Issues in planning and conduct of sample size re-assessment in medical device trials

Hong (Laura) Lu, Ph.D, Xiting (Cindy) Yang, Ph.D, Jie (Jack) Zhou, Ph.D
Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration

Disclaimer: This presentation expresses personal views of the presenter and not necessarily that of the FDA.

Abstract

Sample size re-assessment is gaining popularity in clinical trials in recent years due to the practical reason and maturity of methodology. In the medical device industry, adaptive design is especially appealing as there is usually limited human data available before a pivotal study is designed for making assumptions on treatment effect. Although the methodologies on sample size re-assessment are well developed, additional issues in practice exist due to the complexity in planning and execution of adaptive trials. In this presentation, examples will be provided to address issues we see in sample size re-assessment in medical device trials. We will stress the importance of pre-planning, consistency between trial design, conduct and analysis, and early communication between industry and FDA.

1. Introduction

An adaptive design in a clinical trial allows modification of one or more aspects of the study during its course of conduct based on accumulating study data. Adaptive design has gain popularity in clinical trials in recent years due its flexibility. In medical device industry, adaptive design is especially appealing as there is usually limited human data available before a pivotal or decisive study is designed.

To provide guidance to industry on good clinical practice and to facilitate internal review process, Center of Device and Radiological Health (CDRH) published a draft guidance entitled "Adaptive Designs for Medical Device Clinical Studies" in May 2015. Recently, a survey was also conducted for all adaptive designs submitted to CDRH from January 2007 and May 2013. The survey results show that the most common type of adaptive design is sample size re-assessment at an interim analysis. The need for sample size re-estimation could be due to 1) uncertainty of treatment effect and/or variability due to reasons such as lack of clinical data and change of standard care, 2) avoidance of expensive up-front commitments of investment, and other reasons.

Although the methodologies on sample size re-assessment are well developed, additional issues in practice exist due to the complexity in planning and execution of adaptive trials. In Section 2, we will provide mock examples to illustrate the issued we experienced in medical device trials. In Section 3, we will summarize the findings.

2. Examples and Issues

In this section, we will provide four examples in sample size re-assessment. Although the designs and outcomes of the trial have been modified for propriety reason, the issues remain the same.

Example 1

The trial design is as following:

- A randomized, controlled, open-label superiority trial
- Primary effectiveness endpoint: treatment success through 3 months
- Sample size: initial sample size 80 subjects/arm was planned for achieve 80% power
- Sample size re-assessment at an interim analysis:
  - Planned at 50% information with 80 subjects evaluated at Month 3
  - Based on promising zone approach (Mehta and Pocock (2011)): if the conditional power is $\geq$ 80% (favorable zone), continue with original sample size; if the conditional power is $50\% \leq CP < 80\%$ (promising zone), sample size will be increased to achieve a conditional power of 80% with the maximum sample size caped at 300 total; if the conditional power is <50% (unfavorable zone), then the trail will continue with the initial planned sample size.
  - No Type I error rate adjustment is needed since the sample size is only to be increased when the interim results are promising (Chen, DeMets, Lan, 2004)

The trial was monitored by a Data Monitoring Committee (DMC) and the interim analysis was conducted by an independent statistician in a contract research organization. After reviewing the results of the interim analysis, the DMC concluded that there was no need to increase sample size. The trial continued as planned and the final results demonstrated effectiveness of the device.

The trial described above serves as a good example as it has the following desirable features regarding sample re-assessment:

- The Type I error rate was controlled since the sample size re-assessment was pre-planned in a statistical analysis plan and a valid statistical methods was used for sample size re-assessment.
- Operational bias was mitigated as the trial was monitored and advised by a DMC and the interim analysis was conducted by an independent statistician.
- The study results can be interpreted statistically and clinically as the trial was conducted according to the protocol.

Example 2

The trial design is as following:

- A randomized, controlled, open-label trial
- Primary endpoint: event free rate at 6 months

- Sample size: initially planned at 300 subjects/arm with 80% power for superiority claim
- The trial was originally designed as a superiority study (T vs. C), but was changed to a non-inferiority comparison at two years after enrollment started and within 6 months before study submission. No rationale was provided for the choice of the non-inferiority margin of 10%
- An interim analysis was planned with sample size re-assessment proposed within 6 months of study submission (the same time as the change from superiority to non-inferiority claim) as the following:
    - The analysis was to be conducted at 50% information (150 subjects/arm)
    - Li's (2002) approach was used for sample size re-assessment: if $p<0.005$ ($\alpha 1$), stop for effectiveness; if $p>0.25$, stop for futility; otherwise adjust sample size to reach conditional power 80%

Based on the interim results, the sponsor decided not to stop the trial for effectiveness or futility, but to increase the sample size to 440 in total. However, the final analysis was conducted with 580 in total and the final p-value for effectiveness was 0.016. Based on Li's approach, the nominal α level for the final result should be 0.040. However, the sponsor compared the final p-value with the α level of 0.048, which was based on O'Brien-Fleming (OF) spending function as if no sample size-reassessment was conducted at the interim analysis.

The trial described above has the following issues to be concerned with:

- The change from superiority to non-inferiority design was considered post-hoc since the changes was proposed near the end of trial when more than 50% of subjects were already evaluated.
- The interim analysis was also considered as post-hoc due to the same reasons as above.
- The proposed changes were not submitted to FDA.
- The trial was not conducted according to the decision at interim analysis: the sample size at final analysis was 580 instead of 440.
- The sample size re-assessment was not taken into consideration in the final analysis:
    - Sample size re-assessment was originally based on Li's approach
    - O'Brien-Fleming spending function was used at the final analysis as if no sample size-reassessment was ever conducted at the interim analysis.

Although the final results had a p-value that was smaller than the nominal α based in either Li's or OB approach, the strength of the result is difficult to interpret from statistical aspects due to the post-hoc nature of changes in trial design and inconsistency in trial conduct and analysis from the trial plan.

Example 3

The trial design is as following:

- A randomized (2:1), active controlled, non-inferiority trial
- Endpoint: success rate at 12 weeks
- Non-inferiority margin: 10% (expected success rate in both arms: 85%)
- Initial sample size: 450 (300: 150) subjects with 80% power
- Blinded sample size re-assessment planned at 150 (100:50) subjects

Blinded sample size re-assessment minimally impacts Type I error rate and adjustment for multiplicity is usually not required for the final alpha level. Therefore, the final chance of success won't be affected by blinded sample size re-assessment. However, the sponsor chose not to do the blinded sample size re-assessment due to concern on

- reliability of outcome from 150 subjects
- Type I error inflation (not an issue)
- potential budget

The final results did not demonstrate the non-inferiority of the testing device to the active control. The results were marginal with the observed rates close to 65% in two treatment arms, which are lower than the expected 85% success rate. Had the blinded sample size re-assessment been conducted as planned, interim results could provide valuable information for resizing the trial and increase the chance of trial success. Therefore, the lesson we learned from this example are as following:

- Blinded sample size re-assessment provides valuable information in sizing a study.
- There is no need to adjust final Type I error rate with blinded sample size re-assessment.
- Blinded sample size re-assessment should be encouraged when feasible.

Example 4

The trial design is as following:

- A randomized, controlled, double blind, multi-center, superiority study
- Primary endpoint: 2-year rate of SAE
- Analysis method for primary endpoint: log-rank test for comparing hazard rates
- Initial sample size: 220 subjects/arm with 85% power to demonstrate effectiveness (30% reduction of event rate)
- An interim analysis planned as the following:
    - Planned at 200 subjects/arm, α_1=0.019, α_f=0.021 (OF spending function)
    - Sample size re-assessment at interim: If p>0.019, re-assess sample size based on Cui, Hung and Wang (2009) method with maximum sample size caped at 600 subjects/arm

The trial described above has the following issues to be concerned with:

- The parameter in the statistical hypotheses (2-year rate of SAE) is different than that tested in log-rank test (hazard ratio for SAE)

- Information level at interim analysis was calculated incorrectly based on number of subjects at the interim analysis. For log-rank test, the information level at an interim analysis is determined by the number of events, not the number of subjects.
- The timing of interim analysis was near the end of study when 200 out of the 210 total subjects were evaluated. The utility of the late interim analysis is very limited due to:
  - Even if the results at the interim analysis are promising enough for stopping the trial, little resource can be saved since all 210 subjects are already enrolled.
  - If, based on the results of the interim analysis, the sample size is to be increased to achieve sufficient conditional power at this late stage, the observed treatment effect must be quite small compared with expected treatment effect (<15% reduction vs. 30% reduction in event rate). Whether the observed result at the interim analysis implies clinical significant effect is questionable.
  - If the sample size is to be increased, with a small observed effect and a small α to spend, the amount of increase must be large in order to achieve sufficient conditional power. However, the large number of patients increased will be substantially under-weighted with the CHW method, which leads to inefficiency from statistical point of view.

After reviewing the sponsor's protocol, FDA provided the following comments:

- Parameter tested should be the same as that specified in the statistical hypothesis.
- Information level for log-rank test at interim should be decided by number of events at interim.
- Earlier interim analysis is recommended if the intention is to save subjects with unexpected large effect.
- If the trial not stopped at the interim, the applicant may consider setting futility bound based on whether the observed outcome indicates clinically non-significant effect.

Upon receiving the FDA's comments, the sponsor had a few rounds of communication with FDA and made the following revisions in the final protocol:

- Hazard ratio was the primary parameter of interest.
- Interim analysis was set to the time of occurrence of 50% of planned total events.
- A futility bound was established to terminate study with unfavorable results at interim analysis.
- Sample size re-assessment would be based on conditional power and there would be no sample size increase when CP is below 50%.
- Sample size re-assessment method was changed to promising zone approach (Mehta (2008) without down-weighing the second stage results.

The lesson learned from this example is that early communication with FDA could lead to a more clinically meaningful and statistically more efficient design.

3. Summary and Conclusion

Based on the examples described above, the following are the points to be considered when designing a trial with sample size re-assessment:

- Properly planned sample size re-assessment can potentially save resource/increase the chance of success.
- Preplanning is essential for protecting the integrity and scientific validity of the study.
- Sample size re-assessment should be conducted by an independent statistician and under the advice of a DMC.
- Early communication with FDA is important for developing a valid and efficient plan for sample size re-assessment.
- Blinded sample size re-assessment based on aggregate rate and variance with no cost for Type I error rate is encouraged.
- Valid statistical method should be used in controlling Type I/Type II error rate with unblinded sample size re-assessment.
- Futility bound should be set to avoid spending further resource on device without clinical meaningful effect.

References

- Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics,*1999; **55**:853–7.
- Li G, et al. A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics,* 2002; **3***:* 277-87
- Chen J, DeMets DL, Lan KK. Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine,* 2004; **23**:1023-38.
- Mehta C, Pocock S. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in Medicine,* 2011; **30**: 3267–84.