted OR for EVT of 3.20 and a 95% credible interval of 0.91–8.18. The posterior probability of the OR being >1 is 0.967, which is less than the traditional 0.975 Bayesian threshold for significance and may be thought of as being borderline significant. In contrast to *P* values that fail to reach statistical significance, for

example, a 2-sided *P* value of 0.10, the Bayesian posterior probability has a direct probabilistic interpretation regardless of whether it has met statistical significance. In this example, one could conclude based on the Bayesian analysis that there is a 96.7% probability that EVT provides a positive benefit in this population.

The above examples show how sensitive Bayesian analyses can be to the choice of the priors and highlight the importance of sensitivity analyses that explore a range of plausible priors. It also becomes evident from these examples that in the worst case, selective choice of priors and their weights could be used to manipulate study results. What should have been done in the above example is to identify and combine information from all prior studies into a prior distribution using a systematic, predefined approach; it is essential that this approach should have been described in the ESCAPE trial statistical analysis plan prior of the investigators gaining knowledge of the trial results.

The critical reevaluation of frequentist trials in a Bayesian framework using a systematic approach can be of value. We strongly recommend to prespecify secondary Bayesian analyses, and the respective priors that will be used in the trial protocol and statistical analysis plan to avoid biased post hoc reanalyses. Furthermore, there is some value in standardizing the design and reporting of Bayesian reanalyses of clinical trials, perhaps accompanied by a checklist similar to the CONSORT checklist (Consolidated Standards of Reporting Trials) for reporting parallel group randomized trials or the PRISMA checklist (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for reporting systematic reviews. Zampieri et al[20] have proposed standardized approaches to Bayesian reanalysis of clinical trials in critical care and particularly the choice of priors in this context, much of which applies to stroke medicine as well. Perhaps, the most critical point is to use a predefined, standardized set of priors, including weakly informative, optimistic, and pessimistic priors.

BAYESIAN ANALYSIS FOR SUBGROUP ASSESSMENT

In frequentist statistics, treatment effects and outcomes in subgroups are traditionally estimated by the observed effect within that particular subgroup; that is, only data from that patient subgroup are used to obtain a subgroup treatment effect/outcome. Although this is easy to understand, it may not be an ideal method because it can result in imprecise estimates, particularly when subgroups are small.[21] Shrinkage estimation/partial pooling using Bayesian hierarchical models can be used to incorporate data from other subgroups of a trial weighted by their similarity to the subgroup under investigation to increase the precision of the subgroup estimate[22–24] (Figure 2). Imagine, for example, a sex-based subgroup analysis using a new thrombolytic drug in patients with acute ischemic stroke. When investigating the treatment effect in females, it is likely that the data from male patients are not completely irrelevant; however, it is almost certainly less relevant than the information obtained from female patients. If there is such an a priori belief that the treatment will benefit both males and females, but the size of the treatment effect may be different, a Bayesian hierarchical model allows estimation of separate effects with some degree of partial pooling (ie, shrinkage) between the sex-specific treatment effects. Shrinkage estimation can be performed with Bayesian hierarchical models, which can incorporate multiple levels of detail and, thereby, allow us to partition sources of variability, for example, at the study level, the subgroup level, and the patient level (Figure 2A). Shrinkage is based on a weighted average of the overall trial's treatment effect/outcome and the subgroup of interest's treatment effect/outcome. The weights are determined by within-subgroup variability and between-subgroup variability: with greater within-subgroup variability, borrowing of information from other subgroups increases and the estimate moves closer to the overall trial's estimate (Figure 2B), whereas less within-subgroup variability decreases borrowing, and the estimate will move closer to the subgroup estimate (Figure 2C). In general, the larger the number of groups, the better the ability to estimate the within- and between-group variations using hierarchical models.

The ENRICH trial (Early Minimally Invasive Removal of Intracerebral Hemorrhage) is an example, in which Bayesian hierarchical modeling was used to model differential surgical benefit in 2 distinct subgroups and successfully demonstrated the benefit of minimally invasive surgical removal of supratentorial intracerebral hemorrhage compared with medical management[25] in one of the subgroups. The trial evaluated the benefit of the surgery on the utility-weighted modified Rankin Scale score at 180 days in 2 hemorrhage subtypes: (1) basal ganglia hemorrhage and (2) lobar hemorrhage. It included prespecified interim analyses to adapt the sample size and potentially enrich the study population to 1 of the 2 locations if certain prespecified criteria were met. A Bayesian hierarchical model was used for the interim adaptive decision criteria and for the estimation of subgroup-specific effects at the final analysis. The prespecified prior distribution allowed for dynamic pooling across locations, in which the magnitude of pooling would depend on the similarity of the observed treatment effect between the respective subgroup locations (Figure 2). Using interim enrichment decision criteria based on the Bayesian model, the study stopped enrolling to basal ganglia hemorrhage and enriched to lobar hemorrhage at the second interim analysis. On study completion, the observed treatment effects were different across locations (mean difference of modified Rankin Scale score at 6 months equal to 0.142 in lobar and −0.041 in basal ganglia hemorrhage.[26] Thus, the

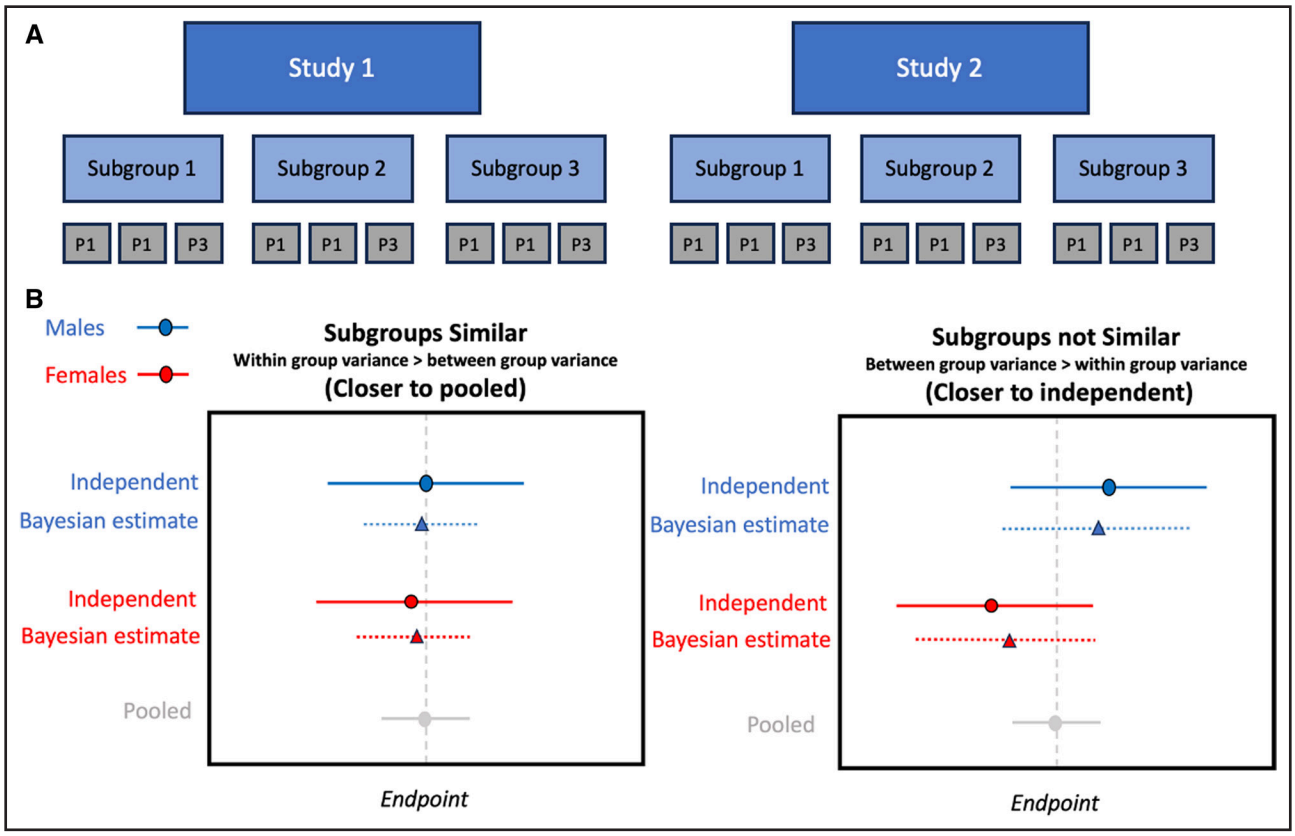

Figure 2. Overview of Bayesian hierarchical modeling and shrinkage estimation.
Schematic overview of Bayesian hierarchical models (**A**) and 2 shrinkage estimation scenarios (**B**). **A**, Bayesian hierarchical model in which parameter variability is partitioned into 3 sources (from top to bottom): the study level, the subgroup level, and the patient (P) level. Shrinkage of the subgroup results of a trial is determined by the ratio of between-subgroup variability and within-subgroup variability. In (**B**), treatment effect is plotted separately for 2 subgroups of a clinical trial, namely, women (red) and men (blue). On the left-hand side, the observed means are similar across subgroups, which implies that the Bayesian hierarchical model has a greater magnitude of partial pooling across the 2 subgroups (greater shrinkage). The Bayesian subgroup–specific estimates are pulled toward the other group, with increased precision resulting in more narrow CIs. Hence, the Bayesian subgroup–specific estimates are closer to the pooled estimates than the independent estimates. On the right-hand side, the observed means are different across subgroups, which implies that the Bayesian hierarchical model has a smaller magnitude of partial pooling across the 2 subgroups (less shrinkage). The Bayesian subgroup–specific estimates are pulled slightly toward the other group with little change in the width of the CIs. Hence, the Bayesian subgroup–specific estimates are closer to the independent estimates than the pooled estimate. In other words, if within-subgroup variability is larger than between-subgroup variability (right), shrinkage estimation will move the subgroup estimates closer to the trial's overall effect size estimate. If between-subgroup variability is larger than within-subgroup variability (left), less weight is given to the trial's overall treatment effect size estimate, and only a little shrinkage of the subgroup estimates toward the overall trial result estimate occurs.

Bayesian model incorporated moderately weak pooling (shrinkage) across the 2 groups, with estimated differences of 0.127 and −0.013 for lobar and basal ganglia hemorrhages, respectively. The trial demonstrated the superiority of the surgical approach in lobar hemorrhage only, with an overall Bayesian probability of benefit for the surgical approach equal to 0.981.

SYNERGY OF BAYESIAN STATISTICS AND ADAPTIVE TRIAL DESIGN

Adaptive trial designs use the evidence that accumulates in the trial to decide how to modify certain aspects of the trial design such as inclusion criteria, randomization algorithm, and sample size based on prespecified rules. Adaptive trial design can apply to both classical frequentist and Bayesian statistical approaches and is a highly complex topic in itself (see previous review article).[27] Advantages of adaptive trial designs include the potential to reduce sample size and drop arms or dosages. Of note, the use of Bayesian adaptive trial designs does not necessarily translate to smaller sample sizes; rather, the objective for many of these designs is to maximize the probability of a successful trial over the inherent uncertainties in the trial; which, sometimes, implies a larger sample size, for example, when a Bayesian interim analysis recognizes that more patients are required to provide sufficient statistical power.

Due to its iterative nature, Bayesian analysis methods lend themselves well to adaptive trial designs, and it is generally easier to combine an adaptive trial design with Bayesian statistics rather than with frequentist statistics although both combinations exist. With Bayesian

methods, a prior distribution can be updated to obtain a posterior distribution at a given interim analysis, which then serves as the prior for the next interim or final analysis.[28] Bayesian methods can be leveraged to define prespecified adaptive trial designs that aim to increase the overall probability of trial success by identifying key parameters of uncertainty in the trial design stage and building a trial design that adaptively responds to the accruing data and information on those key parameters. In addition, Bayesian modeling allows for more interpretable decision rules for adaptations, for example, futility criteria based on a clinically meaningful effect or the use of predictive probabilities.[29]

Examples of adaptive Bayesian stroke trials include the abovementioned ENRICH trial, the DAWN trial (Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo),[16] the BEST-II trial (Blood Pressure After Endovascular Stroke Therapy-II),[30] and the StrokeNet thrombectomy endovascular platform, which is currently enrolling patients.

Finally, Bayesian adaptive trial designs are well-suited for integrated phase 2/3 trial designs. For instance, the BEST-II trial used a Bayesian adaptive design to investigate the benefit of intensive versus standard blood pressure lowering after successful EVT.[31] The phase 2 trial used an uninformative prior, and it was prespecified which criteria need to be met for any active arm to advance to phase 3 and which size the phase 3 trial will have. None of the arms met these prespecified criteria, but had they been met, data-driven priors from phase 2 for phase 3 would have been available.

BAYESIAN STRATEGIES FOR MISSING DATA

Many studies experience a nontrivial amount of missing data; this often complicates the interpretation of analysis results. Many strategies exist for frequentist analyses of missing data both in the setting of frequentist and Bayesian trial designs. These strategies include single imputation/last observation carried forward and multiple imputation in which several complete data sets are imputed, and in a frequentist setting, a single *P* value is derived from the imputed data sets.[32]

Bayesian models can also be used for data-informed imputation. They harness additional variables collected in the study (at baseline and during follow-up) that may be correlated with the end point (Bayesian multiple imputations). The correlation of these additional variables with the end point is modeled to a large number of complete data sets from which a single Bayesian posterior distribution can be obtained, which reflects the uncertainty imposed by the missing data.[33] For example, the ENRICH trial prospectively defined a primary analysis that leveraged 90-day modified Rankin Scale score outcomes for Bayesian multiple imputation of missing 180-day modified Rankin Scale score outcomes.[25] This approach captures the uncertainty of the missing data in the derived Bayesian posterior distribution, whereby the magnitude of uncertainty in the missing data is determined by the observed data: if the observed correlation between the end point and additional variables is large, there is less uncertainty in the Bayesian posterior. However, if the observed correlation is small, there is greater uncertainty in the Bayesian posterior.

BAYESIAN TRIAL DESIGN FOR TRIALS SEEKING REGULATORY APPROVAL

In their guidance for the use of Bayesian statistics in medical device trials, the US Food and Drug Administration (FDA) acknowledges that the Bayesian approach may, in some instances, be less burdensome than a frequentist approach, and reliable prior information can justify a smaller sample size or shorter trial duration of pilot trials.[34] What constitutes reliable information is, of course, subjective, and whether a proposed Bayesian study design is fit-for-purpose is determined on a case-by-case basis by the agency. Therefore, the FDA recommends discussing any Bayesian trial protocol, particularly the choice of the priors, with their expert team early on at the design stage of the trial.[34] Because Bayesian analyses often lack closed-form solutions for statistical power calculations, FDA guidance documents also emphasize the importance of virtual trial simulations to understand the frequentist operating characteristics of a Bayesian design, for example, power and type I error.

Importantly, regulatory agencies do not commit to a single, universal criterion for study success in the setting of Bayesian trials. In other words, there is no clear rule on which posterior probability of a treatment benefit would be considered high enough in a Bayesian trial for a certain drug or treatment to be accepted by the FDA, the European Medicines Agency, and other regular regulatory bodies because the clinical context (eg, disease severity and prevalence, the risk profile of the treatment) is taken into account when determining what constitutes an acceptable criterion for study success. While this may seem confusing to investigators, the FDA seems to make great efforts to establish Bayesian methods in drug and device development: Bayesian methods, and particularly the agency's perspective on their usefulness in different settings, have been discussed in several recent publications and guidance for industry documents.[23,34–36] These efforts, together with the increasing awareness and familiarity of researchers with Bayesian methodology, will likely make the path to drug/device approval using Bayesian trials more clear in the future.

CONCLUSIONS

Stroke research, like all clinical research, must interpret the results of any study in the context of the existing evidence. While this is something that we all naturally do, it is not formalized in frequentist analyses. The iterative nature of Bayesian analysis allows researchers to make probabilistic statements about key analysis parameters, formally incorporate existing evidence into their analysis, pool/borrow information across trial subgroups to obtain more precise subgroup estimates, and, paired with adaptive trial designs, allow for population enrichment according to prespecified rules. Bayesian methods are becoming more popular among stroke researchers, and the potential of Bayesian trial design and data analysis has been recognized by guideline committees and regulatory authorities such as the US FDA.[34]

Due to the subjectivity that is inherent to the choice and weighing of prior information, it is crucial that we as researchers systematically define and justify our choice and weighing of prior information and describe these processes a priori in the study protocol. Efforts should be made to (1) implement standardized evaluation criteria for studies using Bayesian approaches and (2) standardize methodologies for synthesizing existing evidence into prior distributions.

If used appropriately, Bayesian approaches to trial design and statistics have the potential to increase the efficiency of clinical trials, reduce the number of patients that are exposed to harmful treatments, and allow us to adopt effective treatments earlier.

ARTICLE INFORMATION

Affiliations

Department of Diagnostic Imaging (J.M.O., S.B.C., M.D.H., M.G.), Department of Clinical Neurosciences, Cumming School of Medicine, Hotchkiss Brain Institute (J.M.O., J.K.H., A.G., S.B.C., B.M., M.D.H., M.G.), Department of Emergency Medicine (J.K.H.), and Department of Community Health Sciences (J.K.H.), Foothills Medical Center, University of Calgary, AB, Canada. Bright Research LLC, Chicago, IL (S.B.). Department of Neurology, Amsterdam University Medical Centers, the Netherlands (L.R.). Adaptix Trials, LLC, Austin, TX (B.R.S.). Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN (B.R.S.).

Acknowledgments

The authors would like to thank Jeffrey L. Saver.

Sources of Funding

None.

Disclosures

Dr Ospel is a consultant for Nicolab. Dr Goyal is a consultant for Mentice, Medtronic, MicroVention, and Stryker. Dr Brown reports compensation from MicroVention, Inc, for consultant services and compensation from the Faculty of Arts, University of Calgary, for consultant services. Dr Rinkel reports grants from the Niels Stensen Fellowship. Dr Ganesh reports compensation from Figure 1 for consultant services; grants from the Heart and Stroke Foundation of Canada; compensation from Atheneum for consultant services; grants from the Canadian Cardiovascular Society; compensation from Creative Research Designs for consultant services; grants from Campus Alberta Neuroscience; grants from the Government of Canada; stock options in TheRounds.com; compensation from DeepBench for consultant services; grants from the Alzheimer Society of Canada; grants from Rhodes Scholarships; grants from the Canadian Institutes of Health Research; compensation from the Canadian Association of Neuroscience Nurses for consultant services; grants from Alberta Innovates; stock holdings in Collavidence, Inc; grants from Wellcome Trust; travel support from the American Academy of Neurology; travel support from the American Heart Association; travel support from the University of Calgary; service as an editorial board member for the journal *Stroke* for American Heart Association; stock options in SnapDx, Inc; grants from MicroVention, Inc; service as a review editor for *Stroke* for *Frontiers in Neurology*; compensation from Alexion Pharmaceuticals for other services; grants from Panmure House; compensation from Biogen for other services; compensation from MD Analytics for consultant services; compensation from MyMedicalPanel for consultant services; compensation from CTC Communications Corporation for consultant services; service as an editorial board member for the journals *Neurology and Neurology: Clinical Practice* for the American Academy of Neurology; grants from the Fondation Brain Canada; and grants from the Sunnybrook Research Institute. Dr Hill reports grants from Boehringer Ingelheim, Medtronic, MicroVention, Inc, the Canadian Institutes of Health Research, NoNO, Inc, and Medtronic; compensation from Merck for end point review committee services; grants from Stryker Corporation; a patent issued for Systems and Methods for Assisting in Decision-Making and Triaging for Acute Stroke Patients licensed to Circle NVI; compensation from Brainsgate Ltd for consultant services; grants from Biogen, Inc, and Medtronic; and employment by the University of Calgary. The other authors report no conflicts.

Supplemental Material

Supplemental Article Details, Source Code, and Model Specifications
Appendix

REFERENCES

1. Gelman A CJ, Stern H, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3 ed. Chapman and Hall/CRC; 2013.
2. Lewis RJ, Viele K. Revisiting the analogy between clinical trials and diagnostic tests by interpreting a negative trial as a negative test for efficacy. *JAMA*. 2023;329:2023–2025. doi: 10.1001/jama.2023.8972
3. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485. doi: 10.1136/bmj.311.7003.485
4. O'Brien SF, Yi QL. How do I interpret a confidence interval? *Transfusion*. 2016;56:1680–1683. doi: 10.1111/trf.13635
5. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility index. *J Clin Epidemiol*. 2014;67:622–628. doi: 10.1016/j.jclinepi.2013.10.019
6. Gelman A, Stern H. The difference between "significant" and "not significant" is not itself statistically significant. *Am Stat*. 2006;60:328–331. doi: 10.1198/000313006x152649
7. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*. 2001;10:277–303. doi: 10.1177/096228020101000404
8. Bolstad WM, Curran JM. *Introduction to Bayesian Statistics*. Wiley; 2016.
9. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med*. 1999;130:1005–1013. doi: 10.7326/0003-4819-130-12-199906150-00019
10. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. *J Clin Epidemiol*. 2010;63:355–369. doi: 10.1016/j.jclinepi.2009.06.003
11. Rietbergen C, Groenwold RHH, Hoijtink HJA, Moons KGM, Klugkist I. Expert elicitation of study weights for Bayesian analysis and meta-analysis. *J Mixed Methods Res*. 2014;10:168–181. doi: 10.1177/1558689814553850
12. Ibrahim JG, Chen MH. Power prior distributions for regression models. *Stat Sci*. 2000;15:46–60. doi: 10.1214/ss/1009212673
13. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. 2014;13:41–54. doi: 10.1002/pst.1589
14. Holcomb JB, Tilley BC, Baraniuk S, Fox EE, Wade CE, Podbielski JM, del Junco DJ, Brasel KJ, Bulger EM, Callcut RA, et al; PROPPR Study Group. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA*. 2015;313:471–482. doi: 10.1001/jama.2015.12
15. Lammers D, Richman J, Holcomb JB, Jansen JO. Use of Bayesian statistics to reanalyze data from the pragmatic randomized optimal platelet and plasma ratios trial. *JAMA Netw Open*. 2023;6:e230421–e230421. doi: 10.1001/jamanetworkopen.2023.0421

16. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, Yavagal DR, Ribo M, Cognard C, Hanel RA, et al; DAWN Trial Investigators. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2017;378:11–21. doi: 10.1056/NEJMoa1706442
17. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, Roy D, Jovin TG, Willinsky RA, Sapkota BL, et al; ESCAPE Trial Investigators. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. 2015;372:1019–1030. doi: 10.1056/NEJMoa1414905
18. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, Schonewille WJ, Vos JA, Nederkoorn PJ, Wermer MJ, et al; MR CLEAN Investigators. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med*. 2015;372:11–20. doi: 10.1056/NEJMoa1411587
19. Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, Yan B, Dowling RJ, Parsons MW, Oxley TJ, et al; EXTEND-IA Investigators. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015;372:1009–1018. doi: 10.1056/NEJMoa1414792
20. Zampieri FG, Casey JD, Shankar-Hari M, Harrell FE Jr, Harhay MO. Using Bayesian methods to augment the interpretation of critical care trials. An overview of theory and example reanalysis of the alveolar recruitment for acute respiratory distress syndrome trial. *Am J Respir Crit Care Med*. 2021;203:543–552. doi: 10.1164/rccm.202006-2381CP
21. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365:176–186. doi: 10.1016/s0140-6736(05)17709-5
22. Alosh M, Fritsch K, Huque M, Mahjoob K, Pennello G, Rothmann M, Russek-Cohen E, Smith F, Wilson S, Yue L. Statistical considerations on subgroup analysis in clinical trials. *Stat Biopharm Res*. 2015;7:286–303. doi: 10.1080/19466315.2015.1077726
23. FDA. Impact story: using innovative statistical approaches to provide the most reliable treatment outcomes information to patients and clinicians. 2019. Accessed May 2024. https://www.fda.gov/drugs/regulatory-science-action/impact-story-using-innovative-statistical-approaches-provide-most-reliable-treatment-outcomes
24. Lipsky AM, Gausche-Hill M, Vienna M, Lewis RJ. The importance of "shrinkage" in subgroup analyses. *Ann Emerg Med*. 2010;55:544–552.e3. doi: 10.1016/j.annemergmed.2010.01.002
25. Ratcliff JJ, Hall AJ, Porto E, Saville BR, Lewis RJ, Allen JW, Frankel M, Wright DW, Barrow DL, Pradilla G. Early minimally invasive removal of intracerebral hemorrhage (ENRICH): study protocol for a multi-centered two-arm randomized adaptive trial. *Front Neurol*. 2023;14:1126958. doi: 10.3389/fneur.2023.1126958
26. Pradilla G, Ratcliff JJ, Hall AJ, Saville BR, Allen JW, Paulon G, McGlothlin A, Lewis RJ, Fitzgerald M, Caveney AF, et al; ENRICH trial investigators; ENRICH Trial Investigators. Trial of early minimally invasive removal of intracerebral hemorrhage. *N Engl J Med*. 2024;390:1277–1289. doi: 10.1056/NEJMoa2308440
27. Burnett T, Mozgunov P, Pallmann P, Villar SS, Wheeler GM, Jaki T. Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. *BMC Med*. 2020;18:352. doi: 10.1186/s12916-020-01808-2
28. Ruberg SJ, Beckers F, Hemmings R, Honig P, Irony T, LaVange L, Lieberman G, Mayne J, Moscicki R. Application of Bayesian approaches in drug development: starting a virtuous cycle. *Nat Rev Drug Discovery*. 2023;22:235–250. doi: 10.1038/s41573-023-00638-0
29. Saville BR, Connor JT, Ayers GD, Alvarez J. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clin Trials*. 2014;11:485–493. doi: 10.1177/1740774514531352
30. Mistry EA, Hart K, Davis T, Yeatts S, Lindsell CJ, Lewis RJ, Albers G, Wanderer JP, Prestigiacomo C, Bernard GR, et al. Design of the BEST-II randomized clinical trial. *Stroke*. 2022;2:e000249. doi: 10.1161/SVIN.121.000249
31. Mistry EA, Hart KW, Davis LT, Gao Y, Prestigiacomo CJ, Mittal S, Mehta T, LaFever H, Harker P, Wilson-Perez HE, et al. Blood pressure management after endovascular therapy for acute ischemic stroke: the BEST-II randomized clinical trial. *JAMA*. 2023;330:821–831. doi: 10.1001/jama.2023.14330
32. Harel O, Zhou XH. Multiple imputation: review of theory, implementation and software. *Stat Med*. 2007;26:3057–3077. doi: 10.1002/sim.2787
33. Little RJA, Rubin DB. Bayes and multiple imputation. In: *Statistical Analysis With Missing Data*. John Wiley & Sons, Inc; 2002:200–220.
34. Administration UFaD. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. In: US Department of Health and Human Services FaDA, ed; 2013. Accessed May 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials
35. FDA. Adaptive Designs for Clinical Trials of Drugs and Biologics. Accessed May 2024. https://www.fda.gov/media/78495/download
36. FDA. Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. 2020. Accessed May 2024. https://www.fda.gov/media/130897/download