

PP14: Sample Size Re-Estimation in Action: Design Consideration, Charter Development, and Implementation of Analyses in a Trial with Survival Endpoints

Adam Hamm, PhD

Abstract

We detail the operational and statistical considerations for a recent trial that utilized sample size re-estimation in a cardiovascular trial with a time to event endpoint. This case study details beneficial study practices and processes that could be improved, including development of the DMC charter to detail adaptive decision making, maintaining the trial masking of treatments, performing complex statistical analyses, and ultimately informing study personnel of decision made at the interim analysis. This case study is presented from the unblinded statistician's point of view, including how the unblinded statistician operates within the clinical team to ensure the best possible outcome for the sponsor at the end of the trial.

DMC Charter for a Study with an Adaptive Design

Charters used for trials with adaptive designs will sometimes include sections that are not standard for DMCs that are only associated with safety review. Especially for trials with survival endpoints, DMC meetings for efficacy analyses will be event based rather than enrollment based and this must be detailed in the charter. Adaptation rules, including stopping for futility for efficacy or dropping doses (in multiple active arm trials) should be included. An example from our case study is presented as follows:

At the interim analysis, the decision algorithm will be implemented sequentially in the following steps:

1. The conditional power (CP) for primary efficacy will be calculated. The CP is calculated with the initial planned total sample size
2. Futility will be assessed, based on the futility boundary ($CP < 10\%$). If the futility stopping boundary is reached, the DMC will consider all evidence in the interim report in order to make their recommendation.
3. If the trial is not stopped for futility then efficacy will be assessed applying the criteria of one-sided p -value < 0.0005 .
4. If the trial is not stopped for futility or efficacy, the calculated CP will be used to determine if the data are in:
 - The unfavorable zone;
 - The promising zone; or
 - The favorable zone

Specific rules for sample size re-estimation should be provided under access-controlled separate cover if complex (i.e. multiple options for the N for sample size increase) so that the magnitude of the effect size at the interim analysis cannot be assessed by a trained statistician.

The Statistics for the Interim Efficacy Analysis in a Study with a Time to Event Endpoint and Stratified Analyses

When developing the interim statistical plan for an adaptive design with sample size re-estimation (SSR), identifying the statistics used for the conditional power calculation is useful and crucial for those programming the conditional power. In this example, the statistics for use in the conditional power were not readily available from Proc Lifetest in SAS and therefore must be carefully considered and validated by statistical programming. Key statistics required for the interim analysis (stratified log-rank test) in our study were:

The stage 1 stratified log rank statistic $Z_1 = \frac{U_1}{\sqrt{V_1}}$ calculated based on log rank score U_1 with variance V_1 where

$$U_1 = \sum_{s=1}^S O_s - E_s$$

and $O_s = \sum_{j=1}^J O_{sj}$ is the total number of events in experimental group in stratum s ,

and $E_s = \sum_{j=1}^J E_{sj} = \sum_{j=1}^J \left(\frac{r_{Esj}}{r_{sj}} \right) d_{sj}$ is the total number of expected events in experimental group in stratum s .

Similarly, the variance of the score is given by the sum of the variances calculated separately within each stratum:

$$V_1 = \sum_{s=1}^S V_s$$

where $V_s = \sum_{j=1}^J \frac{r_{Csj} r_{Esj} d_{sj} (r_{sj} - d_{sj})}{r_{sj}^2 (r_{sj} - 1)}$

In the equations above, $t_{(1)} < \dots < t_{(J)}$ are the ordered event times within stratum s ; at time $t_{(j)}$ there are d_{sj} events in stratum s , and O_{sj} events in experimental group in stratum s ; and just before time $t_{(j)}$ there are $r_{sj} = r_{Csj} + r_{Esj}$ patients at risk in stratum s , with r_{Csj} in the control group and r_{Esj} in the experimental group.

In summary, the derivations of the key statistics U_1 and V_1 which are key in the calculation of the log rank statistics, interim hazard ratio and conditional power are a counting process where the number of subjects at risk each time at least one event is seen at time $t_{(j)}$ and the total number of events seen at that time are used in the equations for U_1 and V_1 as noted above.

Therefore the counting process can be summarized in the following steps:

1. Identify and order all event times across strata $t_{(1)} < \dots < t_{(J)}$
2. At each time of event (j) and within each strata (s) , tally the number or events d_{sj} and the number of subjects who were at risk in the control and experimental groups (r_{Csj} and r_{Esj} , with r_{sj} being the sum of these two values)
3. Calculate U_1 and V_1 per the formulas above.

The log hazard ratio can also be calculated from this as $\delta = \widehat{Z}_1 / \sqrt{\widehat{V}_1}$.

Calculation of the Conditional Power

If the counting process and derivations of Z_1 , U_1 and V_1 are evaluated correctly, the conditional power can be subsequently derived as

$$CP = P_{\delta}(Z_{CHW} > c_2 | Z_1) = P_{\delta}\left(Z_2 > \frac{c_2 - \sqrt{w} Z_1}{\sqrt{1-w}} \middle| Z_1\right) \\ \approx 1 - \Phi\left(\frac{c_2 - \sqrt{w} Z_1}{\sqrt{1-w}} - \frac{Z_1}{2\sqrt{V_1}} \sqrt{\sum_{i=1}^S \frac{N_{2i}}{N_{1i}} D_{1i}}\right)$$

Additional key statistics in the calculation of conditional power are:

c_2 - the critical value used in determining statistical significance at the second (final) stage

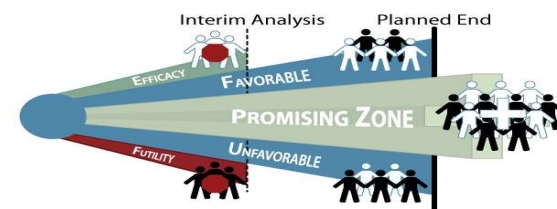
w - the PLANNED information at the time of interim analysis

N_{1i} and N_{2i} are the known N in the i^{th} stratum after stage 1 and the planned n in the i^{th} stratum after stage 2, respectively, in the absence of any adaptations.

D_{1i} is the number of events (total) in the i^{th} stratum.

The DMC Meeting and Evaluation of Interim Results

After evaluation of conditional power and consideration of safety data, the decisions for the DMC can be best summarized after the conditional power is placed into one of the "zones" shown in the following graphic.



In our case study, the upper boundary for futility is 10% conditional power. If the conditional power is less than 10%, the DMC could stop the study due to futility. If the p -value at the interim analysis is less than 0.0005, then the analysis is in the efficacy zone and the DMC could stop the study due to overwhelming efficacy. If the conditional power is in the range of 10-30% (unfavorable zone) or above 90% but not high enough to declare overwhelming efficacy ($p < 0.0005$) then the study would continue with no adaptation. If the conditional power is in the promising zone (30-90%) the sample size could be adjusted in accordance with rules provided to the DMC in a controlled access document.

Questions about this presentation? Please contact Adam Hamm at adam.hamm@Cytel.com