Endpoint selection and sample size recalculation with composite binary endpoints

Vienna-Barcelona

2020-12-04

Main ideas and motivation

Goal

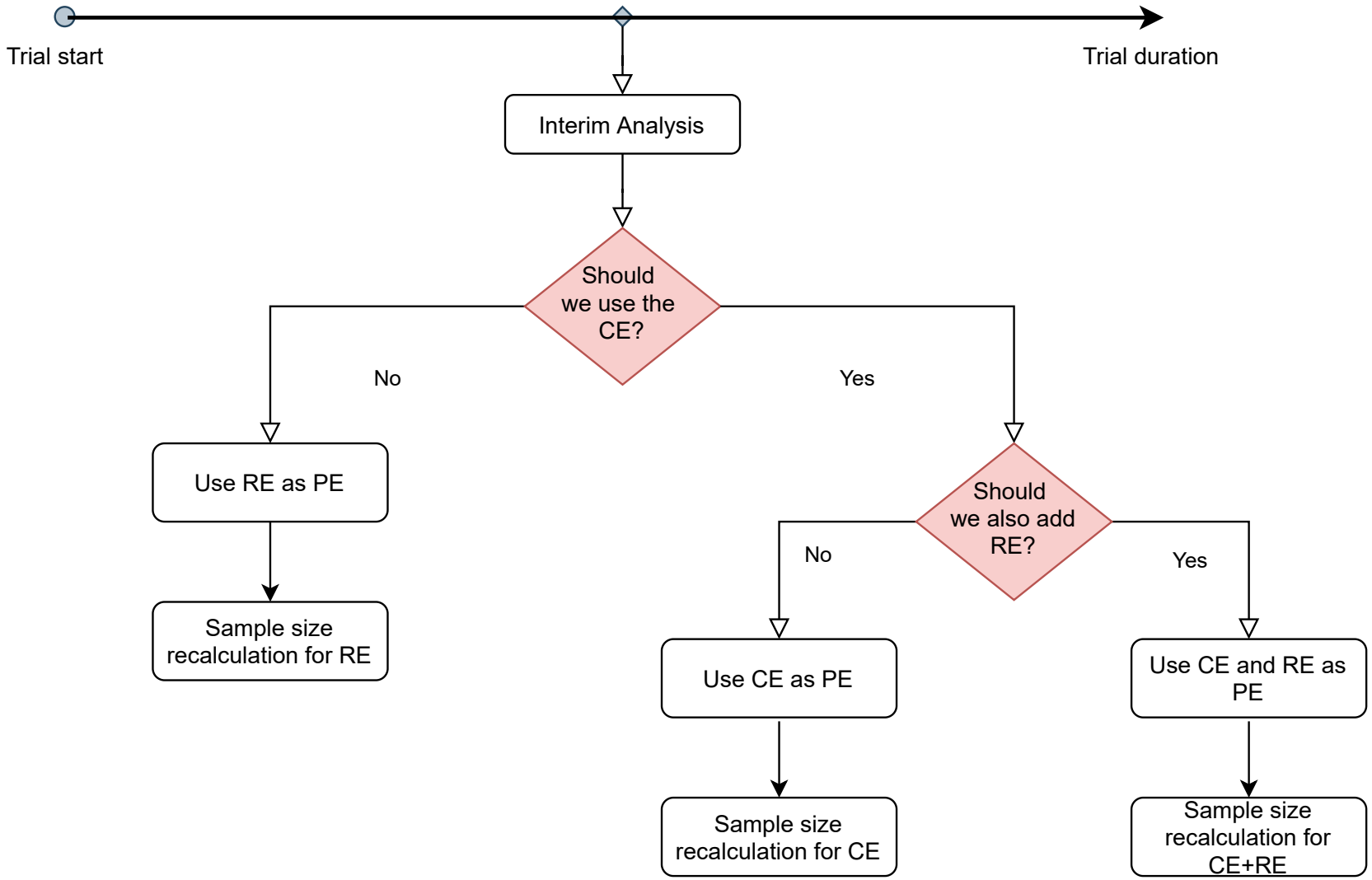
To define a theoretic-decision framework to add/drop components into a primary composite endpoint and to reassess the sample size accordingly.

- Assume composite binary endpoint (CE) formed by two components;
- Suppose that the Endpoint 1 (called relevant endpoint (RE)) is more relevant than Endpoint 2.

Trial design with an interim stage in which we:

1. select the most efficient endpoint to be used as the primary endpoint;
2. determine which testing procedure is more powerful to use;
3. recalculate the sample size in accordance to the aforesaid.

Scheme of the problem



Where are we? – Last meeting (October 29, 2020)

Trial design in which **at the end of the study** we evaluate whether to use the CE or the RE as primary endpoint;

Where are we? – Last meeting (October 29, 2020)

Trial design in which **at the end of the study** we evaluate whether to use the CE or the RE as primary endpoint;

What we did:

- Asymptotic Relative Efficiency (ARE) method: to decide between the composite endpoint and the relevant endpoint.
- Estimation of the correlation and the probabilities under the control group to calculate the ARE method.

Where are we? – Last meeting (October 29, 2020)

Trial design in which **at the end of the study** we evaluate whether to use the CE or the RE as primary endpoint;

What we did:

- Asymptotic Relative Efficiency (ARE) method: to decide between the composite endpoint and the relevant endpoint.
- Estimation of the correlation and the probabilities under the control group to calculate the ARE method.

Some conclusions and ideas:

- ARE method was not selecting the most powerful design when ARE≈ 1.
- Use the **ratio of sample sizes** as the decision criteria, and compare the results with the ones obtained using the ARE criteria.
- Implement the **sample size reassessment** and include an **interim stage**.
- Consider also the comparison in efficiency of the composite design versus the **multiple endpoint approach**.

Today's presentation

1) **Simulation results**: Evaluate whether to use the CE or the RE

Trial designs with:

(Potential) Interim stage:

- Decision based on blinded/unblinded data using 50% of the initial sample size
- Decision based on blinded/unblinded data using 100% initial sample size

(Potential) Sample size reassessment

- With sample size recalculation based on the decision.
- Without sample size recalculation after the decision.

Today's presentation

1) **Simulation results**: Evaluate whether to use the CE or the RE

Trial designs with:

(Potential) Interim stage:

- Decision based on blinded/unblinded data using 50% of the initial sample size
- Decision based on blinded/unblinded data using 100% initial sample size

(Potential) Sample size reassessment

- With sample size recalculation based on the decision.
- Without sample size recalculation after the decision.

2) **Sketch for comparison between multiple endpoints and composite endpoint approaches**

Simulations without sample size reassessment

Two-group comparison based on:

- the **relevant endpoint** $\rightarrow$ Endpoint 1 as the primary endpoint;
- the **composite endpoint** $\rightarrow$ Composite endpoint as the primary endpoint;
- the **selected endpoint** $\rightarrow$ Select the primary endpoint based on the decision criteria and using blinded/unblinded data.

Simulations without sample size reassessment

Two-group comparison based on:

- the **relevant endpoint** $\rightarrow$ Endpoint 1 as the primary endpoint;
- the **composite endpoint** $\rightarrow$ Composite endpoint as the primary endpoint;
- the **selected endpoint** $\rightarrow$ Select the primary endpoint based on the decision criteria and using blinded/unblinded data.

Decision criteria

To choose which primary endpoint(s) to use, we consider:

- ARE method: comparison of the powers under sequences of local alternatives.
- Ratio of the sample sizes: comparison of the required sample sizes.

Simulations without sample size reassessment

Two-group comparison based on:

- the **relevant endpoint** $\rightarrow$ Endpoint 1 as the primary endpoint;
- the **composite endpoint** $\rightarrow$ Composite endpoint as the primary endpoint;
- the **selected endpoint** $\rightarrow$ Select the primary endpoint based on the decision criteria and using blinded/unblinded data.

Decision criteria

To choose which primary endpoint(s) to use, we consider:

- ARE method: comparison of the powers under sequences of local alternatives.
- Ratio of the sample sizes: comparison of the required sample sizes.

We simulate 100000 trials for each scenario, for each decision criteria and for each endpoint considered as primary endpoint.

The sample size is calculated to have 0.80 power to detect an effect of OR_1 on the endpoint 1 (**RE**) at significance level $\alpha = 0.05$.

Simulations with sample size reassessment

Two-group comparison based on:

- the **relevant endpoint** $\rightarrow$ Endpoint 1 as the primary endpoint;
- the **composite endpoint** $\rightarrow$ Composite endpoint as the primary endpoint;
- the **selected endpoint** $\rightarrow$ Select the primary endpoint based on the decision criteria and using blinded/unblinded data.

Decision criteria

- ARE method: comparison of the powers under sequences of local alternatives.
- Ratio of the sample sizes: comparison of the required sample sizes.

We simulate 100000 trials for each scenario, for each decision criteria and for each endpoint considered as primary endpoint.

Simulations with sample size reassessment

Two-group comparison based on:

- the **relevant endpoint** $\rightarrow$ Endpoint 1 as the primary endpoint;
- the **composite endpoint** $\rightarrow$ Composite endpoint as the primary endpoint;
- the **selected endpoint** $\rightarrow$ Select the primary endpoint based on the decision criteria and using blinded/unblinded data.

Decision criteria

- ARE method: comparison of the powers under sequences of local alternatives.
- Ratio of the sample sizes: comparison of the required sample sizes.

We simulate 100000 trials for each scenario, for each decision criteria and for each endpoint considered as primary endpoint.

The sample size is calculated to have 0.80 power to detect an effect of OR_* on the **CE** at significance level $\alpha = 0.05$. OR_* is computed based on the parameters of the composite components and assuming correlation equal 0.

Conclusions simulation study

Conclusions:

- The powers using the selected endpoint (both blinded and unblinded approaches) are larger than the ones obtained using the composite endpoint and the relevant endpoint.
- In general, it works better the blinded approach.
- The significance level is slightly larger than $\alpha = 0.05$ when using the unblinded approach.
- The power in the unblinded approach is in general larger than $1 - \beta = 0.80$ when recalculating the sample size.
- The decision based on the ARE and ratio of the sample sizes is generally the same, with the exception of some cases where ARE≈ 0.95 and SS≈ 1.

Comparison multiple endpoint vs composite endpoint

We want to compare the efficiency of two designs:

- the **composite endpoint** $\rightarrow$ CE as the primary endpoint;
- the **multiple endpoints** $\rightarrow$ CE and the RE as multiple primary endpoints.

[1] Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W., & Rohmeyer, K. (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. BiomJ.

Comparison multiple endpoint vs composite endpoint

We want to compare the efficiency of two designs:

- the **composite endpoint** $\rightarrow$ CE as the primary endpoint;
- the **multiple endpoints** $\rightarrow$ CE and the RE as multiple primary endpoints.

Sketch algorithm for comparing the efficiency of multiple endpoints vs composite endpoints:

- Simulation-based approach to decide between multiple endpoints and composite endpoints:
  - Estimate the correlation and probabilities under control group
  - Calculate the critical values for the adjusted Bonferroni[1]
  - Estimate the power for both designs
- Sample size recalculation for the multiple endpoints $\rightarrow$??

[1] Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W., & Rohmeyer, K. (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. BiomJ.

Structure of the paper and R package

Proposed structure

1. Notation
2. Family hypotheses and test statistics
3. Decision criteria
4. Endpoint selection and sample size reassessment
5. `eselect` package in R (R pkg, Vignette and Examples)
6. Simulations
7. Discussion

To think about: Journal?

To do list

- Decision criteria: ARE vs Ratio sample sizes. Review references Ryan et al. [2],[3] and Cuzick [4].
- Try different ARE approximations.
- Implement sample size recalculation for multiple binary endpoints and run simulations.

[2] Lefkopoulou, M., & Ryan, L. (1993). Global Tests for Multiple Binary Outcomes. Biometrics.

[3] Legler, J. M., Lefkopoulou, M., & Ryan, L. (1995). Efficiency and Power of Tests for Multiple Binary Outcomes. JASA.

[4] Cuzick, J. (1982). The Efficiency of the Proportions Test and the Logrank Test for Censored Survival Data. Biometrics.