Broglio et al1 present an insightful comparison of 2 clinical trial designs: (1) the actual study design that was implemented in the Stroke Hyperglycemia Insulin Network Effort (SHINE) trial2 and (2) an alternative nonimplemented design. This comparison was planned before the SHINE study started enrolling patients. Each design can be described as a sequence of decisions after preplanned interim analyses and a plan for final analyses and reporting. The statistical design specifies how the study adapts to the interim data, including actions such early stopping and variations of the randomization probabilities. Broglio and coauthors1 describe what would have happened if the study team had followed the nonimplemented design. They provide the sequence of decisions that would have been made by using the data that were available for interim analyses during the SHINE clinical trial. Their study precisely mimics the interim decisions that would have been made.

The implemented and nonimplemented study designs had markedly different plans to decide at interim analyses whether to stop or continue the enrollment. In particular, the nonimplemented design had more frequent interim analyses, using bayesian predictions that leverage the available information on early auxiliary outcomes and primary outcomes. If the predictive probability of a positive result at the end of the study becomes small, then it is recommended to stop the study for futility. On the other hand, if there is strong evidence of positive treatment effects, then enrollment is terminated, and later, when all primary outcomes become available, the results of the trial are reported. By contrast, the study design that was implemented had a plan with less frequent interim analyses based only on primary outcomes and involved data imputation.

We congratulate the authors for this prospectively planned comparison, which showed that the nonimplemented design would have stopped the study for futility after the randomization of 800 patients, whereas the implemented design produced a recommendation to stop for futility after 936 patients were randomized. This comparative study achieves several aims.

First, it provides an interpretable estimate of the strong influence of the choice of the study design among candidate options on the number of enrolled patients and more generally on pivotal operating characteristics. A prospectively planned comparison has relevant merits. Indeed, defining the candidate designs before the enrollment period of the SHINE study eliminates potential concerns of a circular logic, with ad hoc components of the designs, such as the number of interim analyses, tailored to the trial data.

Second, it illustrates that patient-level data from completed clinical trials are crucial for evaluating innovative study designs. The methodological strategy used by Broglio et al3 can be used in prospective or retrospective analyses of completed clinical trials to evaluate innovative aspects of other trial designs, such as the integration of pretreatment biomarkers and prognostic variables or the use of external data sets to make interim decisions during the study.3 These empirical evaluations are complementary to the use of simulation studies, which are frequent and useful for investigating innovative statistical designs. Typically, simulation studies present limitations related to the arbitrary choice of the set of scenarios and to underlying assumptions that might be violated during clinical studies. In other words, the toolbox to design future clinical trials includes evaluations based on patient-level data from completed trials in addition to analytic results, for example, for power calculations and simulation studies. The use of real-world data sets (eg, electronic health records) could further support the assessment of the operating characteristics of innovative trial designs.

Third, the article of Broglio et al1 suggests the exciting idea of competitions between study designs or groups of biostatisticians. Competitions in other disciplines, such as machine learning and...
structural biology, have provided strong incentives for developing novel methods.\textsuperscript{4,5} These competitions have led to the introduction of powerful prediction algorithms and the identification of successful strategies by comparison on blinded validation data sets. Similarly, the comparison of emerging ideas in clinical trial design could be achieved by prospective head-to-head evaluations in a competitive landscape. In view of the time and vast resources necessary to develop effective treatments, initiatives dedicated to the evaluation of innovative statistical approaches for clinical trials, using patient-level data or data summaries from comprehensive collections of clinical trials, could be a good investment and contribute to relevant improvements.

The use of data sets from completed clinical trials can be useful to investigate several other operating characteristics of complex designs beyond the sample size. This investigation can be performed throughout the drug development process, from early phase studies to confirmatory phase 3 trials, to assess properties, including power, the risk of false-positive findings, and the accuracy in identifying which patients benefit from the experimental treatments. There are a few other approaches to estimate pivotal operating characteristics and to evaluate innovative designs based on completed clinical studies that have been previously discussed.\textsuperscript{6} These approaches include subsampling schemes to mimic sequential decisions during a clinical trial and the use of simulation scenarios obtained by fitting statistical models to a completed clinical trial.

The primary result of the work by Broglio et al\textsuperscript{1} is a point estimate of the sample size reduction of the nonimplemented bayesian design relative to the actual study design. To support the design of future clinical trials in a specific context, such as patients with hyperglycemia and acute ischemic stroke (ie, the SHINE population), both point estimates of pivotal operating characteristics (eg, sample size or the number of adverse events) and metrics of uncertainty and variability in candidate designs are necessary and equally important. These metrics can capture variability, for example, of the sample size, if a specific study were repeated many times under identical conditions, enrolling the same patient population and testing the same treatments. This type of variability can potentially be assessed using randomization or resampling techniques. We believe this approach will help tackle the limitations of prospective head-to-head comparisons for trial designs with more complex forms of adaptation to primary outcomes.

An additional component of uncertainty exists in the design of clinical trials, stemming from variation among different studies, in a range of aspects, including the enrollment period and the treatments tested. Quantifying this type of uncertainty requires the development of disease-specific, comprehensive, and up-to-date data collections of completed clinical trials. Improved data-sharing efforts and large data collections of completed trials will lead to more robust empirical evaluations and a better comprehension of efficiencies and risks of innovative study designs. We believe the next key steps in the direction taken by Broglio et al\textsuperscript{1} will be the prospective evaluation of general principles for clinical trial design across studies of different treatments and furthering methods for inference and uncertainty quantification.

---

**ARTICLE INFORMATION**

**Published:** May 11, 2022. doi:10.1001/jamanetworkopen.2022.11620

**Open Access:** This is an open access article distributed under the terms of the CC-BY License. © 2022 Trippa L et al. JAMA Network Open.

**Corresponding Author:** Lorenzo Trippa, PhD, Dana-Farber Cancer Institute, 60 Longwood Ave, Brookline, MA 02446 (ltrippa@jimmy.harvard.edu).

**Author Affiliations:** Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Trippa); Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, United Kingdom (Bacallado).

**Conflict of Interest Disclosures:** Dr Trippa reported receiving a grant from the National Institutes of Health (1R01LM013352-01A1) outside the submitted work. No other disclosures were reported.
REFERENCES


