Making Sense of a Negative Clinical Trial Result: A Bayesian Analysis of a Clinical Trial of Lorazepam and Diazepam for Pediatric Status Epilepticus

Daniel B. Chamberlain, BA; James M. Chamberlain, MD*

*Corresponding Author. E-mail: jchamber@cnmc.org.

Study objective: We demonstrate the application of a Bayesian approach to a recent negative clinical trial result. A Bayesian analysis of such a trial can provide a more useful interpretation of results and can incorporate previous evidence.

Methods: This was a secondary analysis of the efficacy and safety results of the Pediatric Seizure Study, a randomized clinical trial of lorazepam versus diazepam for pediatric status epilepticus. We included the published results from the only prospective pediatric study of status in a Bayesian hierarchic model, and we performed sensitivity analyses on the amount of pooling between studies. We evaluated 3 summary analyses for the results: superiority, noninferiority (margin $<-10\%$), and practical equivalence (within $\pm 10\%$).

Results: Consistent with the original study's classic analysis of study results, we did not demonstrate superiority of lorazepam over diazepam. There is a 95% probability that the true efficacy of lorazepam is in the range of 66% to 80%. For both the efficacy and safety outcomes, there was greater than 95% probability that lorazepam is noninferior to diazepam, and there was greater than 90% probability that the 2 medications are practically equivalent. The results were largely driven by the current study because of the sample sizes of our study (n=273) and the previous pediatric study (n=61).

Conclusion: Because Bayesian analysis estimates the probability of one or more hypotheses, such an approach can provide more useful information about the meaning of the results of a negative trial outcome. In the case of pediatric status epilepticus, it is highly likely that lorazepam is noninferior and practically equivalent to diazepam. [Ann Emerg Med. 2017;69:117-124.]

Please see page 118 for the Editor's Capsule Summary of this article.

A **feedback** survey is available with each research article published on the Web at www.annemergmed.com.
A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2016.08.449

INTRODUCTION

Background

The results of the randomized controlled trial "Lorazepam vs Diazepam for Pediatric Status Epilepticus" (the Pediatric Seizure Study) were published in 2014.[1] After enrolling 273 children with status epilepticus, the study found that both medications were approximately 72% effective at stopping convulsive status, and both were associated with an approximate risk of respiratory depression of 16% to 17%. To the clinician reading this article, the results suggest that the 2 medications have equivalent efficacy and safety. However, because this was designed as a superiority trial, the interpretation of these results by classic (frequentist) statisticians is that this trial failed to prove the hypothesis that lorazepam is superior to diazepam. The Food and Drug Administration (FDA) summarized the study results as having "…failed to establish the efficacy of Ativan in the treatment of status epilepticus"[2] because it did not prove superiority to the currently approved medication, diazepam. How is the clinician to reconcile what seems to be equivalence between 2 drugs with the statistical ruling that lorazepam has not been proven effective?

Importance

Bayesian analysis provides an alternative approach to the interpretation of data from clinical trials by directly answering the clinically relevant question, "Given the data we observed, how likely is our hypothesis?" The outputs of Bayesian analysis (posteriors) provide direct estimates of the effectiveness of both treatments and differences between

Editor's Capsule Summary

What is already known on this topic

Despite classic statistics' widespread use, many experts consider them an inappropriate method for analyzing clinical medical research. A classic analysis of a randomized trial "failed to demonstrate the efficacy of lorazepam over diazepam" in pediatric status epilepticus.

What question this study addressed

Although not superior, is lorazepam inferior, noninferior, or equivalent to diazepam in practice?

What this study adds to our knowledge

Using Bayesian analysis, the authors have shown that it is highly likely that lorazepam is noninferior and practically equivalent to diazepam.

How this is relevant to clinical practice

Bayesian analyses can provide clinicians more useful interpretations of a negative clinical trial result.

them.[3] Classic statistical analysis provides no such estimate of the probability of a hypothesis about efficacy or safety. The *P* value provided in classic statistical testing is often misinterpreted as the probability that the investigators are wrong in rejecting the null hypothesis, ie, the false-positive rate. The *P* value actually answers the question "Given the null hypothesis, how likely is it that we observed our data or more extreme data?" and does not measure the probability of the alternate hypothesis.

In addition to providing direct information about the likelihood of the study hypothesis, the Bayesian paradigm offers other advantages. First, Bayesian analyses can test multiple alternate hypotheses, whereas classic statistical analysis can test only one alternate to the null hypothesis. For example, Bayesian investigators can state both the probability that lorazepam is superior to diazepam and the probability that lorazepam is practically equivalent to diazepam, using the same data. This is because the outputs of Bayesian analyses are probability distributions of the likelihood of efficacy and safety. The results stand for themselves and do not depend on how the investigator framed the study hypothesis before collecting the data (as in classic statistical testing). Second, Bayesian analyses can include historical information from past studies directly in the analysis, allowing researchers to improve the accuracy of the estimates of efficacy and safety. To do this, researchers can design Bayesian models in a hierarchic fashion to allow pooling of information from different studies to create a composite estimate.[4,5] This is similar in concept to meta-analysis but offers the additional advantage that data can be incorporated dynamically as a study is being conducted, constantly updating the probability of the study hypotheses. A hierarchic model assumes that measured drug efficacy in one study is only an estimate of the true underlying drug effect. As more studies are conducted, the distribution of measured outcomes can be used to more accurately estimate the true efficacy of a drug. A hierarchic model is more appropriate than pooling of results of multiple studies because patients in different studies are not exactly exchangeable because of differences in study settings or patient populations. If studies are relatively similar, the amount of pooling between studies will increase and the uncertainty about the estimated effect will decrease.

Goals of This Investigation

In this article, we report the results of Bayesian analyses of the Pediatric Seizure Study. The purpose of these analyses is to provide a clinically meaningful interpretation of the results for the comparative efficacy of lorazepam and diazepam for pediatric status epilepticus in the context of a negative superiority trial result. This article also demonstrates the general principles of Bayesian analysis of a clinical trial.

MATERIALS AND METHODS

Setting and Selection of Participants

Methods for the Pediatric Seizure Study have been published previously.[1] Briefly, the study was commissioned in response to the Best Pharmaceuticals for Children Act, which seeks to improve labeling of medications for use in children. Children aged 3 months to younger than 18 years and with status epilepticus were randomized at 11 sites to receive either lorazepam or diazepam. A second dose was administered at 5 minutes if status was still present. The primary efficacy outcome was cessation of status by 10 minutes, without recurrence by 30 minutes. The primary safety outcome was the need for assisted ventilation ("life-threatening respiratory depression") within 4 hours of study drug.

Sample size projections estimated a difference in efficacy of 17% favoring lorazepam, based on a large out-of-hospital trial in adults.[6] In accordance with this estimate, the study was designed with 90% power to detect a 20% difference and 80% power to detect a 17% difference with 120 subjects in each treatment arm. Planned interim analyses at 50% enrollment were presented to an independent data and safety monitoring board, who recommended increasing enrollments to 131 subjects in each arm.

The study was designed as a superiority trial by an expert consensus panel before the request for proposal by