Recommendations to choose the primary endpoint in cardiovascular clinical trials

Guadalupe Gómez Melis¹, Moisés Gómez-Mateu¹, Urania Dafni²

¹Dept. of Statistics and Operations Research
Universitat Politècnica de Catalunya
Jordi Girona 1–3, 08034 Barcelona, Catalonia (Spain)
²University of Athens, Hellas, Athens (Greece)
Report DR 2013/08
December 2013
Report
Recommendations to choose the primary endpoint in cardiovascular clinical trials

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

* From the Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain; and
** University of Athens, Hellas, Athens, Greece.

Correspondence to Guadalupe Gómez, PhD, Jordi Girona 1-3, 08304, Barcelona. E-mail lupe.gomez@upc.edu.
Abstract

**Background** – A composite endpoint is often used as the primary endpoint to assess the efficacy of a new treatment in randomized clinical trials (RCT). In cardiovascular trials, the often rare event of the relevant primary endpoint (individual or composite), such as cardiovascular death (CV death), Myocardial Infarction (MI), or both, is combined with a more common secondary endpoint, such as target lesion revascularization, with the aim to increase the statistical power of the study.

**Methods** – Gómez and Lagakos developed statistical methodology to be used at the design stage of a RCT for deciding whether to expand a study relevant primary endpoint $\varepsilon_1$ to $\varepsilon^*$, the composite of $\varepsilon_1$ and a secondary endpoint $\varepsilon_2$. The method uses the asymptotic relative efficiency of the logrank test for comparing treatment groups based on $\varepsilon_1$ versus the logrank test based on $\varepsilon^*$. The method is used to assess, in the cardiovascular research area, the characteristics of the candidate individual endpoints that should govern the choice of using a composite endpoint as the primary endpoint in a clinical trial.

**Results and conclusions** – A set of recommendations is provided based on the reported values of the frequencies of observing each candidate endpoint as well as on the magnitude of the effect of treatment as expressed by the hazard ratio, supported by cardiovascular RCTs published in 2008.

**Key words:** Asymptotic Relative Efficiency; Composite outcome; Logrank test; Cardiovascular; Randomized Clinical Trial
Introduction

The conclusions of a clinical trial rely heavily on its primary endpoint, and thus, at the design phase is of outmost importance that the most appropriate choice of primary endpoint is made. Composite endpoint (CE) is an event that is considered to have occurred if any one of several different events or outcomes (components) is observed\textsuperscript{1,2}. CEs are nowadays used commonly as the primary endpoint 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 (CV death) or not, such as Myocardial Infarction (MI), stroke and hospitalization. One could argue that the aim of using a CE is to address all efficacy measures deemed relevant to the success of a new treatment without the limitations imposed by multiplicity and competing risks problems. Nevertheless, one of the main objectives is to increase the power to detect a significant benefit induced by the new treatment. This increase, in the case of time-to-event endpoints, is expected to be achieved by the inclusion of component endpoints that occur with higher frequency and/or earlier than the main events of interest\textsuperscript{3}. However, adding less specific components might in fact lead to loss of power to detect the true treatment differences. In addition, improvement in the composite does not necessarily translate to an improvement in the relevant component (e.g., overall survival). The use of a CE is furthermore intricate due to the possibly big difference in the relative importance of the components as well as in the respective magnitude of the treatment effect.
Gómez & Lagakos⁴, provided a methodology to reach an informed decision regarding the primary endpoint at the design stage of a clinical trial. In the current manuscript, an illustration of this statistical methodology is presented and used to provide guidelines for the informed choice of primary endpoint in the context of cardiovascular clinical trials.

**Statistical Method**

Consider a two-arm RCT involving random assignment either to an active treatment or to a control treatment. We have a study relevant endpoint (RE), $\varepsilon_1$, that could be used as the primary endpoint for efficacy and a secondary endpoint which could be viewed as an additional endpoint of interest (AE), $\varepsilon_2$. For example, assume RE is the composite of cardiovascular death and MI, ($\varepsilon_1$), while AE is target lesion revascularization ($\varepsilon_2$). We consider the CE $\varepsilon^*$ of $\varepsilon_1$ and $\varepsilon_2$. The individuals are followed from randomization until the event of interest, or until the end of study, whichever occurs first.

To make an informed decision on whether the RE $\varepsilon_1$ or the CE $\varepsilon^*$ should be the primary endpoint, Gómez and Lagakos⁴ develop a strategy based on the behavior of the asymptotic relative efficiency of the logrank test for comparing treatment groups with respect to $\varepsilon_1$ versus the logrank test with respect to $\varepsilon^*$, denoted by $\text{ARE}(\varepsilon^*, \varepsilon_1)$. The computations for the ARE depend on a) whether or not the two endpoints of interest include a terminal event (death), b) the probabilities $p_1$ and $p_2$ of observing events $\varepsilon_1$ and $\varepsilon_2$, respectively, for the control group, c) the treatment effect with respect to $\varepsilon_1$ and $\varepsilon_2$ given by the hazard ratios $HR_1$ and $HR_2$, and d) the correlation between the times to event $\varepsilon_1$ and $\varepsilon_2$. 
Gómez and Lagakos propose as a general rule to use the CE instead of the RE if $\text{ARE}(\varepsilon^*, \varepsilon_1) > 1.1$; and to retain the RE, if $\text{ARE}(\varepsilon^*, \varepsilon_1) \leq 1.1$.

**Composite Endpoints in Cardiovascular Research**

Freemantle 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. In a meta-analysis exploring the use of CE in cardiovascular research, 114 interventional RCTs, almost half of the total cardiovascular trials examined, were identified to use a CE as the primary trial endpoint. In the conclusions, it is stated that the use of CE is often complicated by the magnitude of the effect of treatment across component endpoints as well as by the relative importance of the different components for the patients. Furthermore, it is discussed how higher event rates and larger treatment effects associated with less important components may lead to a misleading impression of the treatment effect.

**Material and Methods**

The extent of use of CEs in the recent literature was explored through a systematic Medline search covering the 2008 year publication of RCTs in 6 high impact medical journals (see 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 to 216 publications. From these, 87 mentioned in the abstract, title or keywords, a composite or combined endpoint, or the specific endpoints of MACE (Major Adverse Cardiac Events), or NACE (Net Adverse Clinical
Events). Papers that dealt with other diseases, looked at subgroup or nonrandomized comparisons or did not use time to event endpoints were excluded. A total of 61 clinical trials, were considered for exploring the use of a CE. The breakdown by journal is presented in Table 1. See the complete reference list in the appendix.

<table>
<thead>
<tr>
<th>JOURNAL</th>
<th>Total papers</th>
<th>%</th>
<th>CE RCT</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEJM</td>
<td>46</td>
<td>21%</td>
<td>17</td>
<td>28%</td>
</tr>
<tr>
<td>THE LANCET</td>
<td>36</td>
<td>17%</td>
<td>13</td>
<td>21%</td>
</tr>
<tr>
<td>European Heart Journal</td>
<td>54</td>
<td>25%</td>
<td>12</td>
<td>20%</td>
</tr>
<tr>
<td>CIRCULATION</td>
<td>53</td>
<td>25%</td>
<td>10</td>
<td>16%</td>
</tr>
<tr>
<td>JAMA</td>
<td>24</td>
<td>11%</td>
<td>9</td>
<td>15%</td>
</tr>
<tr>
<td>Annals of Internal Medicine</td>
<td>3</td>
<td>1%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Total RCT</strong></td>
<td><strong>216</strong></td>
<td><strong>100%</strong></td>
<td><strong>61</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Table 1: Summary of Medline search, for Cardiovascular terms, for 2008 publications of Randomized Clinical Trials (RCT). CE stands for Composite Endpoint.

A CE was used as primary endpoint 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 cases that CE is the primary, 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), while death is encountered in all of them but one (98%). In addition, among the 14 trials with an individual primary endpoint, in 13 of them death is either the relevant (in 4) or used as an additional endpoint (in 9).
It is thus clear that in the cardiovascular context, the CEs under consideration overwhelmingly include a terminal event either as a relevant or as an additional endpoint. Two clinical trials are used as case studies in the next section: the first, with death as component of the primary CE, and the second with death as component of the secondary CE.

**Case Study 1: Treating patients after an acute coronary syndrome with succinobucol**

A 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 primary endpoint, denoted by $\varepsilon^*$, the composite of $\varepsilon_1$ (time to first occurrence of cardiovascular death, resuscitated cardiac arrest, MI, stroke), and $\varepsilon_2$ (unstable angina or coronary revascularization) (see Figure 1) $^6$. A total of 6144 patients having experienced an acute coronary syndrome up to
one year before recruitment, were randomized to receive succinobucol (n=3078) or placebo (n=3066), in addition to standard of care. The arguably more important components of the primary CE, denoted by $\varepsilon_1$, are cardiovascular death, resuscitated cardiac arrest, MI, or stroke, comprising one of three secondary CEs. A beneficial effect of succinobucol on $\varepsilon_1$ was found [207 events: succinobucol vs 252 events: placebo; hazard ratio (HR) = 0.81, $p = 0.029$]. The less important but frequent outcomes, i.e., hospitalization for unstable angina and coronary revascularization, denoted by $\varepsilon_2$, were included in the primary CE $\varepsilon$. The expectation would be that by the inclusion of these outcomes, the resulting increase in the number of CE events observed would lead to an increase in study power. On the contrary, these endpoints did not differ significantly between the two treatment groups, and their contribution of a high relative number of events in the primary CE (64%), led to the disappearance of the statistically significant benefit of the active treatment on the important outcomes $\varepsilon_1$. Thus, the primary CE, $\varepsilon$, was not found to be significantly different between treatment groups (530 events: succinobucol vs 529 events: placebo).

Figure 1: Composite Endpoint $\varepsilon$ as the union of the Relevant endpoint ($\varepsilon_1$) and the Additional Endpoint ($\varepsilon_2$) in Tardif’s RCT.
Using the notation we have introduced in the previous section, we have that the probability of observing the RE $\epsilon_1$, is $p_1 = 8.2\%$ with observed $HR_1 = 0.81$, while the probability of observing the AE $\epsilon_2$, is $p_2 = 10.4\%$ with $HR_2 = 1.05$ (it corresponds to coronary revascularization, while $HR$ for unstable angina is 1.10).

The ARE($\epsilon_2$, $\epsilon_1$) 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 less than 1.1. Following the rule of Gómez and Lagakos, the benefits of using the CE, $\epsilon_2$, over the RE, $\epsilon_1$, are marginal and probably too small to justify adding $\epsilon_2$.

The use of $\epsilon_2$ would be justified in the case that $HR_2 \leq 0.85$, for all other parameters fixed (i.e., $p_1 = 8.2\%$, $HR_1 = 0.81$, $p_2 = 10.4\%$) (see Figure 2). However, if $HR_2 \geq 0.95$ not even an expected frequency of 20\% for the AE $\epsilon_2$, would justify the use of $\epsilon_2$. If $HR_2=0.9$, $\epsilon_2$ would only be justified if $p_2 \geq 20\%$ and the association between $\epsilon_1$ and $\epsilon_2$ is very weak (not shown).

Figure 2: ARE of composite versus relevant endpoint for a range of Spearman correlation coefficients and different values of $HR_2$ for the parameters of Case Study 1 ($p_1 = 0.082$, $HR_1 = 0.81$, $p_2 = 0.104$) and marginal increasing hazards.
Thus, under these circumstances, the additional components of coronary revascularization or hospitalization for unstable angina on the primary endpoint 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. Based on these findings one should be cautious of adding components to the primary endpoint of relatively little importance.

Case Study 2: Treating haemorrhagic complications during primary percutaneous coronary intervention (PCI) in acute myocardial infarction

The Harmonizing Outcomes with Revascularization and Stents in Acute Myocardial Infarction (HORIZONS-AMI) study is a prospective, open label, randomized, multicenter trial in patients with ST-segment elevation myocardial infarction presented within 12 hours after the onset of symptoms. In this study, 3602 patients were assigned to treatment with heparin plus a glycoprotein IB/IIa 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 endpoints were prespecified: 1) major bleeding, denoted by \( \varepsilon_1 \) and 2) net adverse clinical events (NACE), denoted by \( \varepsilon^* \), a composite of major bleeding and major adverse cardiovascular events (MACE). MACE, denoted by \( \varepsilon_2 \), is composed, in this trial, of death, reinfarction, target vessel revascularization for ischaemia and stroke. In this case, while major bleeding is the relevant event of interest, the composite \( \varepsilon^* \) takes into account all other additional adverse clinical events including death. According to the results, MACE is almost identical in the two groups (98 vs 99 events, \( p = 0.95 \)) while major bleeding is statistically significantly lower in the
bivalirudin alone group (89 vs 149 events, p<0.001). The comparison of \textit{NACE} (166 vs 218 events, \( p = 0.005 \)) between treatment groups is found statistically significant, and as correctly mentioned by the authors, this is entirely driven by the effect on \textit{major bleeding}. The risk taken by the researchers of combining the endpoint of interest with an endpoint on which treatments have no differential effect, is demonstrated using this study.

The probability of observing a \textit{major bleeding event}, \( \varepsilon_1 \), is \( p_1 = 8.3\% \) with hazard ratio \( HR_1 = 0.6 \), while the probability of observing a \textit{MACE} event, \( \varepsilon_2 \), is \( p_2 = 5.5\% \) with \( HR_2 = 1 \). \textit{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(\( \varepsilon^*, \varepsilon_1 \)) is examined 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 \textit{major bleeding event} and a \textit{MACE} event, is less than 1.1, meaning that the use of the CE (\textit{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 \textit{MACE} event, \( \varepsilon_2 \), \( 5.5\% \leq p_2 \leq 8.0\% \), a similar pattern emerges, with a sparse number of cases (5 of 84 scenarios), with ARE above 1.1, leading to the recommendation of \textit{NACE}, with all cases occurring for correlation of 0.75. Second, the ARE(\( \varepsilon^*, \varepsilon_1 \)) was also explored for larger beneficial effects on \textit{MACE} \( 0.3 \leq HR_2 \leq 0.9 \) and the ARE value is less than 1.1 except for 11 cases (out of 105). Figure 3 illustrates the AREs for the values of the parameters of this clinical trial \( (p_1 = 8.3\%, \ HR_1 = 0.6, \ p_2 = 5.5\%) \) and for marginal increasing hazards. Globally, in 88\% of the scenarios, the use of \( \varepsilon_1 \) is recommended.
It is clear that the chosen primary endpoint $\varepsilon^*$ 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 $NACE$ since the ARE can be as low as 0.2 if the beneficial effect on $MACE$ is 0.5, meaning that major bleeding as a primary endpoint 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.
Results

The absolute number and frequency of occurrence of each endpoint (CE and components of interest), the corresponding hazard ratios (HRs) and p-values between groups compared in the trial, were extracted from each publication in Table 1. The choice between CE and RE in the design of future trials on similar populations and questions of interest, can be guided by the estimated ARE based on the observed values of endpoint control group frequency and HRs from the reviewed trials. Interesting cases of trials leading to a significant result for the CE while non-significant for the RE or significant for the RE and non-significant for the CE, are described later in this section.

Reviewed superiority trials leading to non-significant differences both on the RE and CE\textsuperscript{8-11}, are not of particular interest, since at the design stage for any future superiority trial, the anticipated HRs under investigation will differ from one.

In addition, trials that led to opposite effects in the RE and AE, will not be explored in this context, since in such cases, designing the study based on the CE is not useful. An illustration of why such a choice would be problematic is given by a study exploring the effect of metoprol succinate in patients undergoing non-cardiac surgery with (or at risk of) atherosclerotic disease\textsuperscript{12}. In that trial, the primary CE, $\varepsilon_1$, (cardiovascular death, non-fatal MI, and non-fatal cardiac arrest) shows a treatment benefit ($HR_1=0.84$), driven entirely by the AE (MI), $\varepsilon_2$, ($p_2=5.7\%; HR_2=0.73$), while death ($\varepsilon_1$) shows a harmful effect ($p_1=2.3\%; HR_1=1.33$). The interpretation of the results and the study conclusion, as would be expected, rather than focusing on the CE significant result, is instead dramatically affected by the observed harm on survival.
In the reviewed studies, specific combinations of the control group frequencies for the RE($p_1$) and AE($p_2$) with corresponding $HR$ values, emerged. The AREs were estimated for these combinations to serve as a guide for the design of future trials.

*Death from Cardiovascular causes or death from any cause* was included in the composite primary endpoint in 46 trials. When *death* is the RE, the ARE of a CE including any of *MI, Stroke or hospitalization* as AEs, was explored for different shapes of time to event distributions, and a range of correlations between times to $\varepsilon_1$ and $\varepsilon_2$.

For the majority of trials, the frequency of *death* was relatively low (median 4%), with the exception of 3 trials where death was very frequent (above 20%). For observed frequencies up to 12%, it is reasonable to attempt to use a CE and examples of the trials are discussed next.

For the relatively low frequency of *MI* ($\varepsilon_2$) (up to 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* ($HR_2>HR_1$).

For particularly low frequency of *stroke*, $\varepsilon_2$, 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 ($HR_2<HR_1$). The same is true for the higher frequency of *stroke* (12%), while the CE is also justified when the beneficial effect on *stroke* is slightly less than on *death* but *death* presents with lower frequency.

When examining the CE of *death and hospitalization*, in the cases that *hospitalization* is occurring with high frequency (18% to 48%), a different pattern emerges. The CE is justified in the cases that the $HR$ for *death* is above 0.8, while for *hospitalization* is below 0.9. Even in the case of $HR_1=0.70$, for very low frequency of
death ($p_1 = 3\%$), when the frequency of hospitalization is high ($p_2 = 48\%$), the composite is not only justified for $HR_2 = 0.70$ as would be expected, but also for $HR_2 = 0.9$. For a substantial benefit on death coupled with low frequency ($HR_1 = 0.5; p_1 = 6\%$), when the frequency of hospitalization is high ($p_2 = 39\%$) even for a smaller benefit for hospitalization ($HR_2 = 0.70$), the CE is justified, while when the benefit on hospitalization becomes even smaller ($HR_2 = 0.90$), using death as a primary endpoint is preferred. The CE is not justified when $HR_2 > HR_1$ provided that the frequency of death is higher ($p_1 = 12\%$).

Discussion of specific trials
In total, 5 clinical trials used as primary endpoint the CE of death and MI, with frequency of the CE ranging from 5% to 18%\textsuperscript{10-14}. The corresponding follow-up period for 3 of the trials was 30 days: for patients with ST-segment elevation acute MI (STEMI) without reperfusion therapy\textsuperscript{13}, for intermediate to high risk patients with coronary artery bypass graft surgery\textsuperscript{10}, and for patients undergoing non-cardiac surgery with (or at risk of) atherosclerotic disease\textsuperscript{12}. For stable patients after acute phase of MI with late occluded infarction related arteries\textsuperscript{11} and for hypertensive patients\textsuperscript{14}, the follow-up period was 4 and 5.5 years, respectively.

In the clinical trial testing fondaparinux in STEMI patients\textsuperscript{13}, the RE of death and the AE of Myocardial re-infarction at 30 days occurred in 12.5% and 3.7% of control patients, respectively. The CE occurred in 15.1% of control patients, indicating a weak correlation between RE and AE. The corresponding hazard ratios ($HR_1 = 0.83$ and $HR_2 = 0.66$) 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 $HR^*$ of 0.80. In this trial, the use of the CE, is clearly indicated by the ARE in 100% of the simulations.

In testing atorvastatin in hypertensive patients\textsuperscript{14}, a statistically significant benefit on the CE of \textit{fatal coronary heart disease (CHD)} and \textit{non-fatal MI} was found ($p^*=4.8\%, \; HR^*=0.64$), while a trend was identified in the RE of \textit{cardiovascular mortality} ($p_1=3.2\%, \; HR_1=0.84$). Use of the CE would be justified here for the AE of \textit{non-fatal MI}, since the ARE is larger than 1.1 for i) $HR_2$ less than 0.8 and any $p_2$ up to 0.6, ii) $p_2$ larger than 13\% and any $HR_2$ as high as 0.9.

Three prevention studies assessed the benefit on the risk of cardiovascular disease of i) vitamins E & C in men above 50 years old\textsuperscript{15}, ii) intensive glucose control in veterans with type 2 diabetes\textsuperscript{16}, and iii) low-dose aspirin in the prevention of atherosclerotic events in patients with type 2 diabetes\textsuperscript{17}. Composite primary outcome was used in each of these studies, with a \textit{major cardiovascular event} defined as \textit{nonfatal MI, nonfatal stroke, death from CV}\textsuperscript{15}, or MI, stroke, death from CV, \textit{congestive heart failure}, \textit{surgery for vascular disease, inoperable coronary disease, and amputation for ischemic gangrene}\textsuperscript{16}, while the CE of \textit{atherosclerotic events} was defined as \textit{fatal and nonfatal ischemic heart disease, fatal or nonfatal stroke, and peripheral arterial disease}\textsuperscript{17}.

The latter trial, could be considered an outlier, due to the combination of a very low frequency of \textit{fatal CV events} ($p_1=0.008$), yet significantly different between groups ($HR_1=0.10, \; p=0.0037$). Under these extreme conditions, the use of the RE would have been justified based on the very low $HR_1$, while the use of the CE would have been justified based on the very low frequency of events. The ARE points to the clear choice of the CE for $HR_2\leq 0.2$, and the clear choice of the RE for $HR_2\geq 0.8$, while for $HR_2$ values between 0.2 and 0.8, the CE is recommended as $HR_2$ increases for
progressively higher values of $p_2$. The CE occurred in 6.7% of control patients, indicating a weak correlation between RE and AE, leading to an $HR^*$ of 0.80, but not statistically significant. Nevertheless, an assumption of a treatment effect at such an extreme would be difficult to justify at the design stage, although it could be taken under consideration for the next trial designed on this question.

Finally, in only 4 trials, death from CV or death from any cause was used as the individual primary or co-primary endpoint. The frequency of cardiovascular death or any death in two of the trials on patients with NYHA class II-IV Chronic Heart Failure, or Atrial Fibrillation & NYHA 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 CV reasons as co-primary endpoint, would be fully supported by the ARE.

On the contrary, for a particularly low frequency of death ($p_1<1\%$) and moderate or low magnitude of benefit, as already discussed, the CE is almost always justified. Surprisingly, in the Influenza Vaccination study on patients with Confirmed Coronary Artery Disease (CAD), the primary endpoint used was death from CV, although its frequency was $<1\%$. A significant benefit of vaccination was detected on the secondary CE of Coronary Ischemic event (MACE and hospitalization for myocardial ischemia). Use of this CE would have been recommended at the design stage by the calculation of ARE.
Recommendations

We present recommendations for future design choice between CE and RE for cardiovascular clinical trials which use CEs as an option for the primary endpoint, include death as the RE and add other non fatal endpoints such as MI, stroke, hospitalization, etc. We are focusing on the 44 clinical trials out of the 47 in Table 2 having death as part of the primary CE and observed frequency of death less than 15%.

In all cases, computations have been done modeling the marginal laws of the times to $\varepsilon_1$ and $\varepsilon_2$ as Weibull, representing decreasing, constant and increasing hazard functions. Previously examined scenarios for this situation (Case 3, Gómez and Lagakos), are combined with the parameter combinations encountered in the current report to provide recommendations for the cardiovascular area trials.

Gómez and Lagakos reproduced several frequency situations by taking probabilities $p_1$ and $p_2$ equal to 0.05, 0.15, 0.30 and 0.50. The relative treatment effect on the RE was set to $HR_1=0.5$ or 0.7, and it was combined with six different values for $HR_2$ reproducing situations where the beneficial effect on the AE was larger, the same or smaller. These parameter values were combined with different degrees of dependence between times to $\varepsilon_1$ and to $\varepsilon_2$. Based on all these combinations, it is observed that the ARE decreases when the correlation between the two endpoints increases and when the relative effect of treatment on the AE is smaller. The observed relative frequencies of death among the 44 studied clinical trials were between 0.002 and 0.15. The observed relative frequencies of the AEs (MI, Stroke, Hospitalization and TVR) were between 0.002 and 0.48. Concerning the relative treatment effects it was found that some of the RCTs had an observed $HR$ larger than 1 (19 out of 44). At the design stage of a clinical trial an anticipated value of $HR>1$ will
not be a value of interest neither for death nor for the AEs. For computations we are using \(0.99\) to represent negligible relative treatment effects. Among the clinical 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.

Since not all the combinations of frequencies and relative treatment effects \((p_1, HR_1)\) or \((p_2, HR_2)\) were found in the studied RCTs, we did restrict our computations to published pairs of values \((p, HR)\). Figures 4 and 5 reproduce the possible pairs \((p, HR)\) for death and for Stroke, MI, Hospitalizations and TVR, extracted from the 44 clinical trials, after excluding pairs with \(HR > 1\). They represent the range of combinations which have been used to compute the ARE.

Figure 4: Plot for the probability of occurrence of Death in the control group versus the corresponding hazard ratio. Clinical trials with \(HR > 1\) are excluded. Chosen pairs of values for the computations are: \((p_1, HR_1) = (0.005, 0.4), (0.05, 0.99), (0.03, 0.7), (0.03, 0.8), (0.03, 0.99), (0.06, 0.5), (0.06, 0.7), (0.06, 0.8), (0.06, 0.9), (0.06, 0.99), (0.09, 0.7), (0.09, 0.9), (0.12, 0.8), (0.12, 0.9), (0.12, 0.99), (0.15, 0.8).
Figure 5: Plot for the probability of occurrence of Stroke, MI, Hospitalizations and TVR in the control group versus the corresponding hazard ratio. Clinical trials with $HR > 1$ are excluded. Chosen pairs of values for the computations are: $(p_2, HR_2) = (0.005, 0.3), (0.005, 0.5), (0.005, 0.6), (0.005, 0.99), (0.03, 0.6), (0.03, 0.9), (0.03, 0.99), (0.06, 0.7), (0.06, 0.8), (0.06, 0.99), (0.12, 0.8), (0.12, 0.9), (0.12, 0.99), (0.18, 0.9), (0.27, 0.9), (0.39, 0.7), (0.48, 0.9)$. The three last pairs not shown in the figure.

Recommendations for cardiovascular clinical trials

Aiming to provide a useful guide to help the investigator in the planning of a clinical trial, we discuss the recommendations in terms of the values of the anticipated hazard ratios $HR_1$ and $HR_2$, and, when needed, in terms of the anticipated probabilities of occurrence $p_1$ and $p_2$.

It would never be recommended to use a CE by adding an AE with anticipated frequency value as small as 0.005 and/or an $HR$ close to 1. One has also to keep in mind that the association between time to $\epsilon_1$ and time to $\epsilon_2$ could play an important role and that decisions based on hazard plots as the ones in Figures 2 and 3 are recommended. Furthermore, the recommendations are to be taken cautiously since very infrequent events ($p$ in the order of 0.005), frequencies of death with order of magnitude larger than the frequency of AE ($p_1/p_2 > 12$) and/or unlikely very frequent endpoints ($p$ larger than 0.35) could reverse the direction of the recommendation.
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 case studies and the results section, a set of recommendations on whether to use the RE or the CE is outlined below (see figure 6):

- $HR_2 < HR_1$: the relative treatment effect is greater on the AE than on the RE $\rightarrow$ CE should always be used.
- $HR_2 = HR_1$: RE and AE have approximately the same relative treatment effect $\rightarrow$ CE should almost always be used. Only in those cases where the anticipated probability for AE has a low frequency ($p_2 \leq 0.06$) and the frequency for RE is between 2 and 5 times the frequency of the other endpoints ($2 < p_1/p_2 < 5$), RE could be a better choice.
- $HR_2 = HR_1 + 0.1$: AE has a slightly smaller effect on treatment than RE $\rightarrow$ RE should always be used if $p_1/p_2 \geq 3$ and CE should always be used if $p_1/p_2 \leq 0.25$. Whenever $0.25 < p_1/p_2 < 3$ the decision will depend on the anticipated values of the relative treatment effect, the frequency of observation of either endpoint along with its correlation and to a lesser extent on the shape of the marginal density.
- $HR_2 = HR_1 + 0.2$: AE has a smaller effect on treatment than RE $\rightarrow$ RE should almost always be used except when the relative frequency of the AE is extremely higher than that of the RE ($p_1/p_2 \leq 0.06$).
- $HR_2 \geq HR_1 + 0.3$: AE has a much smaller effect on treatment than RE $\rightarrow$ RE should always be used.

Figure 6: The horizontal axis represents the values of the $HR_2$ of AE as a function of the $HR_1$ of 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 CE in preference to RE. * See explanation on text.
Discussion
This study, focused in the cardiovascular research area, explores under which circumstances adding other endpoints to a primary endpoint 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 primary endpoint 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 since, depending on the strength of the association between RE and AE, the ARE is not necessarily greater than 1.

Our methodology implicitly assumes the ARE as the reciprocal ratio of the sample sizes needed to attain the same power for a given significance level. Furthermore, the computation of the sample size is often based on the proportionality of the hazards across the two treatment groups based on the primary endpoint. It is important to emphasize that this assumption for both the RE and the AE does not imply the proportionality on the CE, hence alternative formulas for the computation of the sample size if the CE is chosen are needed.

We also assume that the dependence between the RE and the AE is specified by means of a Frank’s Archimedean copula, but also other copulas could be taken into consideration. Of note, at the analysis stage, the presence of competing risks, the homogeneity and the strength of association among the components of the CE is encouraged.

We have designed a platform called CompARE to calculate ARE values based on the information of the different relevant endpoints together with the anticipated values of $p_1$, $p_2$, $HR_1$, $HR_2$ and $\rho$. The design of CompARE is flexible enough allowing
different scenarios to be shown in plots by combining different range of values for the parameters. This is a free tool to be used to learn which is the most efficient primary endpoint among a set of given ones in an intuitive way. CompARE, written under the GNU/LGPL license, allows the user, through HTML forms from the web-based application, to introduce the parameter values. Such an interactive web site, still in a beta version and available from the second author under request, would allow users to enter their own values when designing a clinical trial.

Finally, it is important to point out that our methodology is only intended for the planning phase of the RCT.

**Sources of Funding**

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

**Disclosures**

None.
References


APPENDIX

List of References of the Cardiovascular Clinical Trials


