Primer on Statistical Interpretation or Methods

Informed Choice of Composite End Points in Cardiovascular Trials

Guadalupe Gómez, PhD; Moisés Gómez-Mateu, MSc; Urania Dafni, ScD

Abstract—A composite end point is often used as the primary end point to assess the efficacy of a new treatment in randomized clinical trials. In cardiovascular trials, the often rare event of the relevant primary end point (individual or composite), such as cardiovascular death, myocardial infarction, or both, is combined with a more common secondary end point, such as target lesion revascularization, with the aim to increase the statistical power of the study. Gómez and Lagakos developed statistical methodology to be used at the design stage of a randomized clinical trial for deciding whether to expand a study-relevant primary end point to the composite of the relevant end point and a secondary end point. The method uses the asymptotic relative efficiency of the logrank test for comparing treatment groups based on the relevant end point versus the logrank test based on the composite end point. The method is used to assess, in the cardiovascular research area, the characteristics of the candidate individual end points that should govern the choice of using a composite end point as the primary end point in a clinical trial. A set of recommendations is provided based on the reported values of the frequencies of observing each candidate end point and on the magnitude of the effect of treatment as expressed by the hazard ratio, supported by cardiovascular randomized clinical trials published in 2008. (Circ Cardiovasc Qual Outcomes. 2014;7:170-178.)

Key Words: asymptotic relative efficiency • cardiovascular drugs • clinical trials, randomized • composite end point • logrank test

The conclusions of a clinical trial rely on its primary end point (PE), and thus, at the design phase, it is of outmost importance to choose the most appropriate PE. Composite end point (CE) is an event that is considered to have occurred if any one of several different events or outcomes (components) is observed.1,2 CEs are nowadays used commonly as the PE to assess the efficacy of a new treatment. In cardiovascular trials, a CE is more often used than not, incorporating either terminal outcomes, death from any cause, cardiovascular death or not, such as myocardial infarction (MI), stroke, and hospitalization. One objective is to increase the power to detect a significant benefit induced by the new treatment. This increase, in the case of time-to-event end points, is expected to be achieved by the inclusion of component end points that occur with higher frequency or earlier than the main events of interest.3 However, adding less specific components might in fact lead to loss of power to detect the true treatment differences.

Gómez and Lagakos4 provided a methodology to reach an informed decision on the PE at the design stage of a clinical trial. The objective of the present study was to examine the use of CE in cardiovascular randomized clinical trials (RCT) systematically, to illustrate this statistical methodology by means of case studies, and to use it to provide guidelines for the informed choice of PE in the context of cardiovascular clinical trials.

Background

Composite End points

CEs, also referred as combined outcomes, are defined as a single measure of effect based on a combination of a variety of clinically relevant individual end points. Many trials measure dichotomous end points after a certain period of treatment or follow-up and combine them into a single composite outcome that is considered to have occurred if any of the individual outcomes is observed. Health indices and total scores, defined as a weighted or unweighted combination of multiple item scores, counts or other end points are also referred as composite outcomes.5 When the focus is time-to-event analysis, a CE is defined as the time from randomization until the time that the first component of the CE occurs. For instance, time to Major Adverse Cardiovascular Event (MACE) is a CE defined as the time to either cardiovascular death or reinfarction or target vessel revascularization for ischemia or stroke, whichever occurs first (other definitions of MACE are possible).

The rationale for using CE is well described in Neaton.2 A CE, which is being used as a PE to provide strong scientific evidence about efficacy, should be clinically relevant, easily ascertainable in all patients, capable of unbiased assessment, sensitive to the hypothesized effects of the treatment and inexpensive to measure. The main reason for considering a CE instead of a single event as PE is the reduction of sample size.
size, consequence of increasing the event rate in the control group. Another reason is to avoid the problem of competing risks, as can be the case for end points that do not include the mortality component that are problematic because patients who died before the end point of interest are likely not at the same risk as the survivors. A third argument is based on the need of combining multiple measurements into a single CE, when a single primary variable cannot be selected from multiple measurements associated with the primary objective, as stated in the International Conference On Harmonization guideline.7

It is well established that a CE in a RCT should only be used if (1) the individual components are clinically meaningful and of similar importance to the patient, (2) the expected effects on each component are similar based on biological plausibility, and (3) the clinically more important components should at least not be affected negatively.8

Although the motivation to use a CE, other than multiplicity, is to get higher event rates, smaller sample sizes, and shorter follow-up times, several authors have cautioned against the use and interpretation of a CE. On one hand, loss of power could arise if the treatment effect for the components (1) is not of similar magnitude or (2) goes in the opposite direction for some of them. On the other, the treatment benefit with a CE on which the component end points are of highly different clinical importance can be problematic because the treatment might only beneficially affect the less important end points and thus give a misleading impression.2

CEs in Cardiovascular Research

In the past 10 years, many authors have addressed the issue of using and interpreting CEs in the cardiovascular research area. In what follows we present a brief summary of relevant readings. Freemantle et al.9 examine the use of CE in major clinical trials, by means of a selection of 167 RCT (with a total of 300 276 patients), that include a primary CE incorporating all-cause mortality, assess the arguments for and against CE, and provide guidance on their applications and reporting. He acknowledges the inadequate reporting of CEs used as primary outcome measures in randomized trials, concluding that, often, the reported results apply to the individual components of the CE rather than to the overall CE.

Ferreira-González et al.10–12 use MEDLINE to conduct 2 systematic reviews to investigate the rationale, potential problems and solutions of using CEs. They point out that the CE, by capturing the net benefit of the intervention, could give a more appropriate reflection of the clinical spectrum of important outcomes associated with the disease being treated than would any component alone. In the conclusions, it is stated that the use of CE is often complicated by the magnitude of the effect of treatment across component end points and by the relative importance of the different components for the patients.

The reader is referred to Huque et al.13 for an excellent introduction together with some key considerations for using a CE. They present as well some solutions through applications of multiple testing strategies.

Statistical Method to Reach an Informed Decision

Consider a 2-arm RCT involving random assignment either to an active treatment or to a control treatment. The individuals are followed from randomization until the event of interest, or until the end of study, whichever occurs first. We have a study relevant end point (RE) that could be used as the PE for efficacy, and a secondary end point that could be viewed as an additional end point (AE) of interest. For example, assume RE is the composite of cardiovascular death and MI, whereas AE is target lesion revascularization. We consider the CE as the combination of RE and AE.

To make an informed decision on whether the RE or the CE should be the PE, Gómez and Lagakos4 develop a strategy based on the behavior of the asymptotic relative efficiency (ARE) of the logrank test for comparing treatment groups with respect to RE versus the logrank test with respect to CE. The ARE is a measure of the relative power of the 2 tests and is interpreted as the ratio of the required sample sizes to detect a specific treatment effect to attain the same power for a given significance level.4 That is, if ARE=0.7, we only need 70% as many cases to attain a given power if we use RE as we would need if we used CE. Values of ARE>1 are in favor of using the CE instead of the RE. However, because the advantage of one end point over the other is small in the vicinity of 1, Gómez and Lagakos4 propose as a general rule to use the CE instead of the RE if ARE>1.1 and to retain the RE if ARE<1.1.

The computations for the ARE depend on (1) whether the 2 end points of interest include a terminal event (death), (2) the probabilities \( P_R \) and \( P_A \) of observing events RE and AE, respectively, for the control group, (3) the treatment effect with respect to RE and AE given by the hazard ratios \( HR_R \) and \( HR_A \), and (4) the correlation between the times to event RE and AE. Gómez and Lagakos4 reproduced several frequency situations by taking probabilities \( P_R \) and \( P_A \) equal to 0.05, 0.15, 0.30, and 0.50. The relative treatment effect on the RE was set to \( HR_R = 0.5 \) or 0.7, and it was combined with 6 different values for \( HR_A \) reproducing situations where the beneficial effect on the AE was larger, the same or smaller than on the RE.

An interactive Website, CompARE, still in a beta version and available from the second author under request, has been designed to calculate ARE values based on the information of the different REs together with the anticipated values of \( P_R \), \( P_A \), \( HR_R \), and \( HR_A \). The design of this free tool allows the user to enter their own values when designing a clinical trial (Figure 1) and shows in plots, as the ones reproduced in Figures 2 and 3, different scenarios by combining different range of values for the parameters.

Survey of the Use of CEs in the Cardiovascular Literature

Identification of Published Clinical Trials that Used CEs

The extent of use of CEs in the recent literature was explored through a systematic Medline search covering the 2008 publication of RCTs in 6 high impact medical journals (Table 1). Medline search, was restricted to randomized controlled
trial and human subjects publications, including the terms coronary artery disease, valvular heart disease, arrhythmia, cardiomyopathy, congestive heart failure, cardiovascular, or cardiovascular disease in the abstract, title, or keywords. The systematic search resulted in 216 publications. The ones that mentioned in the abstract, title, or keywords, a composite or combined end point, or the specific end points of MACE, or Net Adverse Clinical Events (NACE) were selected (87 of 216). Studies that dealt with other diseases looked at sub-group or nonrandomized comparisons or did not use time-to-event end points were excluded (26 of 87). A total of 61 clinical trials were considered for exploring the use of a CE (Figure 4). The breakdown by journal is presented in Table 1. The complete reference list is available in the Data Supplement.

Information Abstracted From Each RCT

The following information was abstracted from each of the published articles: time to follow-up, sample size, components of each primary and secondary end point, frequency of occurrence of each end point (CE and components of interest), the corresponding HRs and P-values between groups compared in the trial.

Method to Set the Recommendations

From the information abstracted for each trial together with previously examined scenarios in Gómez and Lagakos, we establish all the possible parameter combinations. However, because not all the combinations of frequencies and relative treatment effects (P_R, HR_R) or (P_A, HR_A) were found in the studied RCTs, we did restrict our computations to published pairs of values (P, HR). The ARE is computed for each of a total of 320 combinations to provide recommendations for the cardiovascular area trials. In all cases, computations have been done assuming that death is part of the RE, modeling the marginal laws of the times to RE and AE as Weibull, representing decreasing, constant and increasing hazard functions, combining each scenario with different degrees of dependence between times to RE and to AE and using HR=0.99 to represent relative treatment effects of no interest.
Case Studies

Interesting cases of trials leading to a significant result for the RE whereas nonsignificant for the CE, significant for the CE driven by the effect on the RE and nonsignificant for the RE whereas significant for the CE are described next, the first two with greater detail.

Case Study 1: Treating Patients After an Acute Coronary Syndrome With Succinobucol

An RCT to assess the effects of the antioxidant succinobucol (AGI-1067) on cardiovascular outcomes in patients with recent acute coronary syndrome already managed with conventional treatments, uses as PE, denoted by CE, the composite of RE (time to first occurrence of cardiovascular death, resuscitated cardiac arrest, MI, stroke), and AE (unstable angina or coronary revascularization; Figure 5). A total of 6144 patients having experienced an acute coronary syndrome ≤1 year before recruitment were randomized to receive succinobucol (n=3078) or placebo (n=3066), in addition to standard of care. A beneficial effect of succinobucol on RE was found (207 events: placebo). Using the notation we have already introduced, we have that the probability of observing the RE is $P_R=8.2\%$ with observed $HR_R=0.81$, whereas the probability of observing the AE is $P_A=10.4\%$ with $HR_A=1.05$ (it corresponds to coronary revascularization, whereas observed $HR$ for unstable angina is 1.10).

The ARE is explored for these parameter values. For all different shapes of the time-to-event distributions (9 combinations including increasing, constant, and decreasing hazard functions) and correlation values ranging from 0.15 to 0.75 (63 scenarios), it is found that the ARE is always <1.1. Following the rule of Gómez and Lagakos, the benefits of using the CE over the RE are marginal and probably too small to justify adding the AE.

The use of CE would be justified in the case that $HR_R \leq 0.85$, for all other parameters fixed (ie, $P_R=8.2\%$; $HR_A=0.81$; $P_A=10.4\%$; Figure 2). However, if $HR > 20\%$, even an expected frequency of 20% for the AE would justify the use of CE. If $HR_R=0.9$, CE would only be justified if $P_A \geq 20\%$, and the association between RE and AE is weak (not shown).

Thus, under these circumstances, the additional components of coronary revascularization or hospitalization for unstable angina on the PE would had only been recommended if the expected beneficial effect of succinobucol on these components would have been approximately as strong as the expected effect on cardiovascular death, resuscitated cardiac arrest, MI, or stroke (Figure 6).

| Table 1. Summary of Medline Search, for Cardiovascular Terms, for 2008 Publications of RCTs |

<table>
<thead>
<tr>
<th>Journal (Papers and RCT)</th>
<th>Total Articles</th>
<th>% CE RCT</th>
<th>% RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEJL</td>
<td>46</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>The Lancet</td>
<td>36</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>European Heart Journal</td>
<td>54</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>Circulation</td>
<td>53</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>JAMA</td>
<td>24</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Annals of Internal Medicine</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total RCT</td>
<td>216</td>
<td>100</td>
<td>61</td>
</tr>
</tbody>
</table>

CE indicates composite end point; and RCT, randomized clinical trials.

Figure 3. Asymptotic relative efficiency (ARE) of composite versus relevant end point (CE and RE, respectively) for a range of Spearman correlation coefficients and different values of $HR$ for the parameters of case study 2 ($P_R=0.083$; $HR_A=0.6$; $P_A=0.05$) and marginal increasing hazards. $HR$ indicates hazard ratio.

Case Study 2: Treating Hemorrhagic Complications During Primary Percutaneous Coronary Intervention in Acute MI

The Harmonizing Outcomes with Revascularization and Stents in the Acute MI (HORIZONS-AMI) study is a prospective, open-label, randomized, multicenter trial in patients with ST-segment–elevation MI presented within 12 hours after the onset of symptoms. In this study, 3602 patients were assigned to treatment with heparin plus a glycoprotein IIb/IIIa inhibitor (n=1802) or the alternative treatment of bivalirudin alone (n=1800). The interest lies on whether hemorrhagic...
Complications are reduced, when using bivalirudin alone. Two primary 30-day end points were prespecified: (1) major bleeding, denoted by RE and (2) NACE, denoted by CE, a composite of major bleeding and MACEs. MACE, denoted by AE, is composed, in this trial, of death, reinfarction, target vessel revascularization for ischemia and stroke. In this case, while major bleeding is the relevant event of interest, the composite CE takes into account all other additional adverse clinical events, including death. According to the results, MACE is almost identical in the 2 groups (98 versus 99 events; \(P=0.95\)), whereas major bleeding is statistically significantly lower in the bivalirudin-alone group (89 versus 149 events; \(P<0.001\)). The comparison of NACE (166 versus 218 events; \(P=0.005\)) between treatment groups is found statistically significant, and as mentioned by the authors, this is entirely driven by the effect on major bleeding. The risk taken by the researchers of combining the end point of interest with an end point on which treatments have no differential effect is demonstrated using this study.

The probability of observing a major bleeding event, \(RE\), is \(P_R=8.3\%\) with \(HR_R=0.6\), whereas the probability of observing a MACE event, \(AE\), is \(P_A=5.5\%\) with \(HR_A=1\) (Figure 1). MACE is occurring with smaller frequency than the RE and in addition the treatment does not have an effect on it. Under these parameter values the ARE is examined, as above, for 21 scenarios, corresponding to different shapes of the time-to-event distributions (including decreasing, constant, and increasing hazards) and correlation values ranging from 0.15 to 0.75. In the vast majority of cases, the ARE between a major bleeding event and a MACE event is <1.1, meaning that the use of the CE (NACE) is not recommended.

Other scenarios were also explored under all above combinations of distributional shapes and correlation values. First, for higher values of the probability of observing a MACE event (5.5\% \(\leq P_A \leq 8.0\%\)), a similar pattern emerges, with a sparse number of cases (5 of 84 scenarios), with ARE >1.1, leading to the recommendation of NACE, with all cases occurring for correlation of 0.75. Second, the ARE was also explored for larger beneficial effects on MACE (0.3 \(\leq HR_A \leq 0.9\)), and the ARE value is <1.1 except for 11 cases (of 105). Figure 3 illustrates the AREs for the values of the parameters of this clinical trial (\(P_R=8.3\%\); \(HR_R=0.6\); \(P_A=5.5\%\)) and for marginal

**Figure 4.** Flow chart for systematic review of cardiovascular (CV) randomized clinical trials (RCTs). CE indicates composite end point.

**Figure 5.** Pictorial representation of the construction of a composite end point (CE) as the union of the relevant end point (RE) and the additional end point (AE) based on Tardif’s randomized clinical trial.

**Figure 6.** Summary of recommendations for case study 1 as a guide to decide between using composite end point (CE) or relevant end point (RE) as primary end point (PE). Values of treatment effect on RE and relative frequency of RE and AE (\(HR_A=8.01\); \(P_A=8.2\%\) and \(P_A=10.4\%\)) are fixed in advance and correspond to Tardif randomized clinical trial.
increasing hazards. Globally, in 88% of the scenarios, the use of RE is recommended.

It is clear that the chosen PE, NACE, for the efficacy of bivalirudin alone in this study gave unexpected good results and that it was a matter of luck not to have a diluted effect in NACE because the ARE can be as low as 0.2 if the beneficial effect on MACE is 0.5, meaning that major bleeding as a PE can be as much as 5 times more efficient than NACE.

One could wonder under which circumstances the composite NACE would have been a better, more efficient choice, and by running all the ARE computations for different values of the frequency of observing an AE, we find that for the composite NACE to be justified, both a high frequency of observing MACE events as large as 70% and a strong association between bleeding and MACE are needed.

Case Study 3: Testing Fondaparinux in Patients With ST-Segment–Elevation MI

In the clinical trial testing fondaparinux in patients with ST-segment–elevation MI,17 the RE of death and the AE of myocardial reinfarction at 30 days occurred in 12.5% (P = 0.125) and 3.7% (P = 0.037) of control patients, respectively. The CE occurred in 15.1% of control patients, indicating a weak correlation between RE and AE. The corresponding HRs (HR = 0.83 and HR = 0.66) were both not significantly different than 1. The increased number of events for the CE and the same direction of benefit for both components led to a statistically significant HR with respect to CE of 0.80. In this trial, the use of the CE is clearly indicated by the ARE in 100% of the simulations.

Case Study 4: Prevention Studies

A prevention study assessed the benefit on the risk of cardiovascular disease of low-dose aspirin in the prevention of atherosclerotic events in patients with type 2 diabetes mellitus.18 Composite PE was defined as fatal and nonfatal ischemic heart disease, fatal or nonfatal stroke, and peripheral arterial disease. This trial could be considered an outlier because of the combination of a low frequency of fatal cardiovascular events (P = 0.008), yet significantly different between groups (HR = 0.10; P = 0.0037). The CE occurred in 6.7% of control patients, indicating a weak correlation between RE and AE, leading to a HR* of 0.80 but not statistically significant.

Under these extreme conditions, the use of the RE would have been justified based on the low HR, whereas the use of the CE would have been justified based on the low frequency of events. The ARE points to the clear choice of the CE for anticipated strong effects of the aspirin on the nonfatal events (HR ≤ 0.2) and the clear choice of the RE for moderate effects (HR ≥ 0.8), whereas for HR values between 0.2 and 0.8, the CE is recommended as HR increases for progressively higher values of the frequency of nonfatal events. In this particular situation, the choice of the CE based on an assumption of a treatment effect at such an extreme value would be difficult to justify at the design stage although it could be taken under consideration for the next trial designed on this question.

Results and Recommendations

A CE was used as PE for 47 of the clinical trials and as secondary for the remainder of 14 clinical trials. The frequency of use of different CEs, as well as of each individual component, for the 47 cases that CE is the PE, is presented in Table 2. MI and stroke were encountered as components of the CE in over half of these clinical trials (66% and 55%, respectively), Hospitalization and target vessel revascularization are AE in 30% and 13%, respectively, whereas death is encountered in all of them but 1 (46 of 47). In addition, among the 14 trials with an individual PE, in 13 of them death is either the RE (in 4) or used as an AE (in 9).

For all the trials, including death (46 out of 47), the frequency of death was relatively low (median 4%), with the exception of 3 trials where death was frequent (>20%). The observed relative frequencies of death among the 43 low-frequency studied trials were between 0.002 and 0.15 (Table 3). The observed relative frequencies of the AEs (MI, stroke, hospitalization, and target vessel revascularization) were between 0.002 and 0.31. Concerning the relative treatment effects, it was found that some of the component end points had an observed HR > 1 (17 of 43). Among the clinical

Table 2. Frequencies for Different Combinations of End Points for 47 RCTs With Composite End Point as Primary End Point

<table>
<thead>
<tr>
<th>End Point</th>
<th>Combinations</th>
<th>Death</th>
<th>MI</th>
<th>Stroke</th>
<th>Hospitalization</th>
<th>TVR</th>
<th>N With Additional End Points</th>
<th>N Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>...</td>
<td>8</td>
<td>14 (30)</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>5</td>
<td>8 (17)</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>1</td>
<td>6 (13)</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>5</td>
<td>11 (6)</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>2</td>
<td>11 (6)</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>2</td>
<td>9 (4)</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>2</td>
<td>4 (2)</td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>1</td>
<td>2 (1)</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>X</td>
<td>1</td>
<td>2 (1)</td>
</tr>
<tr>
<td>10</td>
<td>...</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>1</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

98% 66% 55% 30% 13% 28 47

MI indicates myocardial infarction; and TVR, target vessel revascularization.
trials with HR<1, we have found relative treatment effects for death as small as 0.1 and as large as 0.98 and between 0.35 and 0.94 for the AEs (Table 3).

In the reviewed studies, specific combinations of the control group frequencies for the RE (P_R) and AE (P_A) with corresponding HR values emerged. The ARE of a CE with death as a RE adding MI, stroke, or hospitalization as AE is computed for different shapes of time-to-event distributions and a range of correlations between times to RE and AE and is described next to serve as a guide for the design of future trials.

### Death Plus MI

For the relatively low frequency of MI (AE; ≤12%) for all HR combinations found in the trials, the CE of death and MI is almost always justified based on the ARE except for the case where death and MI present with the same frequency and the beneficial effect on death is higher than on MI (HR_A>HR_R).

### Death Plus Stroke

For particularly low frequency of stroke found in the trials (0.5%), the CE of death and stroke is always justified in the cases that the beneficial effect on stroke is higher than on death (HR_A<HR_R). The same is true for the higher frequency of stroke (12%), whereas the CE is also justified when the beneficial effect on stroke is slightly less than on death, but death presents with lower frequency.

### Death Plus Hospitalization

The CE is justified in the cases that the HR for death is >0.8 or 0.70 coupled with low frequency of death (P_R=3%), whereas the HR for hospitalization is <0.9. For a substantial benefit on death coupled with low frequency (HR_R=0.5; P_R=6%), when the frequency of hospitalization is high (P_A=39%) even for a smaller benefit for hospitalization (HR_R=0.7), the CE is justified.

The CE is not justified when, even for a substantial benefit on death (HR_R=0.5), low frequency of death (P_R=6%), and high frequency of hospitalization (P_A=39%), the benefit for hospitalization is small (HR_A>0.90). The CE is neither justified when HR_A>HR_R provided that the frequency of death is higher (P_R=12%).

### Death From Cardiovascular or Death From Any Cause as the Individual Primary or Co-PE

In only 4 trials, death from cardiovascular or death from any cause was used as the individual primary or co-PE. The frequency of cardiovascular death or any death in 2 of the trials on patients with New York Heart Association class II–IV Chronic Heart Failure, or Atrial Fibrillation and New York Heart Association class II or IV heart failure, was 25% and 29%, respectively. In such cases of high death frequency, the use of the CE is justified only when the anticipated treatment benefit for the AE is similar or higher than the one for survival. Such is the case in the trial exploring the effect of n-3 polyunsaturated fatty acids in patients with chronic heart failure, where the use of the CE of death and admission to hospital for cardiovascular reasons as co-PE would be fully supported by the ARE.

### Recommendations for Cardiovascular Clinical Trials

We present recommendations for future design choice between RE and CE for cardiovascular clinical trials that use CEs as an option for the PE, include death as the RE, and add other nonfatal end points, such as MI, hospitalization. We discuss the recommendations in terms of the values of the anticipated hazard ratios HR_R and HR_A, and, when needed, in terms of the anticipated probabilities of occurrence P_R and P_A. These guidelines have been based on the scenarios explored by Gómez and Lagakos and on the 43 clinical trials of the 47 (Table 2) having death (observed control group frequency ≤15%) as RE and stroke, MI, hospitalization, and target vessel revascularization as AE. Table 3 shows the observed relative frequencies and relative treatment effects of death and the AEs and Table 4 the possible pairs (P, HR) for RE and AE, after excluding pairs with HR>1 (17 of 43).

Keeping in mind that the specific decision for a given trial has to be based on a thorough study as has been shown in the

### Table 3. Summary of Observed Relative Frequencies and Relative Treatment Effects Among Clinical Trials With Observed Frequency of Death >20%

<table>
<thead>
<tr>
<th>End Point</th>
<th>Relative Frequency</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min, % Median, %</td>
<td>Max, % Min, Median, Max</td>
</tr>
<tr>
<td>Death</td>
<td>0.2 3.8 15</td>
<td>0.1 0.83 0.98</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>0.2 3.7 11.3</td>
<td>0.35 0.78 0.92</td>
</tr>
<tr>
<td>Stroke</td>
<td>0.4 2.2 4.7</td>
<td>0.52 0.83 0.89</td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>0.3 3.6 31</td>
<td>0.59 0.75 0.94</td>
</tr>
<tr>
<td>TVR</td>
<td>0.7 7.3 16.2</td>
<td>0.79 0.79 0.83</td>
</tr>
</tbody>
</table>

Hazard ratios restricted to clinical trials with HR<1. TVR indicates target vessel revascularization.

### Table 4. Chosen Pairs of Values of (P, HR) for RE (Death), and AE (Stroke, MI, Hospitalizations, and TVR) used for the recommendations

<table>
<thead>
<tr>
<th>P-Value: Probability of Occurrence in the Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR 0.005 0.03 0.06 0.09 0.12 0.15 0.18 0.27 0.39 0.48</td>
</tr>
<tr>
<td>0.3 AE ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.4 RE ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.5 AE RE ... RE ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.6 AE ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.7 ... RE RE RE ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.8 AE RE RE RE ... RE ... AE ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.9 AE AE AE AE AE ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
<tr>
<td>0.99 AE AE AE AE AE ... ... ... ... ... ... ... ... ... ... ... ... ...</td>
</tr>
</tbody>
</table>

P stands for the anticipated probability of occurrence in the control group and HR the corresponding hazard ratio. Clinical trials with HR=1 are excluded. AE indicates additional end point; and RE, relevant end point.
case studies and the Results section of this article, a set of recommendations on whether to use the RE or the CE is outlined below (Figures 7 and 8):

- **HR_A>HR_R**: the relative treatment effect is greater on the AE than on the RE→CE should always be used.
- **HR_A=HR_R**: RE and AE have approximately the same relative treatment effect→CE should almost always be used. Only in those cases where the anticipated probability for AE has a low frequency (P_A ≤ 0.06) and the frequency for RE is between 2 and 5 times the frequency of the other end points (2<P_R/P_A<5), CE could be a better choice.
- **HR_A<HR_R+0.1**: AE has a slightly smaller effect on treatment than RE→RE should always be used if P_R/P_A≥3 and CE should always be used if P_R/P_A≤0.25. Whenever 0.25<P_R/P_A<3 the decision will depend on the anticipated values of the relative treatment effect, the frequency of observation of either end point along with its correlation and to a lesser extent on the shape of the marginal density.
- **HR_A<HR_R+0.2**: AE has a smaller effect on treatment than RE→RE should almost always be used except when the relative frequency of the AE is extremely higher than that of the RE (P_R/P_A≤0.06).
- **HR_A<HR_R+0.3**: AE has a much smaller effect on treatment than RE→RE should always be used.
- **HR_A close to 1 and P_A<0.005**→RE should always be used.

One has also to keep in mind that the association between time to RE and time to AE could play an important role (ARE decreases when the correlation between the 2 end points increases) and that decisions based on hazard plots as the ones in Figures 2 and 3 are recommended (ARE decreases when the relative effect of treatment on the AE is smaller). Furthermore, the recommendations are to be taken cautiously because frequent events (P in the order of 0.005), frequencies of death with order of magnitude larger than the frequency of AE (P_R/P_A>12), and unlikely frequent end points (P>0.35) could reverse the direction of the recommendation.

**Discussion**

The use of composite PE in cardiovascular randomized trials has been addressed by many authors who have discussed, among other issues, the suitability of components that are clinically less important and the difficulties in interpreting results.

Our new approach helps the trialist, in the design of a future trial, to choose in an objective manner between candidates of PE, by computing the ARE based on the anticipated values of the control group frequency and HR of each candidate end point.

It is clear that in the cardiovascular context, the CEs under consideration overwhelmingly include a terminal event either as a RE or as an AE. This article explores under which circumstances adding other end points to a RE of death would result in a more efficient choice. It is clear from our results that, contrary to a common belief, adding a frequent event to a RE of death does not always help and, indeed, may even prove harmful. The fact that the CE increases the number of events, does not mean, even in the case of a common event rate and similar magnitude of the treatment effects, that the required sample size of a trial is reduced because, depending on the strength of the association between RE and AE, the ARE is not necessarily >1.

It is important to point out that our methodology is intended for the planning phase of the RCT. The reader should be aware of the presence of competing risks and how the analysis should appropriately take care of this issue. Chi5 describes how to properly analyze a RCT based on CE. They recommend to use all-cause mortality instead of cause-specific mortality to prevent from informative censoring, and although not strictly necessary if the CE is valid, to analyze separately the individual components, and to gain a more accurate assessment and interpretation of the clinical benefits and risks involved. They propose 2 basic formats for the presentation of trial data and the results of the analysis.

Finally, although the ARE method has been developed with a RCT in mind, well-planned observational studies, viewed as conditionally randomized experiments could take advantage of an appropriately adjusted version of the ARE method. Recommendations about reporting completely and accurately an observational study have been developed by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) initiative.23 The explanation of

**Figure 7.** The horizontal axis represents the values of the HR_A of additional end point (AE) as a function of the HR_R of relevant end point (RE). Each tick summarizes several scenarios, corresponding to different shapes of the marginal hazards and different degree dependences between RE and AE. For each tick, we indicate whether it is advisable to adopt Composite end point (CE) in preference to RE. See explanation in text for scenarios with CE* and RE*.

**Figure 8.** Summary of recommendations as a general guide for using composite or relevant end points (RE) as primary end point in cardiovascular clinical trials. AE indicates additional end point; and HR, hazard ratio.
how the study size was arrived at is among the requirements included in the STROBE Statement Checklist.24 Admittedly, the importance of sample size determination in observational studies depends on the context. When planning a new study, formal, a priori calculation of sample size is useful, especially for studies that will gather new data and will be planned for the purpose to overcome potential problems with previous reports. There is even a call for registration of observational studies on a World Health Organization-compliant registry before they begin to lend greater credibility to the study findings.25 In the case of large, hypothesis-driven cohort studies, there is no doubt that a solid protocol, including sample size and power justification, is required, and in that context, the ARE method is as useful for the informed choice of the end point as for any well-designed RCT.

As a conclusion, if a well-defined experiment is conducted and if the censoring patterns of both groups can be considered similar, the ARE method could be a valid option to discriminate between a RE and a CE.

Sources of Funding

This work is partially supported by grants MTM2008-06747-C02-01 from the Ministerio de Ciencia e Innovación and MTM2012-38067-C02-01 from the Ministerio de Economía y Competitividad, Spain.

Disclosures

None.

References

Informed Choice of Composite End Points in Cardiovascular Trials
Guadalupe Gómez, Moisés Gómez-Mateu and Urania Dafni

Circ Cardiovasc Qual Outcomes. 2014;7:170-178; originally published online January 14, 2014;
doi: 10.1161/CIRCOUTCOMES.113.000149
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272
Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circoutcomes.ahajournals.org/content/7/1/170

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/