Evaluation of early percutaneous coronary intervention vs. standard therapy after fibrinolysis for ST-segment elevation myocardial infarction: contribution of weighting the composite endpoint

Jeffrey A. Bakal1, Cynthia M. Westerhout1, Warren J. Cantor2, Francisco Fernández-Avilés3, Robert C. Welsh1, David Fitchett4, Shaun G. Goodman4, and Paul W. Armstrong1*

1Division of Cardiology, 2-132 Li Ka Shing Centre for Health Research Innovation, University of Alberta, Edmonton, Alberta, Canada T6G 2E1; 2Southlake Regional Health Centre, Newmarket, Ontario, Canada; 3Hospital General Universitario Gregorio Marañón, Madrid, Spain; and 4St Michael’s Hospital, Toronto, Ontario, Canada

Received 13 August 2012; revised 17 October 2012; accepted 27 November 2012; online publish-ahead-of-print 19 December 2012

Aims
The selection of optimal endpoints for cardiovascular clinical trials continues to be challenging. We examined an alternative interpretation of a series of trials when the individual event severity is considered.

Methods and results
We analysed three contemporary myocardial infarction (MI) trials of early percutaneous coronary intervention after fibrinolysis, using a weighted composite method. This method allows the examination of the heterogeneity in the direction and magnitude of component endpoints, and multiple events (vs. first event). We incorporated a physician-assessed severity of each component endpoint in all patients for the five-item composite in the largest study, Trial of Routine Angioplasty and Stenting after Fibrinolysis to Enhance Reperfusion in Acute Myocardial Infarction (TRANSFER-AMI), which enrolled 1059 ST-elevation MI patients. The traditional approach yielded event-free survival probabilities of 0.89 [95% confidence interval (CI) 0.86–0.91] for the early invasive arm and 0.83 (95% CI 0.79–0.86) for the standard care arm (P = 0.004). After accounting for the clinician-investigator-determined weights, the effective survival probabilities were 0.93 (95% CI 0.91–0.95) for the early invasive arm and 0.93 (95% CI 0.90–0.95) with no significant difference (P = 0.54). The same pattern was observed in the three-trial cohort using a four-item composite with an observed improvement in event-free survival outcomes (P = 0.01), which was no longer apparent after the severity weights were considered (P = 0.44).

Conclusion
This analysis highlights the importance of considering the relative severity and multiple events in the evaluation of a clinical trial.

Keywords
Myocardial infarction • Angioplasty • Fibrinolysis • Trials

Introduction
Randomized clinical trials (RCTs) continue to be essential for the evaluation and approval of novel therapies and systems of care in medicine. Yet at a time when there is a compelling need to advance unmet clinical needs through clinical trials, their complexity and the enormous costs required to conduct them have become a major impediment to future research. This has engendered exploration of novel strategies that could result in more efficient and cost-effective approaches.
The selection of outcome measures is pivotal to trial design and it ultimately affects both their subsequent interpretation and potential incorporation into clinical practice. When the traditional composite endpoints methodology is employed, once a subject experiences any pre-specified component of the composite they are considered to have had an ‘event’ at the time of its first occurrence. In addition to combining differing endpoints within a single primary outcome, the use of composite endpoints has the benefit of increasing statistical power by adding more events for a given sample size.2 However, these advantages may come at the expense of a coherent interpretation of the treatment effect since not all components of the composite may be of equal importance: moreover, the assumption that they behave in a directionally similar fashion may not always hold.4

Based on recent calls to improve the metrics associated with RCTs, we developed a method whereby severity weights derived from the responses of a modified Delphi panel of experienced clinician-investigators were incorporated.4 This method has demonstrated promising properties with respect to the inclusion of additional information while maintaining or improving on statistical precision, particularly in the case of heterogeneity in the differences in endpoints.5

In the current study, we applied this weighted approach to the largest single trial comparing early percutaneous coronary intervention (PCI) against standard therapy in ST-segment elevation myocardial infarction (STEMI) patients receiving fibrinolysis.6 We then augmented this trial with two more large contemporary STEMI trials for further evaluation.7,8 The objective in assembling these trials was to evaluate whether an early invasive strategy offers an improvement over standard treatment after accounting for the severity of the individual endpoints and multiple events.

**Methods**

**Data**

The Trial of Routine Angioplasty and Stenting after Fibrinolysis to Enhance Reperfusion in Acute Myocardial Infarction (TRANSFER-AMI) trial is described in detail elsewhere.6 The first analysis in this study examines the 30-day outcomes parallel with the primary endpoint of the trial. The components involved in the composite are death (Death), shock (Shock), congestive heart failure (CHF), recurrent myocardial infarction (Re-MI), and recurrent ischaemia (RI). In order to augment the TRANSFER-AMI trial, we added the data from the two next largest trials in which these components were also captured, i.e. two arms of the Which Early ST-elevation myocardial infarction therapy (WEST) trial7 and Grupo de Análisis de la Cardiopatı´a Isque´-mica Aguda (GRACIA)-1.8 In GRACIA-1, alteplase was administered in STEMI patients presenting within 12 h of symptom onset with the possibility of rescue PCI in 251 patients and planned invasive strategy in 248 patients. In the two fibrinolytic arms of WEST, STEMI patients presenting within 6 h of symptom onset were given teneplase and randomized to either standard therapy (100 patients) or PCI within 24 h (104 patients). In this combined cohort, a four-item composite of Death, Shock, CHF, and Re-MI was examined in a total of 1762 patients within 30 days. In other trials in this area, the data were either not available or did not capture all of these endpoints.9–12

**Weights**

The weights for four of the five elements of the composite endpoint considered here and as the primary endpoint in TRANSFER-AMI used a five-component composite endpoint; we have considered four of these previously.9 To determine the weights, a Delphi panel of clinician-investigators were consulted to determine the relative clinical severity of Death, Shock, CHF, and Re-MI. They concluded that these events should have weights of 1, 0.5, 0.3, and 0.2, respectively. In order to more directly compare the primary composite of TRANSFER-AMI, we included the additional endpoint of RI which had not been considered by our Delphi panel. In order to address this, we created three scenarios, one where RI was given a moderate weight of 0.1 and then as a sensitivity analysis it was given lesser and greater weights of 0.05 and 0.3 as exaggerated limits on the weights. In the analysis of the augmented data from the three combined trials, we utilized the four endpoints from the original survey. We treated subjects with multiple endpoints multiplicatively so that, for example, a patient with a Re-MI on day 2 and CHF on day 4 would have a residual score of 1 – (1 – 0.2) × (1 – 0.3) = 0.44 on day 5. Using this approach, each patient has a score on each day: if a patient had no events, their score was 1; if events occurred, their score was reduced according to the type and number of events, thereby providing a contemporaneous weighted event rate.

**Statistical analysis**

The TRANSFER-AMI and the combined trial data were analysed and presented in two ways, both as a traditional time-to-first-event composite endpoint analysis and then using the weighted composite method.5 The weighted composite method utilizes the pre-specified weights and also allows for multiple endpoints. All results are those acquired within 30 days after the index event and presented as (modified) Kaplan–Meier curves and 95% confidence interval (CI), and modified log-rank scores. All analyses were generated using R 2.12.3 (Foundation for Statistical Computing, Vienna, Austria). All P-values were two-sided with statistical significance set at 0.05.

**Results**

The event rates, weights, and conditional times to event for the 1059 patients from TRANSFER-AMI are given in Table 1. The event rates are reported both as first event, similar to how they were reported in the original analysis of the trial, and as the total occurrence of each component event. The rates of Death and Shock in the early PCI group were directionally different compared with the less severe endpoints with Re-MI and RI. The largest difference in the time to event endpoints between the two treatment groups was observed for RI and Re-MI.

Table 2 provides the 30-day event rates for the combined cohort of 1762 patients from three trials, GRACIA-1, WEST, and TRANSFER-AMI, according to the two randomized arms. As expected, Re-MI comprised the largest fraction of the composite endpoint in the standard first-event analysis and thus is least affected when all events are considered. In contrast, both Shock and Death become more prominent and a larger component of the composite over the observation period: The rates of Death and Shock are numerically lower in the standard treatment arms of these trials whether we consider first-events or all-events in the analysis. The specific weights for each component of the composite endpoint are those used in the weighted analysis.
described above. The right panel of Table 2 provides median and inter-quartile range (IQR) in days to each event among those who experienced that event.

In Figure 1, the event-free survival curves and 95% confidence region for the standard treatment (red) and early PCI (blue) are shown, using both the traditional (left panel) and weighted composite (right panel) methods in the TRANSFER-AMI trial. In the left panel, the early PCI arm demonstrated a superior event-free survival at 30 days (0.89, 95% CI 0.86–0.91) compared with the standard treatment arm (0.83, 95% CI 0.79–0.86; \( P = 0.004 \)), which was based on the time to the first (and only) event. In the weighted composite analysis (right panel), the slightly higher rates of Death and Shock in the early PCI group outweigh the reduced number of less severe endpoints of Re-MI and CHF, leading to a weighted survival rate of 0.93 (95% CI 0.91–0.95), which was comparable with the standard treatment arm (0.93, 95% CI 0.90–0.95; \( P = 0.54 \)). This finding is also accompanied by narrower confidence limits than observed in the traditional time-to-first-event analysis.

The event-free survival curves for the combined cohort are given in Figure 2. In the left panel, differences were detected between the early PCI arm of 0.90 (95% CI 0.87–0.91) and the standard care arm (event-free survival 0.87; 95% CI 0.84–0.89; \( P = 0.01 \)) in the traditional analysis. After the weighting scheme was applied, the difference between the treatment arms was no longer apparent [early PCI arm: 0.94; 95% CI 0.93–0.96; standard arm: 0.94; 95% CI 0.92–0.95; \( P = 0.44 \) (right panel)], due in part to the difference in the least severe endpoint (Re-MI). As a sensitivity analysis, we examined a set of increased weights of 1.0, 0.7, 0.5, and 0.4 for Death, Shock, CHF, and Re-MI, as well as a set of more clustered weights of 1.0, 0.6, 0.4, and 0.3 for the endpoints and found similar results.

### Discussion

Using a novel method, we have presented an event-severity-weighted analysis in the GRACIA-1, WEST, and TRANSFER-AMI trials. When the severity of the endpoints is considered according to a pre-specified weighted scheme, the increments in the rates of the most clinically significant events (i.e. Death and Shock) in the control arms outweigh the value of the larger decreases in the rates of less severe events (Re-MI and CHF), leading to an overall attenuation of the treatment effect associated with the experimental intervention. This weighted approach also increased the precision of the estimated effect as demonstrated by the narrowing of the confidence regions for these new estimates.

There are two fundamental assumptions for the use of a composite endpoint: (i) each component is clinically meaningful; and
the directional change for all endpoints is similar. Notwithstanding the clinical relevance of the composite endpoints, when combined in the traditional composite the un-weighted combination has the potential to attenuate the differences in direction between the component endpoints.

The potential liability of using composite outcomes is that any overall benefit attributable to a novel intervention is presumed to relate proportionally to all elements of the reported composite result. However, this is not always the case. Similarly, the reporting of the individual components can make it difficult to determine the overall clinical implications of the result when their directions diverge: this challenge may be further compounded when the magnitude of the contribution of the components to the composite differs. Of note, a recent analysis of seven trials comparing an early invasive vs. standard approach to STEMI patients treated with fibrinolysis demonstrated no evidence of a change in the death rate (odds ratio [OR] 0.87; 95% CI 0.59–1.30), a reduction in the composite endpoint of death and Re-MI (OR 0.65; 95% CI 0.49–0.88), and also demonstrated the same heterogeneity in the effect sizes between the component endpoints (OR 0.86; 95% CI 0.57–1.30, for death alone; and OR 0.55; 95% CI 0.35–0.83 for Re-MI alone).

The interpretation of individual events within a composite can be hazardous, as there is a high likelihood that the study is underpowered to detect differences at that level. Moreover, the comparison of the component endpoints requires consideration of
multiple hypothesis testing. The weighted methodology does not invalidate single-outcome analysis, but it does offer some assistance in the interpretation of the value of the reduction relative to averting death events. The incorporation of the relative weights described herein has the advantage of permitting a single comprehensive summary that can address differing directions and magnitudes of the components of the composite associated with new treatments beyond the standard measures.

This integration of outcome severity and multiple events is a unique feature that the traditional method, based on uniform weights and consideration of only the first event, cannot address. This analysis demonstrates the importance and ultimately the efficiency of including the additional events, thereby allowing a significant ‘catch-up’ between the treatment groups once multiple events are considered. It is important to emphasize that our analysis does not undermine the findings of the original studies that were based on a traditional pre-specified design. However, we contend that it does provide a novel alternative platform for the future construct of new trials. Indeed, this approach has now been prospectively incorporated into the planned analysis of two ongoing clinical trials.

An alternative method for addressing the issues of composite endpoints in trials has been proposed. The win-ratio methodology relies on the subject-wise pairing of observations followed by a comparison of relative time between pairings to worst event (death) and then in succession comparing times to second and third worst event. After all of these paired comparisons are made, the number of times the worst event occurs first in each arm is compared to compute a ratio. Although attractive, there are several limitations of the win-ratio. For instance, it requires the ad hoc matching of patients, which is a subjective process and is data dependent (i.e. determining on which parameters to match and how to evaluate the ratio if the matching is altered). There is also a real possibility that a number of patients will remain unmatched and thus excluded from the analysis. The evaluation of endpoints is also based on a sequential selection of worst to least severe events (i.e. death, then non-fatal events) which require subjective ranking.

As in quality-adjusted survival studies, which are typically of longer follow-up, critics of value-based methods will frequently cite the subjectivity of the weights. However, a similar objection applies to the decision to include or exclude various components within a composite. Properly determined weights allow for the additional discrimination beyond rank alone. The consistency of the results of our panel-derived weights indicates that the panel members likely combined their assessment of the STEMI complications associated with the non-fatal outcomes and their previously established prognostic implications. The ultimate goal of this type of analysis is to create a platform whereby the weights can be determined by investigators a priori and regulatory boards during the design phase of a trial. Consistent definitions of endpoints such as suggested by the Clinical Investigation Group of the Joint Task Force on the Universal Definition of MI are likely to facilitate cross-trial comparisons. In this regard, a recent literature review and the evaluation of the commonly employed multiple adverse cardiac events (MACE) by Kip et al. is relevant. These authors concluded that the substantial heterogeneity among individual study-specific outcomes used to define MACE led to markedly different results for patients with vs. without MI during 1-year follow-up that occurred in conjunction with randomized trials comparing drug eluting with bare metal stents in PCI. They recommended that the term MACE be abandoned in this context, and that efficacy and safety outcomes be evaluated separately with well-defined internally coherent components (before their integration), appropriately aligned with the diversity of clinical entities to be studied. These changes would then allow for improved estimation of sample sizes and ultimately facilitate discussions around net clinical benefit, in which safety endpoints (e.g. haemorrhagic complications) not addressed in the current study could be similarly weighted in the context of anticipated efficacy. Additionally, further refinement should emerge by addressing the inherent prognostic heterogeneity that exists within individual elements of the composite endpoints such as Re-MI. Ultimately, the designation of the weights deserves extension to a broader range of health providers and importantly should incorporate the view of patients.

The current study demonstrates the additional insights that may emerge by using a more informative evaluation of clinical endpoints for interpreting the results of clinical trials. Most will agree about the important gains made through the evidence-based medicine paradigm. By incorporating the event severity through methods such as the weighted composite, we are better able to make use of all the available evidence.

Funding
The current study did not receive restricted funding. The following discloses the funding sources for the trials which provided data for the current study as they were reported in the primary publication of their results. The TRANSFER-AMI trial was supported by a grant from the Canadian Institutes of Health Research and an unrestricted research grant from Roche Canada. Coronary stents were provided free of charge by Abbott Vascular Canada. The GRACIA-1 trial was funded by unrestricted grants from the Spanish Ministry of Health, the Spanish Network for Cardiovascular Research (National Health Institute ‘Carlos III’), the Spanish Society of Cardiology, Guidant, and Eli Lilly. Additional support was obtained from Guidant to reimburse interventional centres for the cost of stents. The WEST trial was supported by unrestricted research grants from Hoffmann-La Roche and Sanofi-Aventis Canada and also Eli Lilly Canada.

Conflict of interest: W.J.C. reports receiving consulting fees from Sanofi-Aventis and speaking fees from Sanofi-Aventis and Roche Canada; D.F., receiving consulting fees and speaking fees from Sanofi-Aventis; S.G.G., receiving consulting fees from Sanofi-Aventis, Bristol-Myers Squibb, and Roche Canada, speaking fees from Sanofi-Aventis and Bristol-Myers Squibb, and grant support from Roche Canada, Sanofi-Aventis, and Bristol-Myers Squibb; P.W.A., receiving speaker honoraria from Hoffman-LaRoche and Sanofi-Aventis; and R.C.W., receiving speaker honoraria from Hoffmann-LaRoche and Sanofi-Aventis. J.A.B., F.F.-A., and C.M.W. have no conflicts of interest to disclose.

References