PASSION. SCIENCE. EXPERIENCE.

WEBINAR

CLINICAL ENDPOINTS AND TREATMENT EFFECT IN IMMUNO-ONCOLOGY

QUESTIONS

1. Since one may not know before a trial whether the proportional hazards (PH) assumption will hold, what is a more conservative (highest power) way to compute the sample size when using overall survival as primary endpoint?

This is a difficult question, without a simple answer. For instance, the logrank test will lose power in a non-PH setting, while the "late-difference" logrank test will lose power if the PH assumption holds. One could consider a "combination" of weighted tests, which would offer an improved power for non-PH alternatives (as compared with logrank), with (possibly) a limited reduction in power if PH does hold (as compared with, for instance, the "late-difference" logrank). It is not clear, however, how this would compare with the power of, for instance, an RMST-based test. Thus, it is difficult to say a priori what would constitute a "conservative" approach.

2. Can one use generalized pairwise comparisons to compute a sample size?

In theory, yes, though (at least, at this point) one will most likely have to resort to simulations rather than computations. Intensive research is being conducted at IDDI, Hasselt University in Belgium, and Claude Bernard University in Lyon, France, to better define the mathematical properties of generalized pairwise comparisons. This includes sample-size requirements. Read more at: https://www.iddi.com/news/news-events/statistical-method-for-personalized-medicine/

3. Can duration of response be a co-primary endpoint (for example, with response rate or PFS/OS being the other co-primary)?

Strictly speaking, duration of response cannot be used as primary or co-primary endpoint. If a co-primary endpoint will undergo formal hypothesis testing and be subject to rules based on the type I error, as usually done, it is hardly conceivable to use this metric under the current statistical framework. Duration of response is essentially a comparison based on a post-randomization factor, which introduces bias.

However, duration of response can be analyzed in an exploratory fashion, especially if precautions are taken, as explained in our webinar (see also Ellis et al., Contemporary Clinical Trials 2008), to minimize bias.

On the other hand, if duration of response is analyzed informally and used as corroborative evidence in favor of a treatment that has demonstrated benefit in other endpoints, especially the primary, then it is fair to say that an increased duration of response can add to the evidence provided by that treatment. However, it is possible to use the rate of durable responses, as explained in the webinar, as primary or co-primary endpoint. In this case, the endpoint is a categorical one (rather than time-to-event), and all patients contribute to the analysis.