Applying novel methods to assess clinical outcomes: insights from the TRILOGY ACS trial

Jeffrey A. Bakal1, Matthew T. Roe2, E. Magnus Ohman2, Shaun G. Goodman1,3, Keith A.A. Fox4, Yinggan Zheng1, Cynthia M. Westerhout1, Judith S. Hochman5, Yuliya Lokhnygina2, Eileen B. Brown6, and Paul W. Armstrong1*

1Division of Cardiology, University of Alberta and the Canadian VIGOUR Centre, 2–132 Li Ka Shing Centre for Health Research Innovation, Edmonton, AB, Canada T6G 2E1; 2Division of Cardiology, Department of Medicine, Duke University School of Medicine and the Duke Clinical Research Institute, Duke University, Durham, NC, USA; 3Division of Cardiology, Department of Medicine, St Michael’s Hospital, Toronto, ON, Canada; 4British Heart Foundation Centre for Cardiovascular Sciences, University of Edinburgh, Edinburgh, UK; 5Cardiovascular Clinical Research Center, Leon H. Charney Division of Cardiology, New York University School of Medicine and NYU Langone Medical Center, New York, NY, USA; and 6Eli Lilly and Company, Indianapolis, IN, USA

Received 2 December 2013; revised 9 May 2014; accepted 4 June 2014; online publish-ahead-of-print 10 July 2014

See page 340 for the editorial comment on this article (doi:10.1093/eurheartj/ehu311)

Aims

Several methods provide new insights into understanding clinical trial composite endpoints, using both conventional and novel methods. The TRILOGY ACS trial is used as a contemporary example to prospectively compare these methods side by side.

Methods and results

The traditional time-to-first-event, Andersen–Gill recurrent events method, win ratio, and a weighted composite endpoint (WCE) are compared using the randomized, active-control TRILOGY ACS trial. This trial had a neutral result and randomized 9326 patients managed without coronary revascularization within 10 days of their acute coronary syndrome to receive either prasugrel or clopidogrel and followed them for up to 30 months. The traditional composite, win ratio, and WCE demonstrated no significant survival advantage for prasugrel, whereas the Andersen–Gill method demonstrated a statistical advantage for prasugrel [hazard ratio (HR), 0.86 (95% CI, 0.72–0.97)]. The traditional composite used 73% of total patient events; 40% of these were derived from the death events. The win ratio used 66% of total events; deaths comprised 57% of these. Both Andersen–Gill and WCE methods used all events in all participants; however, with the Andersen–Gill method, death comprised 41% of the proportion of events, whereas with the WCE method, death comprised 64% of events.

Conclusion

This study addresses the relative efficiency of various methods for assessing clinical trial events comprising the composite endpoint. The methods accounting for all events, in particular those incorporating their clinical relevance, appear most advantageous, and may be useful in interpreting future trials. This clinical and statistical advantage is especially evident with long-term follow-up where multiple non-fatal events are more common.

Clinical Trial Registration

NCT00699998.

Keywords

Composite endpoints • Clinical trial • Acute coronary syndromes

Introduction

Phase three randomized clinical trials (RCTs) are intended to provide reliable and objective estimates of treatment interventions. The costs and complexity of RCTs, including those done in the setting of acute coronary syndromes (ACS), have increased steeply over the past decade, thereby setting a higher standard for the creative development of novel therapeutic strategies. The ability to demonstrate additional gains in treatment efficacy has also been attenuated by major improvements in treatment with corresponding reductions in mortality and morbidity rates in ACS over the past two decades.

These developments have engendered the adoption of an increasingly diverse set of related clinical outcomes gathered together to form a single composite endpoint. The selection of components...
for the composite should be clinically relevant. The assumptions of the composite endpoint are that each component endpoint should be meaningful; each endpoint should be equivalent in severity and that the treatment effect should be similar for each component. However, these assumptions are seldom met in practice. The time to the first event within the components of that endpoint often serves as the defining unit of measurement in a given trial. A composite endpoint is typically used to construct a larger signal, thereby reducing overall sample size. In response to further calls to improve the value of the information from RCTs, statisticians and trialists have developed novel methodologies designed to better reflect the overall patient experience and afford more meaningful trial interpretation.

The Targeted Platelet Inhibition to Clarify the Optimal Strategy to Medically Manage Acute Coronary Syndromes (TRILOGY ACS) trial (NCT00699998) evaluated 9326 study participants with unstable angina/non-ST-segment elevation myocardial infarction (UA/NSTEMI) who were managed medically without revascularization to investigate the effects of the thienopyridine platelet inhibitors prasugrel vs. clopidogrel. In TRILOGY ACS, the primary analysis was based on a composite endpoint of cardiovascular death, myocardial infarction (MI), and stroke and a neutral result was found. In contrast with many prior trials of dual antiplatelet therapy, longer-term follow-up was employed; additionally, a concomitant exploration of a lower dose therapy (i.e. a 5 mg prasugrel maintenance dose in patients < 60 kg and/or ≥ 75 years) was undertaken. Our intention herein was to use the neutral results of the TRILOGY trial as a template to explore various methodologies beyond the traditional time to event approach.

Given the complexity of composite endpoints and the simplicity of the traditional composite analysis, there is an opportunity to improve upon the use of trial outcomes information to more efficiently capture potential outcomes of interest. Hence the objectives of this study were to explore the attributes of different analytical approaches using the TRILOGY ACS trial as an example.

**Methods**

**Study design and participants**

The design and primary results of the TRILOGY ACS trial have been described previously. In the present study, we utilize all-cause mortality as an endpoint rather than cardiovascular death in each of the analyses. Patients with UA/NSTEMI were eligible for enrolment if their treating physician decided on a final treatment strategy of medical management without revascularization within 10 days of their index event; patients with unstable angina/non-ST-segment elevation myocardial infarction (UA/NSTEMI) who were managed medically without revascularization, but if such a procedure was planned, it had to be performed before randomization. Major exclusion criteria included a history of transient ischaemic attack or stroke, coronary revascularization within the previous 30 days, renal failure requiring dialysis, and concomitant treatment with an oral anticoagulant. Participants were randomly assigned to receive either a loading dose of 30 mg of prasugrel (n = 4663) or 300 mg of clopidogrel (n = 4663), which was followed by daily blinded maintenance administration of a study drug: 10 mg of prasugrel (or 5 mg for those < 60 kg of body weight or ≥ 75 years) or 75 mg of clopidogrel daily. Study treatment was administered for at least 6 months and up to 30 months. Concomitant low-dose aspirin treatment was strongly encouraged.

**Statistical analysis**

Participant characteristics for this trial were summarized according to randomly allocated treatment arm (i.e. intention-to-treat). The efficacy outcomes of interest in this study were time to all-cause death (death; to use with the prior weighted composite method and minimize censored cases), MI, and stroke. Multiple imputation techniques were utilized for missing baseline covariates. All analyses were done using SAS (version 9.3) and R (version 2.16).

**Models**

In addition to a comparison of the primary analysis method using time-to-first-event (all cause death, MI, stroke), a secondary analysis describing repeated ischaemic events among all components of the primary endpoint for the overall period and using a time-dependent model with separate hazard ratios (HR) before and after 30 days, 6 months, and 12 months (Andersen–Gill model) was undertaken. We also provide the secondary analyses of the data using the win ratio and the WCE.

**Win ratio**

The win ratio methodology provides a rank-based approach for assessing treatment superiority by first ranking and then pairing the patients between treatment groups according to risk score; in the current study, the Global Registry of Acute Cardiac Events (GRACE) Risk Score for 6-month mortality was applied to all patients and each group was ranked by the score independently, the patients were then paired by rank. Multiple imputation was used for patients missing data required to compute their GRACE Risk Score (n = 540, 6%). Each patient pair was then evaluated as to which (if either) member of the pair had a death event first. The remaining pairs (i.e. those with no death event) were then evaluated for the stroke events and then subsequently MI events. If the treatment arms were unbalanced in number, then the larger arm had members left unpaired. If one member of the pair was censored prior to the time of the event in the other member, the event was considered ‘unused’. With the intent to include these ‘unused’ events, new pairs were generated according to length of follow-up and then GRACE Risk Score.

Once the pairs were evaluated, the number of ‘wins’ (i.e. pairs in which the clopidogrel arm had the event first) were divided by the number of ‘losses’ (i.e. pairs in which the prasugrel arm had the event first) to provide the win ratio and corresponding 95% confidence intervals (CIs).

For pairs that had one member censored early or were unmatched, these subjects were then recomputed and paired by time groups (rather than GRACE score) to form a time stratified re-ranking which allows for more of the pairs to be used in the win-ratio calculation.

**Weighted composite endpoints**

The WCE methodology provides a generalization of the standard time to event methodology by determining a weight for each of the nonfatal events. In this analysis, each patient begins with a weight of 1.0, patients with non-fatal events were considered to have their contribution to the cohort size reduced in weight, such that the additional weight was lost for subsequent events and the full (or residual) weight was lost for a death event. For example, a patient with an MI on Day 2 and stroke on Day 3 would have a cumulative weighting of 0.67 = 1 − [(1 − 0.38) × (1 − 0.47)]. We can then create a modified life table with a weighted number of patients at risk. Two sets of weights were determined by the Delphi panel, which comprised 23 cardiologist/clinician-investigator members of the TRILOGY ACS steering committee during the planning stage of the trial. The first round of results was received from the larger
Table 1  Classification of endpoint severity

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition</th>
<th>Rule</th>
<th>Weight for WCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>Small periprocedural</td>
<td>Periprocedural (tick box on CRF) OR &lt; 5 × ULN Troponin OR (if no troponin available) &lt; 2 × ULN CK-MB OR (if no CK-MB available) &lt; 2 × CK</td>
<td>0.17</td>
</tr>
<tr>
<td>Moderate</td>
<td>Medium spontaneous</td>
<td>5–30x ULN troponin OR (if no troponin available) 2–10 × ULN CK-MB OR (if no CK-MB available) 2–10 × CK</td>
<td>0.38</td>
</tr>
<tr>
<td>Severe</td>
<td>Large with major ST change and substantial myocardial necrosis with accompanying left-ventricular dysfunction</td>
<td>&gt; 30 × ULN troponin OR (if no troponin available) &gt; 10 × ULN CK-MB OR (if no CK-MB available) &gt; 10 × CK</td>
<td>0.59</td>
</tr>
<tr>
<td>Stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>Transient ischaemic attack with brief quadratic visual field loss</td>
<td>Derived from text comments</td>
<td>0.23</td>
</tr>
<tr>
<td>Moderate</td>
<td>Stroke with significant speech and right arm weakness that entirely recovers in 3 months (non-disabling)</td>
<td>Derived from text comments</td>
<td>0.47</td>
</tr>
<tr>
<td>Severe</td>
<td>Severe right hemiplegia that permanently persists (disabling)</td>
<td>Derived from text comments</td>
<td>0.82</td>
</tr>
</tbody>
</table>

CK, creatine kinase; CK-MB, creatine kinase MB; CRF, case report form; ULN, upper limit of normal.

Results

There were a total of 1913 efficacy events (death, MI, stroke) and 210 safety events (moderate and major bleeds) among 9326 participants enrolled in TRILOGY ACS (Table 2). Using the traditional composite endpoint, there was no difference between prasugrel and clopidogrel at 30 months [HR, 0.96 (95% CI, 0.86–1.06)]. The traditional composite analysis incorporated 1389 of the overall total of 1913 events (i.e. 73%; first event per patient). The Andersen–Gill method, which counts multiple events per patient equally (n = 794), revealed a significant difference in favour of prasugrel [HR, 0.86 (95% CI, 0.72–0.97)] as in the primary analysis of the trial.

Win ratio

Ninety-four percent of patients had complete information on GRACE Risk Score. Missing events were imputed for the patients without data and their individual risk scores were then calculated. The number of patients was balanced between the two arms such that all 4663 patients in the prasugrel group were matched to the clopidogrel group based on their GRACE Risk Score rank. Following the initial ranking by GRACE Risk Score, there were 245 pairs with events (117 death events, 26 strokes, and 102 MIs) unused from the analysis due to the other member of the pair being censored prior to the event time of the first. After the time stratified re-ranking of the 245 missed pairs, 149 of these events were re-entered in the analysis. There were 331 pairs with a death in the prasugrel arm first compared with 348 in the clopidogrel arm first. The non-fatal outcomes were at the similar pattern between arms, with 348 vs. 225 for stroke and 225 vs. 37 for MI in the prasugrel and clopidogrel arms, respectively. In the win ratio, the interpretation of the treatment effect is the reverse of a conventional analysis; that is, conventionally we are concerned with the odds or hazard associated with an event. In the win ratio, a ‘win’ is determined by one member of the pair not having an event before the other. In our analysis, there were numerically more ‘wins’ for
prasugrel, with a ratio of 1.05 [95% CI, 0.94–1.18], which did not meet statistical significance.

**Weighted composite endpoint**

The Delphi panel determined the weights of 0.379 for MI, 0.469 for stroke, and 1.0 for death. Unlike the win ratio which considers only the relative time to event, for the WCE, we consider the time to weighted event between ‘each’ event; the modified Kaplan–Meier curves are provided in Figure 1. Compared with the traditional time to first (equally weighted) event, we observed a change in the location of divergence of the curves (as noted by the vertical line) at Day 420 (Figure 1B) rather than at Day 470 (Figure 1A). This is an indication of the impact of weighting the events and incorporates the differential in both the timing and type of event between the two treatment arms. There was no significant difference between treatment groups.

---

**Table 2 Distribution and weighting of events by analytical method**

<table>
<thead>
<tr>
<th>Method</th>
<th>Deaths</th>
<th>Stroke</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment arm</td>
<td>Prasugrel</td>
<td>Clopidogrel</td>
<td>Prasugrel</td>
</tr>
<tr>
<td>All events (max/pt)</td>
<td>385 (1)</td>
<td>409 (1)</td>
<td>62 (1)</td>
</tr>
<tr>
<td>Primary analysis</td>
<td>272 (70.6)</td>
<td>276 (67.4)</td>
<td>55 (88.7)</td>
</tr>
<tr>
<td>Andersen–Gill (%)</td>
<td>385 (100)</td>
<td>409 (100)</td>
<td>62 (100)</td>
</tr>
<tr>
<td>Win ratio (%)</td>
<td>306 (79.5)</td>
<td>338 (82.6)</td>
<td>32 (51.6)</td>
</tr>
<tr>
<td>Win ratio-time stratified (%)</td>
<td>343 (89.1)</td>
<td>362 (88.5)</td>
<td>37 (59.7)</td>
</tr>
<tr>
<td>WCE BASE (%)</td>
<td>385 (100)</td>
<td>409 (100)</td>
<td>62 (100)</td>
</tr>
<tr>
<td>WCE augmented</td>
<td>385 (100)</td>
<td>409 (100)</td>
<td>2/44/16 (100)</td>
</tr>
</tbody>
</table>

CV, cardiovascular; max, maximum; MI, myocardial infarction; pt, patient; sm/med/lg, small/medium/large.

*aAll deaths rather than CV deaths were used.

**Figure 1** (A) Survival curves for traditional composite. (B) Weighted composite endpoint modified-survival curves. (C) Weighted composite endpoint severity-modified severity curves. Pts, patients.
Weighted composite with severity weights

In this analysis, the two curves also diverge on Day 420 (Figure 1C) indicating that many patients had early non-fatal events in both arms that were then succeeded by more severe events later in follow-up. In Figure 2, these differences are shown by treatment in modified Kaplan–Meier curves for the three scenarios (traditional time to event, WCE, and severity-weighted composite result). Note that when the severity of weights is incorporated, most of the non-fatal events are seen to be of the less severe variety, as indicated by the flatter slopes of the curves (Figure 2).

Efficiency of event use

The distribution and use of events according to the various methods is provided in Table 2. Of note, the Andersen–Gill and weighted composite approaches include all of the endpoints, whereas the traditional analysis and the win ratio used the fewest of all collected data on events.

Figure 3A depicts the percentage of total events used by each method and the relative importance of each event type within that total. For example, in the traditional composite method, we observed that ~73% of the total events were used and that 40% of the relative importance was placed on death. In comparison, the Andersen–Gill method used 100% of events and shows that 42% of all events were deaths. Like the Andersen–Gill, the WCE used all events but instead death comprised 60% of the overall result. The win ratio used the smallest percentage (66%) of overall events and derived ~50% of the relative information from the death events.

Discussion

The current comparison of analytical approaches within the TRILOGY ACS trial provides insight into the relative efficiency and effectiveness of different methodological approaches to clinical outcome study data and potential implications regarding the interpretation of results. The strengths and weaknesses of each are summarized in Table 3. The various methodologies differ in their uses of the available information when the assumptions of a traditional composite outcome have been met. We consider these results in light of a previous observation from a simulation study which indicated that there was no additional loss in power resulting from the use of a weighting methodology.

The traditional time-to-first-event analysis, Andersen–Gill, and WCE methods all identify patients having at least one event. In the application to the TRILOGY data, the Andersen–Gill analysis was the only one to demonstrate statistical significance. We contend, however, that the more important difference in these methods relates to the manner in which events are counted; i.e. for Andersen–Gill, weights are considered equal, whereas in the WCE they are differentiated. Considering subsequent events appropriately is important as their occurrence have clear implications for both the health care costs and the quality of life for the patients: this is especially true when long-term follow-up is planned such as was the case in TRILOGY ACS. The incorporation of additional information...
beyond simple event-free survival (as in the case of the traditional composite) does miss potentially important time dependent information related to subsequent events. Moreover, each of these methods remains limited by the duration of follow-up where events occurring after the time frame are censored. Prior studies comparing the performance of the Andersen–Gill to the traditional models in the presence of heterogeneity across individuals have shown that in cases where there is a restriction on the maximum number of events per subject and censoring at the end of the follow-up, the Anderson–Gill method may give a biased result. Ultimately, these are key considerations in the design and execution of a clinical trial. The use of repeated events in general is also an important consideration as the rates of individual events should be reported to provide an understanding of which events occurred and when (e.g.

---

Figure 3  (A) Percentage use of events and types and relative importance of each event type by treatment. (B) Effective number and types of events used per 1000 patients enrolled by treatment expressed as percentage use. (C) Weighted non-fatal event type (%) by treatment arm for severity-modified weighted composite endpoint. AG, Andersen–Gill; MI, myocardial infarction; mod, moderate; WCE, weighted composite endpoint.
Applying novel methods to assess clinical outcomes

Table 3 Comparison of analysis attributes

<table>
<thead>
<tr>
<th>Treatment arm</th>
<th>Traditional time to first event</th>
<th>Andersen–Gill</th>
<th>Win ratio</th>
<th>WCE base</th>
<th>WCE augmented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses first event</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Uses all events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death as most important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiates within event types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to event used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A priori attribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10 individuals with one event are different than one individual with 10 events. An important consideration in the present study is that neither the traditional nor the Andersen–Gill model provide any differentiation on event type irrespective of potential bias or discrimination ability. Although there have been suggested modifications which allow the baseline hazard rate to vary with each event, these methods still count events similarly. For example, if we consider a patient with an MI and subsequent death, the Andersen–Gill counts these two events as identical in a single patient, whereas all of the other methods count only a single event; in this example, the traditional method would only count the MI, the win ratio would use only the death event, and the WCE would count the MI as a fraction and death as the remainder. Hence, in this scenario, the Andersen–Gill provides the only method which identifies a statistical difference but this comes at the expense of not discriminating between types of event. The Andersen–Gill analysis is intended to address repeated events of the same type and as such both events are considered equally important; this approach inherently adds additional weight to non-fatal events since patients are eligible for inclusion for multiples of non-fatal events. Further in the TRILOGY ACS trial, some patients in both treatment arms experienced six or more MIs. The results from this study are also similar to the primary TRILOGY ACS study which essentially utilized the same outcomes, except it assessed ‘cardiovascular’ death rather than all death. By utilizing all death as an outcome, we facilitated the analysis by reducing the censoring events.

An advantage of the design of the TRILOGY ACS trial was that there were no limitations on how many non-fatal events could be reported, and there was a relatively long (30-month) follow-up. Methods that do not account for the relative severity of non-fatal endpoints may overstate the difference between treatments. In the win ratio, if death occurs, only it would be counted. With the WCE method, multiple events are given importance based on their severity. When multiple MIs occur with continued survival, this scenario is considered an ‘improved state’ as compared with the death of a patient. The differing analyses explored in the current study demonstrate similar results between treatment groups, due in large part to the homogeneity in the magnitude and direction of the individual components. As observed in Figure 3C, late events in the clopidogrel arm appeared more severe than those in the prasugrel arm. Notwithstanding these findings, our major interest was to pursue the optimal use of available data rather than necessarily seeking a difference in the interpretation of the results. We have chosen to use weights derived by clinician-investigators, who are familiar with the impacts, severity, and implications to treatment of each one of these outcomes. Since these weights were utilized for comparative analysis in the current trial, we felt that, until there were well validated weights for the relative states, the outcome weights should be calculated based on their relative impact to the others.

By design, the win ratio approaches the data differently, and amounts to a ‘worst-event’ methodology that places increased relative importance on death events when compared with traditional time-to-first-event analysis. However, because it continues to weight the individual components such that if the worst event in the pairing was a death, then death is counted and other preceding non-fatal events experienced by a subject are excluded. Indeed, the largest percentage of events that went ‘unused’ by the differing methods we evaluated was observed in the win ratio setting. Further, if the worst event is an MI in a matched pair of patients, it is considered equivalent to death. An additional consideration when the win ratio is employed relates to the matching process itself. If not specified a priori, there is an opportunity to undermine the randomization process by changing the way the risks are matched. Further it relies on the capability of an existing risk score to quantify the risk of each of the outcomes accurately, and risks residual confounding through variables not included. Additionally, the proposed template aims to include the most patients using a rank-based matching procedure which optimizes sample size, rather than a risk-based match, which can easily shift the comparison between groups if there is even a small group of lower risk patient in one arm of the trial.

There have been recent developments suggesting the number needed to treat and number needed to harm framework13 deserve emphasis. A key issue arising from this approach is that the definition of ‘treat’ and ‘harm’ needs to be extended beyond the conventional dichotomous (yes/no) analysis in order to better understand its application and the incorporation of multiple events.

Because TRILOGY ACS was a randomized and blinded trial patients were evenly assigned to each arm the chance of differences in observed or unobserved covariates was minimized. Notwithstanding this process, systematic differences may exist and impose potential bias in all of the methods we evaluated including the win ratio where the model is selected post hoc. Another key methodological difference evident in the current report is how the results are portrayed. In both the traditional composite and WCE approaches,
time to event information is exhibited both graphically through a modified Kaplan–Meier curve and a confidence region is provided, whereas the win ratio and Andersen–Gill methods present only a single value and confidence interval for the trial results. Although the single value is sufficient to determine a potential treatment difference, understanding the pathway whereby this is derived is challenging (e.g. did the new treatment prevent more early events?).

In the present analysis, we also extended our prior use of the WCE approach by considering the differential severity of the non-fatal endpoints and identified two important novel features. First, in the survey we noted a 3.5-fold difference in the weight assigned to the mild vs. severe types of both MI and stroke; secondly, we noted that most of the adjudicated non-fatal events were of the mild type, highlighting the importance of differentiating types of non-fatal events. There were numerically fewer severe events within the prasugrel vs. the clopidogrel group.

Of further interest in the primary analysis of TRILOGY ACS was the differentiation of the events in the under 75 years of age vs. 75 years of age and older (i.e. those who received the reduced dose of prasugrel). We examined the groups separately for all of the included analyses and found that the results were similar in magnitude and interpretation.

There are both strengths and limitations to the use of the WCE methodology. Clinically it is understood that there is a benefit to the meaningful incorporation of all events and event severity in the analysis. The inclusion of the non-fatal event severity was included after the trial protocol was finalized and as such our classification of event severity was done with the best available information but deserves further assessment. Similarly, future validation of the associated weights is an exercise that should be completed before employing the method demonstrated here as a primary analysis. We have demonstrated that the WCE method is fairly robust to sensitivity analyses around the weights used. The longer-term follow-up in the TRILOGY ACS trial also demonstrates the importance of including all events, given that many of the early events were of the least-severe non-fatal types. Should the WCE be incorporated into future trial design? It does require some additional effort: (i) it would require that weights be established a priori, ultimately a standardized library of weights should be created and (ii) additional understanding of how the treatment is expected to impact each candidate endpoint is required as opposed to a single global effect estimate. However, based on our previous work there is unlikely to be a penalty in the number of subjects required and there may even be an advantage. In the simulation analysis presented previously, we demonstrated an improved ability to identify a difference in treatment effects in cases where the assumption that all component endpoints do not change in the same direction or magnitude. Precategorized weighting of the severity of differing non-fatal endpoints may also be worthy of prospective consideration by clinical events committees charged with adjudicating non-fatal outcomes in clinical trials.

Each of these methods studied herein offers a different analysis of typical clinical trial data. In order to maximize the efficiency of these analyses, we suggest that a WCE be considered since it not only incorporates all events but also addresses their clinical relevance. In this way, a more efficient and informed (i.e. appropriate priority given to more important components such as death) use of all events occurring in a clinical trial could help to generate a more optimal evidence-based approach to the interpretation of studies aimed at improving care.

Supplementary material

Supplementary material is available at European Heart Journal online.

Funding

The TRILOGY ACS study was funded by Eli Lilly and Daiichi Sankyo. S.G.G. is supported by the Heart & Stroke Foundation of Ontario Polo Chair, Department of Medicine at the University of Toronto.

Conflict of interest: E.M.O. reports grants from Eli Lilly and grants from Daiichi Sankyo, during the conduct of the study; grants from Datascope, grants from Gilead Sciences, personal fees from AstraZeneca, personal fees from Boehringer Ingelheim, personal fees from Bristol–Myers Squibb, personal fees from Gilead Sciences, personal fees from Janssen Pharmaceuticals, personal fees from Liposcience, personal fees from Merck, personal fees from Pozen, personal fees from Roche, personal fees from Sanofi Aventis, personal fees from The Medicines Company, personal fees from WebMD, outside the submitted work; M.T.R. reports grants and personal fees from Eli Lilly, personal fees from Daiichi Sankyo, during the conduct of the study; grants and personal fees from Bristol–Myers Squibb, grants and personal fees from Kowa Pharmaceuticals, personal fees from Novartis, personal fees from AstraZeneca, personal fees from GlaxoSmithKline, personal fees from Janssen Pharmaceuticals, personal fees from Kowa Pharmaceuticals, personal fees from Merck, personal fees from Sanofi Aventis, personal fees from Helsinn Pharmaceuticals, personal fees from Regeneron, outside the submitted work; S.G.G. reports personal fees and other from Duke Clinical Research Institute, during the conduct of the study; grants and personal fees from AstraZeneca, grants and personal fees from Bristol–Myers Squibb, grants and personal fees from Eli Lilly, personal fees from Sanofi Aventis, outside the submitted work; K.A.A.F. reports grants and personal fees from Eli Lilly, during the conduct of the study; grants and personal fees from Bayer, personal fees from Johnson & Johnson, personal fees from Janssen Pharmaceuticals, personal fees from Sanofi Aventis, grants and personal fees from AstraZeneca, outside the submitted work; P.W.A. reports grants and personal fees from Eli Lilly, during the conduct of the study; grants and personal fees from Regado Biosciences, personal fees from Eli Lilly, personal fees from Hoffmann La Roche, personal fees from Alexion Research/ Orexigen, grants from GlaxoSmithKline, grants and personal fees from Boehringer Ingelheim, grants from Sanofi Aventis, grants from Amylin Pharmaceuticals, grants from Merck, personal fees from Bayer, outside the submitted work; E.B.B. reports other from Eli Lilly (she was an employee) during the conduct of the study; J.S.H. reports personal fees from Eli Lilly, during the conduct of the study; personal fees from GlaxoSmithKline, outside the submitted work; J.A.B., Y.L., C.M.W., Y.Z. have nothing to disclose.

References


Downloaded from https://academic.oup.com/eurheartj/article-abstract/36/6/385/445850 by guest on 06 March 2019


