You May Have Worked on More Adaptive Designs Than You Think

Christopher S. Coffey, PhD

According to the clinicaltrials.gov (accessed March 29, 2014), there are >1500 open clinical trials in the field of stroke. During the design of a clinical trial, several important design decisions must be made. Although study success depends on their accuracy, there may be limited information to guide the decisions. Adaptive designs address this uncertainty by allowing a review of accumulating data during an ongoing trial, and modifying trial characteristics accordingly if the interim information suggests that some of the original decisions may not be valid. Correspondingly, adaptive designs have received a great deal of recent attention in the statistical, pharmaceutical, and regulatory fields. However, it is well known that implementing many of the proposed adaptations will require the clinical trials community to address several statistical, logistical, and operational hurdles. For that reason, although some recent stroke trials have used adaptive methods, it may appear on first glance that adaptive designs are not being widely used. Yet, nearly all current stroke trials use group sequential methodology (interim monitoring) to some extent. Many researchers might think that these group sequential methods are a separate concept from adaptive designs. However, there is a close connection between the 2. I give a brief description of each of these methods and explain how these 2 statistical approaches are related.

Group Sequential Designs

Having a periodic review of interim efficacy and safety data by an oversight group has become an integral part of modern clinical trials. For the purposes of this article, we will refer to these groups of individuals as a Data and Safety Monitoring Board (DSMB). However, other terms are also used to describe these groups (Data Monitoring Committees, Data Safety Monitoring Committees, etc). The decision to stop a trial early for efficacy (interim data suggest a clear difference between groups) or futility (interim data suggest no significant difference probable by end of study) is complex and requires a combination of statistical and clinical judgment. For example, stopping an efficacious trial too late may needlessly delay some patients receiving the better treatment. However, stopping an efficacious trial too early may not provide data convincing enough to persuade a change in practice or provide sufficient safety information. To minimize the role of subjective judgment, statistical methods have been developed that allow for valid interim analyses before the completion of the trial.

For assessing efficacy, it is well known that repeated testing at a particular α level under the null hypothesis (generally that there is no difference between the groups) inflates the probability of making a type I error, rejecting the null hypothesis when it is true (or finding a treatment difference when none actually exists), for the entire study as a whole. The solution to this problem is to compare each of the interim test statistics to adjusted critical values that allow the overall family of tests to maintain the desired level of significance. Different types of group sequential tests give rise to different stopping boundaries, based on the amount of type I error spent at each interim look. Pocock bounds use stopping boundaries with the same critical value at each interim look (ie, they spend the same amount of type I error at each interim look). A downside of these bounds is the fact that the final stopping boundary is well below the desired level of significance, a situation that might cause some confusion if the observed P value is less than the desired significance level but not below the adjusted stopping boundary. In other words, one could obtain a P value <0.05 but not declare statistical significance at the final look. Correspondingly, these boundaries are seldom used in practice. O’Brien-Fleming bounds use more conservative stopping boundaries at early stages. These bounds spend little α at the time of the interim looks and lead to boundary values at the final stage that are close to those from the fixed sample design, avoiding the problem noted above with the Pocock bounds. The classical Pocock and O’Brien-Fleming boundaries require a prespecified number of equally spaced looks. However, a DSMB may require more flexibility. Alternatively, one could specify an α spending function that determines the rate at which the overall type I error is to be spent during the trial. At each interim look, the type I error is partitioned according to this α spending function to derive the corresponding boundary values. Because the number of looks does not have to be prespecified nor equally spaced, an O-Brien-Fleming type α spending function has become the most common approach to monitoring efficacy in clinical trials.

Received April 14, 2014; final revision received December 9, 2014; accepted December 15, 2014.
From the Department of Biostatistics, University of Iowa, Iowa City.
Correspondence to Christopher S. Coffey, PhD, Department of Biostatistics, University of Iowa, 2400 University Capitol Centre, Iowa City, IA 52240-4034, E-mail christopher-coffey@uiowa.edu

© 2015 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.114.004288
Stochastic curtailment methods are generally used for assessing futility. With this approach, a trial should be stopped if one can predict the outcome of the trial with high probability given the current data at an interim stage. For example, if the interim data suggest that the trial is unlikely to be positive, strong consideration should be made to terminating the trial. The most common approach for assessing futility is the use of conditional power, the probability that the test statistic at the final stage will be rejected given the observed statistic and assuming the prespecified effect for future observations. Hence, if an unfavorable trend is observed at an interim analysis, the conditional power represents the probability that the unfavorable trend might be reversed by the end of the trial. If the conditional power is below some prespecified threshold, typically 10% to 20%, the trial may be stopped for futility. However, this approach has been criticized because computing conditional power under the originally assumed alternative when the observed effect is near the null value may falsely overstate the true power and subsequently make it less likely the trial will stop for futility. A predictive power approach alleviates this problem by computing a weighted average of the conditional power values of the posterior distribution of the treatment difference given the observed data. As with conditional power, predictive power can be used to define a formal futility stopping rule.

**Adaptive Designs**

As described above, adaptive designs address uncertainty surrounding design choices made during study planning by allowing a review of accumulating information during an ongoing trial. There are essentially an infinite number of adaptive design possibilities. Some of the more commonly proposed adaptive approaches include adaptive dose–response methods (such as the continual reassessment method), adaptive randomization, sample size re-estimation, enrichment designs, and adaptive seamless designs (which combine phases usually considered in separate studies into a single trial). However, the increased attention given to adaptive designs in the scientific literature has come with a good bit of confusion about similarities and differences between the various types of proposed adaptations. To address some of this confusion, an adaptive design working group, consisting of individuals from industry and academia, was started in 2001. The group was associated with Pharmaceutical Research and Manufacturers of America until 2011, when the group transferred to the Drug Information Association and became the Adaptive Design Scientific Working Group. In 2006, this group published the first formal definition of an adaptive design in the literature: “By adaptive design we refer to a clinical study design that uses accumulating data to modify aspects of the study as it continues, without undermining the validity and integrity of the trial.” This publication also specified that “…changes are made by design, and not on an ad hoc basis” and that adaptive designs are “…not a remedy for inadequate planning.”

Maintaining the integrity of the trial involves both scientific components (such as whether the trial will answer the question it was designed to address) and statistical components (control of type I and II error rates, unbiased estimates of treatment effect). Properly designed simulations are often needed to explore whether the proposed adaptations introduce bias into these statistical components. When these simulations suggest that the proposed adaptations may introduce bias, additional simulations may also be critical to assure regulatory bodies that proper adjustments have been implemented to correct for this bias. This approach reinforces the importance of the concept of adaptive by design, because the adaptation rules must be clearly specified in advance to properly define the required simulations (ie, if A happens, then B will occur). Although simulations can be conducted for unplanned adaptations implemented after data have been observed, one cannot adequately capture the randomness that occurred before the decision to implement the unplanned adaptation. Therefore, only planned adaptations that have been adequately assessed in a rigorous simulation study can be guaranteed to avoid bias.

In 2010, the Food and Drug Administration (FDA) released “Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics.” This draft guidance document also included a definition for an adaptive design that was similar to that of the Adaptive Design Scientific Working Group: “…a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study.” Thus, both the Adaptive Design Scientific Working Group and the FDA support the notion that changes are based on prespecified decision rules. However, FDA defines this more generally: “The term prospective here means that the adaptation was planned (and details specified) before data were examined in an unblinded manner by any personnel involved in planning the revision….

This can include plans that are introduced or made final after the study has started if the blinded state of the personnel involved is unequivocally maintained when the modification plan is proposed.” During an ongoing trial, different individuals become unblinded to data at different time points and the FDA document left open some gray areas that merit further discussion. For instance, investigators typically remain blinded until the end of the study, whereas DSMB members may be partially or fully unblinded at the time of the first interim analysis. Suppose an investigator proposes a design change after the time of the first interim analysis based on external factors, such as the release of results from a similar trial, one could argue that the impetus for the proposed adaptation was not based on the results of unblinded data, which would fit the FDA definition for a valid adaptive design. However, if the proposed adaptation has to be reviewed and approved by the DSMB, the fact they have seen unblinded data would seem to imply that the definition may not be met. The role of a blinded versus unblinded statistician in the process may also be important in determining whether the definition has been met. Further clarification of these types of areas is needed in the future to ensure that researchers and regulatory authorities agree on what constitutes a valid adaptive design.

It is also important that researchers and patient communities fully understand what adaptive designs are not. The proper use of adaptation cannot in itself lead to an effective treatment but has the potential to increase the efficiency with which the correct answer is found. Interestingly, original interest in adaptive
designs is often driven by a desire to obtain positive results more quickly. However, the major benefit of adaptation seems to be the opposite, the ability to more quickly identify ineffective treatments. This is an important aspect of drug development because patients with stroke are a valuable resource. Stopping development of an ineffective treatment earlier in the process allows a redistribution of resources to more promising treatments.

**Group Sequential Designs Are Adaptive Designs**

As implied by the adaptive design definitions presented here, a group sequential design is an adaptive design that allows premature termination of a trial because of efficacy or futility, based on the results of an interim analysis. Hence, group sequential designs are some of the most commonly used adaptive designs in clinical trials. Accordingly, any stroke researcher that has worked on a recent clinical trial has likely been involved in an adaptive design, whether they recognized that at the time or not. Furthermore, widely used methods exist which allow these interim analyses to be conducted in a way that preserves the validity and integrity of the trial.

**Summary**

The path from having group sequential methods first proposed in the literature to having them widely used in clinical trials was a long, hard path that required implementing major changes to the overall infrastructure of the clinical trials community. For example, the implementation of these methods required the development of structure to support DSMBs, which are relatively standard for modern clinical trials. This also required substantial training of clinical trialists to ensure that they understand the intricacies of the methods, as well as the potential pitfalls associated with the use of the methods. There is clearly growing interest in extending stroke clinical trials to more complex types of adaptations. For many of the more complex, but potentially beneficial adaptations, the clinical trials community finds itself in a similar situation to the early days of group sequential methodology. To increase the usage of these more complex adaptations, similar types of infrastructure changes (more efficient data management, ability to implement complex simulation studies, ability to quickly respond to changes in drug distribution, etc) may need to take place.

**Sources of Funding**

Dr Coffey’s work was partially supported by National Institutes of Health U01 NS077352.

**Disclosures**

Dr Coffey is a consultant to ZZ Biotech, LLC.

**References**


**Key Words:** adaptive designs ■ clinical trials ■ clinical trials data monitoring committee ■ early termination of clinical trials ■ group sequential designs
You May Have Worked on More Adaptive Designs Than You Think
Christopher S. Coffey

Stroke. 2015;46:e26-e28; originally published online January 8, 2015;
doi: 10.1161/STROKEAHA.114.004288

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/46/2/e26

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

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

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