Reanalysis or redefinition of the hypothesis?

Jody D. Ciolino¹ and Rickey E. Carter²*

¹Department of Preventive Medicine, Northwestern University, Chicago, IL, USA; and ²Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

This editorial refers to ‘Applying novel methods to assess clinical outcomes: insights from the TRILOGY ACS trial’, by J. A. Bakal et al., on page 385

Though a seemingly simple notion, defining a clear research question or hypothesis at the conception of a clinical study can be a daunting task for investigators. Through the TRILOGY ACS study, Bakal and colleagues illustrate that subtle changes in the primary outcome definition for a clinical trial may result in substantial changes in the analytical methods, altering the overall interpretation of the study’s result. Perhaps as an unintended consequence, their article also emphasizes a pitfall of post-hoc analyses. This editorial is intended to provide some insight into the statistical considerations in their article. In doing so, it aggregates their ideas to provide recommendations for appropriate analytical techniques driven by specific research questions.

Most disease progressions are associated with the realization of multiple complications. Cardiovascular disease, for example, may have elevated blood pressure, atherosclerosis, altered perfusion, ischaemia, changes in activities of daily living, etc. From a clinical perspective, a myocardial infarction resulting in death may be more important than reduced cardiopulmonary function that only limits a patient’s activities of daily living. That said, these extreme examples are intrinsically linked in a common disease process and represent different aspects of disease progression. These inter-relationships, when simultaneously modelled, often introduce the need for more sophisticated statistical analyses. These analyses can sometimes obfuscate the findings of the study. Thus, investigators often seek to balance clinical relevance with interpretability of the primary outcome measure. Composite outcome measures help with this process.

The TRILOGY ACS study, like many contemporary clinical trials, utilized a composite primary outcome measure of time-to-first occurrence of any of the following: death, myocardial infarction, or stroke. The appropriate statistical analysis based on this outcome compares the time-to-first-event across intervention arms without differentiating event severities (e.g. a cardiovascular death from a mild, non-disabling stroke). This analytical method addresses a research question that assumes that the severity of any such event in relation to others is unimportant, and the true question pertains to ‘event-free survival’.

Composite endpoints may be useful not only when the investigator is interested in event-free survival, but also when investigators cannot choose between equally important event definitions, the events or components all carry equal weight such that they may be exchangeable, and there is concern of type I error rate inflation due to multiple testing of multiple endpoints. By pooling the occurrence of any events, a trial typically gains power by increasing the number of events observed that may contribute to analyses. On the other hand, power may further increase if analyses account for any recurrent events (i.e. those after the first event where the first event was not death). For this scenario, the authors recognize the utility of the ‘Anderson–Gill’ method. Another challenge we face with the use of composite endpoints in the traditional manner lies in the difficulty in interpretation when components exhibit effects in different directions and/or magnitudes across intervention arms. Further, the research question may pertain not only to time-to-event, but also to overall event counts across arms.

Bakal and colleagues discuss the Anderson–Gill method as a potential alternative to the traditional survival analysis techniques since it uses all events (e.g. death, myocardial infarction, and stroke) from the TRILOGY ACS study to contribute to analyses (table 2 in Bakal et al.). Nevertheless, the authors caution that this method also carries drawbacks in that it treats all events as exchangeable with one another (i.e. death and a mild stroke may be given equal weight), and the methods may yield an overall biased result. In fact, the reanalysis of the TRILOGY ACS data using this method results in a statistically significant difference between intervention arms while the other methods did not. The reader is left to speculate on the possibility of a type I error.

The authors have proposed an alternative approach in the weighted composite endpoints (WCEs) in order to combat the shortcomings noted with the Anderson–Gill and traditional methods. The WCE, as the name implies, does not adhere to the traditional composite definition and each component of the composite outcome carries a weight. Analyses for this type of outcome address the research question pertaining to time-to-event and...
overall event counts while simultaneously taking into consideration the event severities in relation to one another. The WCE method requires that investigators select and provide weights for clinical events and/or measurements that are essential to understanding how an intervention affects the disease progression a priori. All participants begin the study with a ‘score’ of 100%, and any observed event results in a decrease in score with magnitude depending on the severity of the event. The most severe event, death, results in a score of 0% for the participant. The less severe events will diminish a participant’s score, but at a smaller scale such that his/her information may still contribute to the statistical analysis. VCE analyses replace the traditional survival curves with a decaying process that allows for multiple events.

The selection of the weights requires careful consideration. Through judicious choice of weights (and variables included in the composite outcome), one may be able to reach any desired conclusion that is more reflective of his/her working hypothesis. Bakal and colleagues emphasize that investigators must agree upon and document the weighting schedule (including the events comprising the composite endpoint) prior to study implementation. As with any element surrounding the statistical analyses of study data, the International Conference on Harmonization guidelines recommend avoidance of post-hoc determination of the elements of the composite endpoint.

Bakal and colleagues discuss yet another endpoint, the win ratio, which addresses the research question regarding event rates between arms while taking into consideration differing event severities. As the authors acknowledge, the win ratio method ‘approaches [study] data differently’ from the other methods mentioned, and it involves a matching mechanism which may lend itself to conflicting results and interpretations. Thus, the reader is cautioned with this approach.

This brings us back to the interpretation of Bakal’s reanalysis. Should the results of the TRILOGY ACS be reconsidered? Bakal and colleagues show that if one considers a recurrent event survival model (Andersen–Gill model), the observed difference between two intervention arms reaches statistical significance. The WCE results essentially fall between this recurrent event analysis (that did not involve any relative weighting mechanism) and the traditional time-to-first-event analysis. Of more relevance is perhaps whether the analyses properly addressed the research question(s). The article of Bakal and colleagues is important as it demonstrates the need for objective primary endpoints that maximize the statistical efficiency of a clinical trial, address a specific research question, and are defined a priori.

In an attempt to apply the results from Bakal et al., we suggest that investigators first focus on the true research question and the nature of the hypotheses: are they interested in differences in (i) time-to-first-event (regardless of the severity of the event) across arms; (ii) the time-to-event while incorporating overall event counts (again, regardless of severity) across arms; (iii) the event rate across arms based on matched pairs while incorporating event severity; or (iv) time-to-event while incorporating overall event counts and relative severity of events across arms. In the cases of (i), (ii), (iii), and (iv), we recommend the traditional time-to-first-event analyses, Andersen–Gill method, the win ratio method, and the WCE method, respectively. The research question, though sometimes difficult to specify unambiguously, should ultimately drive the choice of primary outcome which will in turn dictate the appropriate analyses for a given clinical trial. The results from Bakal et al. suggest that any method must be pre-specified and interpreted with caution.

Conflict of interest: none declared.

References


