

Changing Endpoints after a Clinical Trial Begins

Submitted by Scott Evans, Ph.D., Harvard School of Public Health
May 27, 2006

It is well documented that pre-specification of endpoints is optimal for establishing and maintaining the scientific integrity of a clinical trial. Occasionally, information obtained after trial initiation (e.g., results from other trials) may lead trial leadership to consider adding endpoints during the course of a trial. Such additions would allow the incorporation of up-to-date knowledge into trial design. However, does the addition of such endpoints jeopardize trial integrity? Would trial results be credible?

Flexible and adaptive designs including amended study endpoints are common. Chan et.al.¹ compared published articles with protocols for 102 randomized trials and reported that 62% of the trials had at least one primary endpoint that was changed, introduced, or omitted. The prevalence of such endpoint revisions/additions/omissions in practice underscores the importance and magnitude of this issue and invites the evaluation of if and when the addition of endpoints is appropriate.

A key consideration in the evaluation of the ramifications of adding trial endpoints is whether the decision is independent of data obtained in the trial to date. If it can be shown that the decision to add endpoints is independent of trial data, then such additions have merit. In fact, Wittes² encourages consideration of changes in long-term trials as medical knowledge evolves or when assumptions made in design of the trial appear questionable. Wittes further argues that “the triallist may consider changes to the primary endpoint when the trial has airtight procedures to guarantee separation of the people involved in making such changes from data that could provide insight into treatment effect”.

Successful endpoint additions after trial initiation have been made. The Post-CABG trial³ was a randomized trial that compared two lipid-lowering regimens with respect to lipid deposition in the coronary arteries in patients that had coronary artery bypass surgery. A primary endpoint was not identified in the design of the trial. An angiogram to assess lipid deposition in the coronary arteries was conducted at entry and then again 5 years later. Changes over the 5 year interval were to be compared between regimens. Since endpoint results were not available for 5 years the protocol team used this time to define the endpoint and develop methods for analysis (and thus the endpoint selection was independent of trial data).

However, if the decision is not independent of the data, then the results associated with these endpoints have questionable validity. For example, if the endpoint has been identified or suggested based on examination of trial data, then there is a serious concern for selective inclusion or “cherry-picking”. Cherry-picking can inflate the Type I error rate since newly added endpoints may have been selected because they displayed a trend towards significance, while other trial data may have been examined but failed to display such a trend and thus were not selected and reported. In the Physicians Health study^{4,5}

the FDA did not approve an indication for aspirin in the prevention of myocardial infarction (MI) since MI was not the pre-specified endpoint.

Three important questions arise when evaluating if the addition of endpoints is independent of trial data. First, what is the source of the new information that causes consideration of the endpoint additions? If the source is external to the trial in question (e.g., results from another trial), then the addition of endpoints may be credible. Secondly, has the proposed endpoint (or related data) been reviewed by the individuals making the decision (e.g., by study investigators)? The addition of endpoints may be credible as long as there is no knowledge of the endpoint or related data (e.g., no interim analyses have been conducted) by the decision makers. Note however that even if no formal interim look at the data has been conducted, investigator impressions of the trial to date may influence decisions regarding endpoint additions (e.g., investigators may have a “sense” of the endpoint result or a related variable even though formal analysis of the endpoint has not been conducted). Thirdly, who is making the decision regarding endpoint addition (e.g., trial sponsors vs. a Data Monitoring Committee (DMC))? It is important that the decision to add endpoints be made by a non-conflicted and blinded body. Trial leadership may wish to convene an independent committee (e.g., a Data Monitoring Committee) that has not reviewed trial data (and has no “sense” of the trial data) to assess the potential impact on trial integrity and to make recommendations regarding the addition of endpoints. Note that if the DMC has reviewed interim data, then a DMC’s decision regarding the addition of trial endpoints may no longer be independent of trial data. In this case, a separate committee may be convened to evaluate the ramifications of endpoint additions.

Another consideration for the evaluating the addition of endpoints is the biological plausibility of the mechanism of action given current knowledge. As illustrated in this discussion, the addition of endpoints is a complex decision. Endpoint additions should be done selectively (e.g., based on sound biological theory).

However, one should also consider the ramifications of not adding the endpoints in question. Does the current state of knowledge (which may be different than that when the study was designed), render the results of the currently designed trial non-informative or inefficient? Is the currently designed trial now scientifically uninteresting or irrelevant? If so, then it may be prudent (and perhaps even ethically required) to add endpoints to ensure that the study is productive, constructive, and provides a scientific contribution.

One should be aware of potential operational bias induced by the addition of endpoints. The actions of clinical investigators and/or patients may be affected with knowledge of such trial revisions as they anticipate the reasons for such revisions. Note also that operational bias can also be induced with failure to add such endpoints.

If trial leadership decides to add endpoints, then the new endpoints should be specified in a revised protocol and analysis plan as soon as possible. Furthermore, responsible analysis and reporting is crucial. Importantly, analyses should be conducted on all

endpoints (i.e., original endpoints should be analyzed as originally planned). If endpoints are added as part of a composite endpoint, then the primary analysis should be conducted using the original pre-specified definition of the composite while analysis of the revised composite should be considered secondary. Selective reporting of endpoints has been documented¹, and can lead to misguided research and suboptimal patient care.

Furthermore, the analysis of the added endpoints should be accompanied by: (1) a very clear statement describing the fact that due to information obtained after trial initiation, these endpoints were added after trial initiation, (2) a description of the reasons (e.g., whether the endpoint was suggested by the data) and decision procedure (e.g., who made the decision and whether data were unblinded), (3) a discussion of the potential biases induced by the addition of the endpoints (including a discussion of data that were examined but were not added as endpoints), and (4) if warranted (i.e., if the decision to add endpoints was not independent of the data), a disclaimer that the results should be interpreted with caution and should be confirmed in future trials. Furthermore, to avoid overstating the significance of the results, the analysis should be descriptive, focusing on estimates of effect with corresponding confidence intervals. These intervals should be interpreted as providing plausible ranges of effect sizes that are consistent with the trial data. One should avoid using the intervals to assess statistical significance due to the potential inflation of Type I error created by the selective addition of endpoints.

It is important to note that a common abuse when pre-specifying an endpoint is vagueness or ambiguity in specification. For example, a protocol designed to study the effects of a new medicinal product on immune function may specify CD4 count as an endpoint. However, this is problematic. The endpoint can be interpreted as: (1) CD4 count at week 24, (2) changes from baseline in CD4 count at week 24, (3) the occurrence of a doubling of CD4 count from baseline, (4) the occurrence of at least a 50 cell increase in CD4 count from baseline, or any number of other functions of CD4. If this is not pre-specified, then there is no protection against examination of all of these combinations and then selecting and reporting the version of the endpoint that displays the most desirable result. This is another form of “cherry-picking” and inflates the type I error rate. The cherry-picking syndrome is common and leads to an under-reporting of negative evidence in addition to over-interpretation of positive evidence. To address this issue, editors of medical journals are discussing the potential requirement of submission of trial data and all analyses performed with manuscript submissions (Workshop on Assuring the Integrity in Reporting and Patient Safety in Therapeutic Trials, Department of Biostatistics, Harvard School of Public Health, Schering-Plough Workshop, June, 2006).

However valid statistical tests of an endpoint added after trial initiation can be obtained even if the decision to add the endpoint were based on data from the trial. If a trial is very large and of long duration, then one may base statistical testing regarding the endpoint in question only on data obtained after the decision to add the endpoint was made (e.g., only on subjects enrolled after the decision was made). This strategy essentially splits the trial into two phases: (1) a hypothesis generating phase in which endpoints are identified, and (2) a hypothesis testing phase. The two phases are separated

by the decision process. Valid statistical tests can be obtained using data only from the hypothesis testing phase.

It is very clear that scientific validity is optimized when endpoints are pre-specified. Clinical trialists should strive to pre-specify endpoints whenever possible. However, if important scientific knowledge has been gained since trial initiation, it seems prudent to incorporate this knowledge into the trial in a careful and responsible manner. Thus we should be open-minded and flexible to unique situations which may warrant the addition of endpoints. The decision to add endpoints should be well justified and a thorough evaluation of the effects on trial integrity should be conducted. Potential problems associated with at least some endpoint additions after trial initiation can be addressed through appropriate analysis and clear reporting.

References:

1. Chan, AW, Hrobjartsson, A, Haahr, MT, Gotzsche, PC, and Altman, DG, "Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials", *Journal of the American Medical Association*, 291:20:2457-2465, 2004.
http://www.splf.org/rmr/admin/JAMA2004_291_2457outcome.pdf
2. Wittes, J, "On Changing a Long-Term Trial Midstream", *Statistics in Medicine*, 27: 2789-2795, 2002.
<http://www3.interscience.wiley.com/cgi-bin/abstract/98517660/ABSTRACT>
3. The Post Artery Bypass Graft Trial Investigators, "The Effect of Aggressive Lowering of Low-Density Lipoprotein Cholesterol Levels and Low Dose Anticoagulation on Obstructive Changes in Saphenous-Vein Coronary Artery Bypass Grafts", *New England Journal of Medicine*, 336:153-162, 1997.
<http://content.nejm.org/cgi/content/abstract/336/3/153>
4. Young, F, Nightingale, S, Temple, R, "The Preliminary Report of the Findings of the Aspirin Component of the Ongoing Physicians Health Study. The FDA Perspective on Aspirin for the Primary Prevention of Myocardial Infarction", *Journal of the American Medical Association*, 259:3158-3160, 1988.
[Abstract not available on-line]
5. Cairns, J, Cohen, L, Colton, T, DeMets, D, Deykin, D, Friedman, L, Greenwald, P, Hutchison, G, Rosner, B, "Issues in the Early Termination of the Aspirin Component of the Physicians Health Study", *Annals of Epidemiology*, 1:395-405, 1991.
[Go to <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=PubMed> and enter title of the article)

Submitted by Scott Evans, Ph.D., Harvard School of Public Health