Selection of composite binary endpoints in clinical trials

Journal:	*Biometrical Journal*
Manuscript ID	bimj.201600229.R2
Wiley - Manuscript type:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Bofill Roig, Marta; Universitat Politecnica de Catalunya, Department of Statistics and Operations Research Gomez, Guadalupe; Universitat Politecnica de Catalunya, Department of Statistics and Operations Research
Keywords:	Asymptotic Relative Efficiency, Binary Endpoint, Clinical Trial, Composite Endpoint, Treatment Effects

Biometrical Journal **??** (2016) ??, zzz–zzz / DOI: 10.1002/bimj.200100000

Selection of composite binary endpoints in clinical trials

Marta Bofill Roig [1] and **Guadalupe Gómez Melis** [*,1]

[1] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain

Received zzz, revised zzz, accepted zzz

The choice of a primary endpoint is an important issue when designing a clinical trial. It is common the use of a composite endpoints as primary endpoint because it increases the number of observed events, captures more information and is expected to increase the power. However, combining events which have not similar clinical importance and have different treatment effects makes the interpretation of the results cumbersome and might reduce the power of the corresponding tests. Gómez and Lagakos proposed the ARE (Asymptotic Relative Efficiency) method to choose between a composite or one of its components as primary endpoint comparing the efficacy of a treatment based on the times to each of these endpoints. The aim of this paper is to expand the ARE method to binary endpoints. We show that the ARE method depends on six parameters including the degree of association between components, the event proportion and the effect of therapy given by the corresponding odds ratio of the single endpoints. A case study is presented to illustrate the methodology. We conclude with efficient guidelines for discerning which could be the best suited primary endpoint given anticipated parameters.

Key words: Asymptotic Relative Efficiency; Binary Endpoint; Clinical Trial; Composite Endpoint; Treatment Effects.

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1022/bimj.XXXXXXX (please delete if not applicable)

1 Introduction

Nowadays, randomized clinical trials guide the advance of medical knowledge. They are the most well-grounded procedure for evaluating the applicability of clinical research and also comparing the safety and effectiveness of a new intervention against the standard of care. The protocol formalizes the medical question and specifies the design of the trial. One key decision that has to be defined is the choice of the primary endpoint which measures the clinical evidence by quantifying the treatment effect. Sample size requirement, analysis and conclusions on efficacy are based on the primary endpoint.

Clinical trials often take into account two or more efficacy endpoints. If we use multiple co-primary endpoints, we could capture different attributes. Moreover, multiple co-primary endpoint might provide a better explanation about how the disease behaves under treatment and an improvement of the evaluation of whether there are the differences in the efficacy between different interventions. However, the use of multiple endpoints entails challenges in analyses and planning. On the one hand, we need a multiplicity adjustment for avoiding an inflation of the overall type I error (Pocock, et al. (1987)). On the other hand, multiple co-primary endpoints are usually correlated between them. Since the correlation affects parameters estimation and sample size calculation, we should correctly specify them, because if we define the endpoints as if they were not correlated, we will not achieve the desired power (Sozu, et al. (2010)).

One possible approach to deal with these challenges is to transform the multivariate problem into a univariate one combining several outcomes in a single summary indicator (Rauch, et al. (2014)). In this

*Corresponding author: e-mail: lupe.gomez@upc.edu

regard, it is common to use the combination of several responses into a unique variable, specially in the settings of time-to-event and binary outcomes. Then, the focus is on the time of first event between a set of possible adverse events which are assumed to be relevant to the disease progress, or, in the binary context, the composite collapses the information into a binary endpoint which takes value 1 if whenever one of the outcomes has occurred.

Composite endpoints are frequently chosen as primary endpoint in many health areas and specially in cardiovascular and oncology trials. There are three key advantages for using a composite. First, they avoid the need of multiplicity adjustment. Second, a composite endpoint contains more information on the course of the disease than a single endpoint providing a better explanation about the differences between treatment groups. Third, the increment of the number of observed events is expected to increase the power (Rauch, et al. (2014)).

The main drawback of using a composite endpoint is its interpretation since its components rarely have comparable clinical importance and similar treatment effects (Ferreira-González, et al. (2007)). Besides, a combination of different events changes the mean and also the variance of the response upon which the analysis is based (Lefkopoulou and Ryan (1993)). The addition of components which are not relevant enough could compromise the interpretation of results and reduce power. Hence, the choice of the particular components for the composite has a great importance in the design phase (Mascha and Sessler (2011)) and a deeper study about the meaning of the composite response is needed (Rauch and Kieser (2013)).

Legler, et al. (1995) and Lefkopoulou and Ryan (1993) presented a framework for comparing the performance of several tests based on multiple binary endpoints. They compare as well those tests to the test that results of collapsing the data into a composite. In the framework of survival analysis, Gómez and Lagakos (2013) proposed the Asymptotic Relative Efficiency (ARE) method to compare the efficiencies of two trial designs according to the chosen primary endpoint. The motivation lied in deciding between one relevant primary endpoint or the compound of this relevant and one additional endpoint as primary endpoint of the trial. Their methodology, referred to as the ARE method, provides a tool to quantify the improvement in efficiency of adding a secondary endpoint to the primary endpoint. However, the ARE method has not been studied for binary outcomes.

Assume a binary composite endpoint and also the most severe and relevant of its components. Aiming to provide statistical guidelines that would indicate when it is more efficient to use the composite endpoint over one of its components as the primary endpoint of the trial, we expand the ARE method proposed by Gómez and Lagakos to binary endpoints. We show that the ARE method for binary composite endpoint depends on six parameters including the degree of association between the components of the composite endpoint on each group, the event proportion, and the effect of therapy on each component.

The paper is structured as follows, first, we relate the parameters of the composite to the parameters of the components and the correlation between them. Then, we find a relationship between the odds ratio of the composite and the odds ratios of its components. After that, we define the extension of the ARE method for binary endpoints and we explain the applicability of the method. Next, a clinical trial is used to illustrate the use of the methodology. Finally, we present recommendation guidelines in order to assess in which cases a composite endpoint should be preferred because is more efficient than the most relevant of its components. A short discussion concludes the paper.

2 Notation and main assumptions

2.1 Binary endpoints

Consider a randomized clinical trial comparing two treatment groups, control group ($i = 0$) and treatment group ($i = 1$), each group composed of n_i patients and denoting by $n = n_0 + n_1$ the total number of patients. We assume two different binary endpoints of potential interest, ε_1 and ε_2. Let X_{ijk} denote the response of the k-th binary endpoint for the j-th patient in the i-th group of treatment ($i = 0, 1$,

$j = 1, ..., n_i$, $k = 1, 2$). The response X_{ijk} is defined by 1 if the event, ε_k, has occurred and 0 otherwise. Then, for all $j \in \{1, ..., n_i\}$, $i \in \{0, 1\}$, $k \in \{1, 2\}$, the event rates are defined as:

$$p_k^{(i)} \quad = \quad \mathrm{P}(X_{ijk} = 1) \; = \; 1 - q_k^{(i)}$$

where $p_k^{(i)}$ and $q_k^{(i)}$ are the probabilities that ε_k occurs or not, respectively, for a patient belonging to the i-th group of treatment.

We consider a binary composite endpoint of two components, $\varepsilon_* = \varepsilon_1 \cup \varepsilon_2$, defined as the event that occurs whenever one of the endpoints is observed. Moreover, we assume that there exists one endpoint which is more relevant for the scientific question than the other, with no loss of generality, consider ε_1 the relevant endpoint and ε_2 the additional one. Denote by X_{ij*} the composite response defined as a Bernoulli random variable of parameter $p_*^{(i)} = \mathrm{P}(X_{ij*} = 1) = 1 - q_*^{(i)}$, where

$$X_{ij*} \quad = \quad \begin{cases} 1, \text{ if } X_{ij1} + X_{ij2} \geq 1 \\ 0, \text{ if else } X_{ij1} + X_{ij2} = 0 \end{cases}$$

In order to quantify the differences in efficacy between the two treatments, we might use the odds ratio for each k-th endpoint defined as:

$$\mathrm{OR}_k \quad = \quad \frac{p_k^{(1)}/q_k^{(1)}}{p_k^{(0)}/q_k^{(0)}}$$

Hereafter, we assume that both the composite endpoint and the relevant endpoint could lead to answer the clinical question, that is, both might be used as the primary endpoint of the trial.

2.2 The relevant endpoint as primary endpoint

If we test the treatment effect on the relevant endpoint, we establish the following hypothesis test:

$$\mathcal{H}_1 : \quad \begin{cases} H_0 : \quad \log(\mathrm{OR}_1) = 0 \\ H_1 : \quad \log(\mathrm{OR}_1) < 0 \end{cases} \tag{1}$$

where the null hypothesis of no-treatment effect is stated as $\mathrm{OR}_1 = 1$ or equivalently $\log(\mathrm{OR}_1) = 0$ and the alternative hypothesis assumes a risk reduction of the relevant endpoint, then, a negative $\log(\mathrm{OR}_1)$. For addressing the problem, we consider the score test defined as:

$$T_{1,n} \quad = \quad \frac{\hat{p}_1^{(0)} - \hat{p}_1^{(1)}}{\sqrt{\frac{1}{n_0}\hat{p}_1^{(0)}\hat{q}_1^{(0)} + \frac{1}{n_1}\hat{p}_1^{(1)}\hat{q}_1^{(1)}}} \tag{2}$$

where $\hat{p}_1^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij1} \; = \; 1 - \hat{q}_1^{(i)}$, that is, the proportion of relevant events in the i-th group of treatment.

Under the null hypothesis, the score test asymptotically follows the standard normal distribution. Under contiguous alternatives of the form $H_{1,n} : \log(\mathrm{OR}_1)_n = \frac{v_1}{\sqrt{n}}$, where $v_1 < 0$, the score test is asymptotically normal with unit variance and mean δ_1, called non-centrality parameter of the test, given by:

$$\delta_1 \quad = \quad v_1\sqrt{p_1^{(0)}q_1^{(0)}\pi(1-\pi)} \tag{3}$$

where π denotes the proportion of patients allocated to control group, that is, $\pi = \lim_{n\to+\infty} n_0/n$.

For any finite sample size, $\log(\mathrm{OR}_1)_n = \frac{v_1}{\sqrt{n}}$ is the treatment effect assumed as alternative. The effects are determined by the sample size and the constant v_1 which is interpreted as the limiting treatment effect as sample size increases, i.e.:

$$\lim_{n\to+\infty} \sqrt{n}\log(\mathrm{OR}_1)_n \quad = \quad v_1$$

Since the contiguous alternative changes with n, it forms a sequence that converge to 0, that is, to the null hypothesis as $n \to +\infty$. Whereas the power of the score test under any fixed alternative goes to 1 as sample size goes to infinity, under contiguous alternatives the limiting power is strictly less than 1.

2.3 The composite endpoint as primary endpoint

If the treatment effect is evaluated at the composite endpoint, the hypotheses are defined by:

$$\mathcal{H}_* : \quad \begin{cases} H_0 : & \log(\mathrm{OR}_*) = 0 \\ H_1 : & \log(\mathrm{OR}_*) < 0 \end{cases} \tag{4}$$

where under the null hypothesis, we assume that there is not treatment effect on the composite endpoint and, under the alternative hypothesis, we state a reduction of the risk evaluated on the composite event.

Following the same procedure as above, the difference between treatment groups is tested by means of the score test, $T_{*,n}$, namely:

$$T_{*,n} \quad = \quad \frac{\hat{p}_*^{(0)} - \hat{p}_*^{(1)}}{\sqrt{\frac{1}{n_0}\hat{p}_*^{(0)}\hat{q}_*^{(0)} + \frac{1}{n_1}\hat{p}_*^{(1)}\hat{q}_*^{(1)}}} \tag{5}$$

The score test $T_{*,n}$ under the null hypothesis is asymptotically $N(0,1)$, and under a sequence of contiguous alternatives, $H_{*,n} : \log(\mathrm{OR}_*)_n = \frac{v_*}{\sqrt{n}}$, is asymptotically normal with unit variance and mean δ_* (non-centrality parameter) given by:

$$\delta_* \quad = \quad v_* \sqrt{p_*^{(0)} q_*^{(0)} \pi (1-\pi)} \tag{6}$$

3 Binary Composite Endpoint defined from the margins

3.1 Parameters

Bahadur's theorem (Bahadur (1961)) allows to determine the joint distribution of multiple correlated binary endpoints and shows that the joint distribution is uniquely determined by the marginal probabilities and the degree of association between the endpoints. As noted by Sozu, et al. (2010), the association degree among the correlated binary endpoints might be defined by different measures. We consider *Pearson's correlation coefficient* as the association measure between endpoints. Let $\rho^{(i)}$ be the correlation coefficient, also referred as phi coefficient, in the i-th group defined as:

$$\rho^{(i)} \quad = \quad \frac{p_\cap^{(i)} - p_1^{(i)} p_2^{(i)}}{\sqrt{p_1^{(i)} q_1^{(i)} p_2^{(i)} q_2^{(i)}}}$$

Note that the correlation coefficient is represented by the underlying probabilities and the overlap between these marginal events, expressed by $p_\cap^{(i)} = \mathrm{P}(X_{ij1} = 1, X_{ij2} = 1)$. Applying results from Bahadur (1961), the probability of the composite endpoint in the i-th group of treatment, $p_*^{(i)}$, is uniquely determined by the probabilities of the single endpoints, $p_1^{(i)}$, $p_2^{(i)}$, and the correlation between them, as follows:

$$p_*^{(i)} \quad = \quad 1 - q_1^{(i)} q_2^{(i)} - \rho^{(i)} \sqrt{p_1^{(i)} p_2^{(i)} q_1^{(i)} q_2^{(i)}}\,, \qquad i = 0, 1$$

The odds ratio of the composite endpoint, OR_*, is given in terms of the correlation between endpoints for each group, the event proportions in the control group given by the respective odds and the therapy effect given by the corresponding odds ratio, as follows:

$$\mathrm{OR}_* = \frac{\frac{\left(1+\frac{\mathrm{OR}_1 p_1^{(0)}}{1-p_1^{(0)}}\right)\left(1+\frac{\mathrm{OR}_2 p_2^{(0)}}{1-p_2^{(0)}}\right)-1-\rho^{(1)}\sqrt{\frac{\mathrm{OR}_1\mathrm{OR}_2 p_1^{(0)} p_2^{(0)}}{(1-p_1^{(0)})(1-p_2^{(0)})}}}{1+\rho^{(1)}\sqrt{\frac{\mathrm{OR}_1\mathrm{OR}_2 p_1^{(0)} p_2^{(0)}}{(1-p_1^{(0)})(1-p_2^{(0)})}}}}{\frac{\left(1+\frac{p_1^{(0)}}{(1-p_1^{(0)})}\right)\left(1+\frac{p_2^{(0)}}{(1-p_2^{(0)})}\right)-1-\rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{(1-p_1^{(0)})(1-p_2^{(0)})}}}{1+\rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{(1-p_1^{(0)})(1-p_2^{(0)})}}}} \tag{7}$$

The full derivation is to be found in Appendix A. Observe that OR_* depends on the following six parameters $(p_1^{(0)}, p_2^{(0)}, \mathrm{OR}_1, \mathrm{OR}_2, \rho^{(0)}, \rho^{(1)})$ and that the parameters associated to each component together with the correlation between them is what we only need to assess the effect on the composite endpoint.

A special property of binary endpoints is that the correlation takes values between two bounds which are defined according to the marginal probabilities (Prentice (1988)) –the parametric space of $\rho^{(i)}$ is more confined than $(-1, 1)$–, that is:

$$\rho^{(i)} \quad \in \quad [\, m(p_1^{(i)}, p_2^{(i)}), \;\; M(p_1^{(i)}, p_2^{(i)}) \,] \subseteq [-1, 1]$$

where:

$$\begin{aligned} m(p_1^{(i)}, p_2^{(i)}) &= \max\left\{ -\sqrt{\frac{p_1^{(i)} \cdot p_2^{(i)}}{q_1^{(i)} \cdot q_2^{(i)}}}, -\sqrt{\frac{q_1^{(i)} \cdot q_2^{(i)}}{p_1^{(i)} \cdot p_2^{(i)}}} \right\} \\ M(p_1^{(i)}, p_2^{(i)}) &= \min\left\{ +\sqrt{\frac{p_1^{(i)} \cdot q_2^{(i)}}{p_2^{(i)} \cdot q_1^{(i)}}}, +\sqrt{\frac{p_2^{(i)} \cdot q_1^{(i)}}{p_1^{(i)} \cdot q_2^{(i)}}} \right\} \end{aligned}$$

3.2 Treatment effects and non-equivalence between hypotheses

It can be easily proved by inspection of (7) that if the treatment has no effect on any of the marginal components and the correlation between them is the same in the two groups, then, the treatment has no effect on the composite endpoint, that is:

$$\mathrm{OR}_1 = 1, \;\; \mathrm{OR}_2 = 1, \;\; \rho^{(0)} = \rho^{(1)} \quad \Longrightarrow \quad \mathrm{OR}_* = 1$$

Note that this result could be restated in terms of the event proportions:

$$p_1^{(0)} = p_1^{(1)}, \;\; p_2^{(0)} = p_2^{(1)}, \;\; \rho^{(0)} = \rho^{(1)} \quad \Longrightarrow \quad p_*^{(0)} = p_*^{(1)}$$

However, the reciprocal is not necessarily true, that is to say, the effect of treatment on any endpoint ($\mathrm{OR}_1 < 1$ or $\mathrm{OR}_2 < 1$) could be diluted on the composite ($\mathrm{OR}_* = 1$). This complex relationship between the odds ratios of each component and the composite shows how the treatment effects are differently measured on each endpoint, and cannot be taken as equivalent. Thus, the two hypothesis tests being considered to test the treatment effect on either endpoint, $\mathcal{H}_1$ (stated in (1)) and $\mathcal{H}_*$ (stated in 4)), are not equivalent.

4 Asymptotic Relative Efficiency

We extend the ARE method developed by Gómez and Lagakos for time-to-event endpoints to binary endpoints. The extension relies on the asymptotic behaviours of the score tests $T_{1,n}$ for the relevant endpoint given in (2) and $T_{*,n}$ for the composite endpoint given in (5) presented in section 2, instead of the log-rank test that was used for survival endpoints. In the following sections we present the ARE method and its version for fixed alternatives.

4.1 ARE method for contiguous alternatives

Consider the following not equivalent hypothesis tests based on the relevant endpoint and on the composite endpoint:

$$\mathcal{H}_{1,n}: \begin{cases} H_0: & \log(\mathrm{OR}_1) = 0 \\ H_{1,n}: & \log(\mathrm{OR}_1)_n = \frac{v_1}{\sqrt{n}} \end{cases} \qquad \mathcal{H}_{*,n}: \begin{cases} H_0: & \log(\mathrm{OR}_*) = 0 \\ H_{*,n}: & \log(\mathrm{OR}_*)_n = \frac{v_*}{\sqrt{n}} \end{cases}$$

Let $T_{1,n}$, $T_{*,n}$ be the score tests corresponding to $\mathcal{H}_{1,n}$ and $\mathcal{H}_{*,n}$, respectively. Whereas under the null hypothesis both tests asymptotically follow the standard normal distribution, under contiguous alternatives, they are asymptotically $N(\delta_1, 1)$ and $N(\delta_*, 1)$ with δ_1 and δ_* presented in (3) and (6), respectively. Both tests behave as a displaced normal distribution according to the non-centrality parameter of the test, δ_1 and δ_*. Since the power of both tests is governed by the non-centrality parameters δ_1 and δ_*, and the larger the parameter is the greater the power (see Figure 1), a comparison between them yields a criterion for relative efficiency. We define the ARE as the square of the ratio of the non-centrality parameters, that is:

$$\mathrm{ARE}(T_{*,n}, T_{1,n}) = \left(\frac{\delta_*}{\delta_1}\right)^2 = \frac{v_*^2 p_*^{(0)} q_*^{(0)}}{v_1^2 p_1^{(0)} q_1^{(0)}}. \tag{8}$$

$\mathrm{ARE}(T_{*,n}, T_{1,n}) > 1$ would imply larger powers if using ε_* while $\mathrm{ARE}(T_{*,n}, T_{1,n}) \leq 1$ would be in favour of using ε_1 as the best option for primary endpoint. Hence, choosing between ε_1 or ε_* is reduced to a comparison between the two means of the asymptotic law under contiguous alternatives. The best primary endpoint would be the one which has the greatest non-centrality parameter.

The method quantifies the differences in efficiency of using the composite or the relevant as primary endpoint to lead the trial and, moreover, provides a decision rule to define the primary endpoint. If the ARE is larger than 1, the composite endpoint may be considered the best option as primary endpoint. Otherwise, the relevant endpoint is preferred. However, when the ARE value is in the vicinity of one, the advantages of the composite endpoint over the relevant endpoint are too small to counteract the complicate interpretation of the composite endpoint. Thus, under this circumstance, the relevant endpoint could be used instead as primary endpoint.[new]

Summarizing, for every endpoint, given their event rates in the control group and their limiting treatment effect, the ARE value captures which endpoint is more efficient for designing a clinical trial and provides a criterion to choose among them.

4.2 ARE method for fixed alternatives

The ARE criterion to choose the primary endpoint given in (8) is based on alternative odds ratios which are close to 1. From a practical point of view, the interest might often be on detecting treatment effects OR_k ($k = 1, *$), not necessarily near 1, and to address this we propose an approximated ARE value.

The efficiency criterion for fixed alternatives, based on the two-sample score tests $T_{*,n}$ and $T_{1,n}$, is defined as:

$$are(p_1^{(0)}, p_2^{(0)}, \mathrm{OR}_1, \mathrm{OR}_2, \rho^{(0)}, \rho^{(1)}) = \frac{(\log(\mathrm{OR}_*))^2 p_*^{(0)} q_*^{(0)}}{(\log(\mathrm{OR}_1))^2 p_1^{(0)} q_1^{(0)}} \tag{9}$$

Expression (9) approaches the ARE definition in (8) if for each endpoint, we would consider the fixed treatment effect stabilized for the sample size as an approximate value for the limiting treatment, that is: $\sqrt{n}\log(\mathrm{OR}_1) \cong v_1$ and $\sqrt{n}\log(\mathrm{OR}_*) \cong v_*$.

Hence, the decision on whether to use a composite binary endpoint versus its most relevant component as the primary endpoint can be assessed by computing the ARE for fixed alternatives, referred to as are. The are depends on the joint law of (X_{ij1}, X_{ij2}) $(i = 0, 1)$ and can be determined by the following

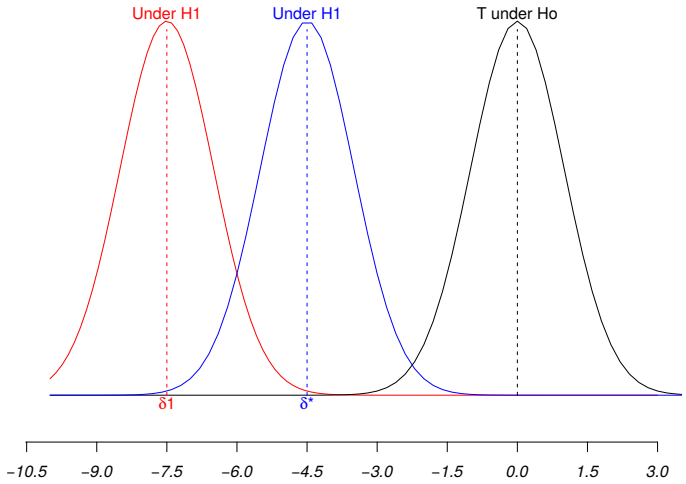


Figure 1 Asymptotic behavior of the score test under the null hypothesis (most right curve) and under contiguous alternatives for each endpoint ε_1 (most left curve) and ε_* (second left).[new]

anticipated parameters: (i) $p_1^{(0)}$ and $p_2^{(0)}$, event rates in control group for the relevant endpoint, ε_1, and the additional endpoint, ε_2; (ii) OR_1 and OR_2 fixed treatment effects for ε_1 and ε_2; (iii) $\rho^{(0)}$ and $\rho^{(1)}$, correlation between X_{ij1} and X_{ij2} for each group.

5 Case Study

Drug-eluting stents have been proved to reduce respenosis in noncomplex lesions, even so, their utility has not been studied in a patient population with more complex lesions. TAXUS-V was a prospective, multicenter, randomized trial to investigate the safety and efficacy of a placlitaxel-eluting stent in a patient population with more complete lesions than previously studied (Stone, et al. (2005)). The trial was conducted from February 2003 to March 2004 at 66 academic and community-based institutions with 1156 patients who underwent stent implantation in a single coronary artery stenosis, including 664 patients (57.4%) with complex or previously unstudied lesions and 9-month clinical and angiographic follow-up. Patients were randomly assigned to receive one or more bare metal stents ($n = 579$) or identical-appearing paclitaxel-eluting stents ($n = 577$).

The primary endpoint was the 9-month incidence of ischemia-driven target vessel revascularization, ε_1. As a secondary endpoint, major adverse cardiac events, ε_*, were defined as ischemia-driven target-vessel revascularization, ε_1, or death from cardiac causes or myocardial infarction, ε_2. The study shows that compared with a bare metal stent, implantation of the paclitaxeleluting stent in a patient population with complex lesions effectively reduces the rate of vessel revascularization.

For illustrative purposes, we assume that a study in a similar setting is to be planned, and the question that arises is which primary endpoint should be used to lead the trial. We also assume that the results of TAXUS-V are used for this purpose. Aiming to study whether it would be more efficient to base the study on major adverse cardiac events, ε_*, instead of ischemia-driven target vessel revascularization, ε_1, we exemplify the use of the ARE method.

The frequency of target vessel revascularization in bare metal group is $p_1^{(0)} = 0.173$, whereas the frequency of death from cardiac causes or myocardial infarction is $p_2^{(0)} = 0.055$. Furthermore, the frequencies under the test group are $p_1^{(1)} = 0.121$ and $p_2^{(1)} = 0.057$, respectively. We discuss the use of the composite endpoint as primary endpoint, for given values $p_1^{(0)} = 0.173$, $p_2^{(0)} = 0.055$ and $p_1^{(1)} = 0.121$ and for the

values for the parameter $p_2^{(1)}$ and ρ presented in Table 1. For given pairs $(p_1^{(0)}, p_2^{(0)})$ and $(p_1^{(1)}, p_2^{(1)})$ and assuming equal correlation in both groups, the eligible values for ρ lie in the interval $(-0.09, 0.53)$.

Table 1 Values of $p_1^{(0)}$ and $p_1^{(1)}$, probability of target vessel revascularization in bare metal group and in placitaxel-eluting group; $p_2^{(0)}$ and $p_2^{(1)}$, probability of death from cardiac causes or myocardial infarction in bare metal group and in placitaxel-eluting group; ρ, correlation among target vessel revascularization and death from cardiac causes or myocardial infarction; OR_1, odds ratio for target vessel revascularization; OR_2, odds ratio for death from cardiac causes or myocardial infarction, used for the discussion. The left part of the table shows the treatment effects in terms of p, the right part shows the treatment effects in terms of OR.

Parameter	Values	Parameter	Values
$p_1^{(0)}$	0.173	$p_1^{(0)}$	0.173
$p_1^{(1)}$	0.121	OR_1	0.67
$p_2^{(0)}$	0.055	$p_2^{(0)}$	0.055
$p_2^{(1)}$	$0.057, 0.050, 0.045, 0.040, 0.035$	OR_2	$1.04, 0.90, 0.81, 0.72, 0.62$
ρ	$(-0.09, 0.53)$	ρ	$(-0.09, 0.53)$

Figure 2 depicts the ARE values (in log scale) in terms of the correlation for each of the five different values of the treatment effect on ε_2. Observe that for a fixed correlation the ARE takes greater values as the odds ratio OR_2 for death from cardiac causes or myocardial infarction decreases. Therefore, the composite endpoint becomes more effective and more useful when the odds ratio for the additional endpoint, OR_2, shows a greater treatment effect. Furthermore, notice that the ARE decreases when the correlation increases, that is, the more correlated among target vessel revascularization and death from cardiac causes or myocardial infarction are, the less appropriate necessary is the composite as primary endpoint.

Especially, when the odds ratio for death from cardiac causes or myocardial infarction, OR_2, is equal or larger than 0.81, the ARE is almost always less than 1 (see Figure 2). Hence, the use of target vessel revascularization, ε_1, provide more efficient detection of the differences between treatments. In the case that $\mathrm{OR}_2 \leq 0.62$[new], the ARE is greater than 1. Then, the primary endpoint major adverse cardiac events, ε_*, would have been more efficient instead of relevant endpoint. Finally, note that when OR_2 is around 0.72, the decision depends on the value that correlation has.

6 Statistical efficiency guidelines

We have seen that the relative efficiency to choose between a composite endpoint or one of its relevant components can be expressed in terms of the following anticipated parameters: treatment effects, event rates and correlations. In this section, we discuss the influence that these parameters have on the relative efficiency value. For example, which role does play the correlation between the two components in preferring the composite as primary endpoint? We conclude reporting guidelines which could be of some help when designing a randomized clinical trial and facing the choice between several binary endpoints or their combination.

6.1 Design

Our efficiency guidelines will be based on [new] event rates, $p_1^{(0)}, p_2^{(0)}$,[new] smaller than 0.1, odds ratios, OR_1, OR_2,[new] between 0.5 and 1, and positive correlations (see Table 2)[new]. This choice is in accordance with the values that are usually encountered in clinical trials. From now on, we assume that the correlations are the same in the two groups and we denote it by ρ. Although Table 2 yields 436810 scenarios, since[new]

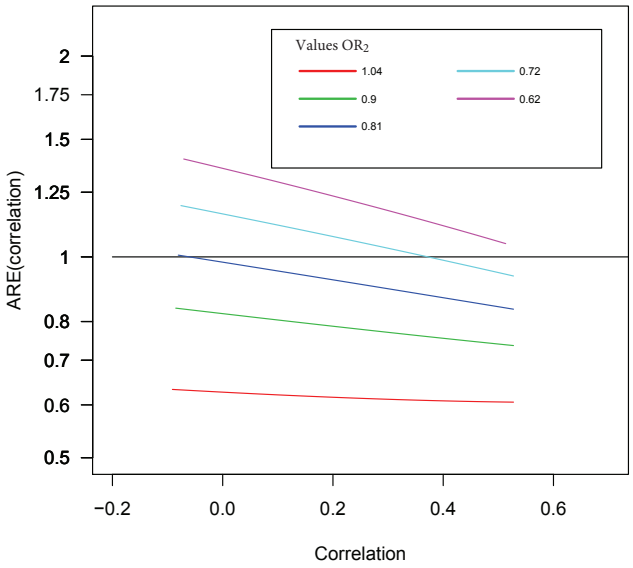


Figure 2 ARE of major adverse cardiac events (death from cardiac causes, myocardial infarction, or target-vessel revascularization) versus target-vessel revascularization for a range of correlation coefficient and different values of OR_2 for the parameters: $p_1^{(0)} = 0.173$, $p_2^{(0)} = 0.055$ and $p_1^{(1)} = 0.121$. The plot shows the curves of the are for each OR_2 depending on the assumed ρ.

for every pair $(p_1^{(0)}, p_2^{(0)})$ and $(p_1^{(1)}, p_2^{(1)})$ not all the correlations are feasible, the total number of possible scenarios is reduced to 315348.

Since the ARE method for fixed alternatives given in (9) depends on the parameters $p_1^{(0)}$, $p_2^{(0)}$, OR_1, OR_2 and ρ, we calculate the are for each scenario. The ARE values that we have obtained has 1.52 as a median, and 0.81 and 4.82 as first and third quartile. We follow the principle that if $are > 1$, the use of the composite endpoint is recommended, and if $are \leq 1$, the relevant endpoint should be used as primary endpoint. At last, we compute the percentage of cases on which the composite is preferred over the relevant endpoint. We conclude with recommendations for the choice of the primary endpoint in terms of the values of the correlations, the treatment effects and the event rates in control group for each individual component. We have performed all computations using `R` software tool (Version 0.98.1087), the time required to perform the considered scenarios was 16.58h.

As said earlier, when the ARE values are close to one,[new] in particular if[new] $are \in (1, 1.1)$, the benefits of using the composite endpoint over the relevant endpoint are small. Despite the value one is regarded as the threshold of our study and is the focus of the subsequent discussion, guidelines using 1.1 as the threshold for the decision can be viewed in Appendix B.

6.2 General pattern of the percentage of cases in which $are > 1$

We study the influence that the value of certain anticipated parameters, such as the treatment effect on the relevant endpoint or the event rate of the additional endpoint, has on the selection between a composite endpoint or its more relevant component as primary endpoint. As we will see, the most well-suited primary endpoint might differ according to the anticipated parameters of the clinical trial.

Table 2 Values of parameters $p_1^{(0)}, p_2^{(0)}, \mathrm{OR}_1, \mathrm{OR}_2$ and ρ for the settings used for the efficient guidelines.

Parameter	Values
$p_1^{(0)}$, $p_2^{(0)}$	$0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055, 0.060, 0.065,$ $0.070, 0.075, 0.080, 0.085, 0.090, 0.095, 0.100$
OR_1, OR_2	$0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.99$
ρ	$0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$
Total scenarios	436810
Possible scenarios	315348

We have computed the are values for each of the 315348 scenarios described in Table 2 and in each case we have recorded whether $are > 1$ –the composite endpoint would be recommended[new]– or $are \leq 1$ –the relevant endpoint should be kept as primary endpoint–. A given scenario is characterized by the following 5 parameter values $\theta = (\mathrm{OR}_1, \mathrm{OR}_2, \rho, p_1^{(0)}, p_2^{(0)})$. Let $P_1(a)$ indicate the percentage of cases yielding $are > 1$ among all the scenarios with $\mathrm{OR}_1 = a$. Analogously define $P_j(a)$ as the percentage of cases yielding $are > 1$ among all the scenarios with $\theta_j = a$ $(j = 2, \cdots, 5)$.

We have examined $P_1(\mathrm{OR}_1)$ for $0.5 \leq \mathrm{OR}_1 < 1$, $P_2(\mathrm{OR}_2)$ for $0.5 \leq \mathrm{OR}_2 < 1$ and $P_3(\rho)$ for $0 \leq \rho \leq 1$. We observe that the percentage of situations in which $are > 1$ increases whenever: i) the relative effect of treatment on the relevant endpoint increases, ii) the relative effect of treatment on the additional endpoint decreases and iii) the correlation between the two endpoints decreases. In other words, the number of situations where the composite endpoint is preferred is larger i) for larger values of OR_1, ii) for smaller values of OR_2 and iii) for weakly correlated endpoints. Figure 3 and Figures 5, 6 (in Appendix B) summarize these findings.

We have studied the behavior of $P_2(\mathrm{OR}_2)$ as a function of OR_1. Figure 4 represents $P_2(\mathrm{OR}_2 = \mathrm{OR}_1 + a)$ for $\mathrm{OR}_1 = 0.6$ and $-0.10 \leq a \leq 0.35$. We observe that the percentage of cases in which the composite is preferred drops off rapidly when the effect of treatment is not as strong on the additional endpoint as it is on the relevant endpoint (see Figures 7 and 8 when $\mathrm{OR}_1 = 0.7$ and $\mathrm{OR}_1 = 0.8$ in Appendix B).

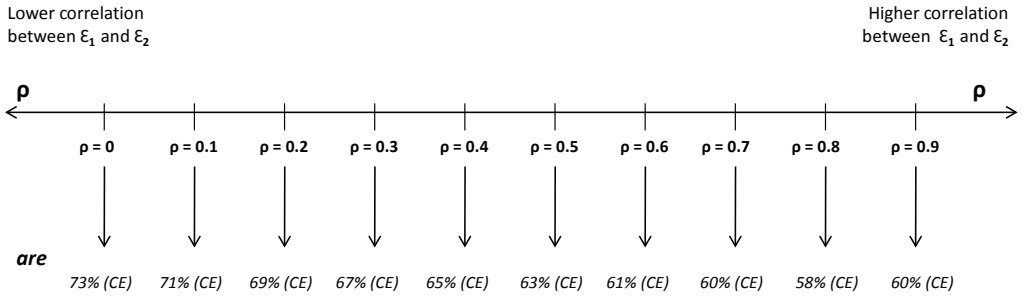


Figure 3 Percentage of scenarios in which the composite endpoint should be used depending on ρ.

We have also studied the behavior of $P_4(p_1^{(0)})$ for $0.01 \leq p_1^{(0)} \leq 0.1$ and of $P_5(p_2^{(0)})$ for $0.01 \leq p_2^{(0)} \leq 0.1$. There is a certain trend showing (plots not provided) that $P_4(p_1^{(0)})$ decreases with $p_1^{(0)}$ while $P_5(p_2^{(0)})$ increases with $p_2^{(0)}$, indicating that less frequent event rates for the relevant endpoint and more frequent for the additional endpoint are in the direction of preferring the composite endpoint. However, the values for $P_4(p_1^{(0)})$ or $P_5(p_2^{(0)})$ are between 60% and 75%, implying that the other parameters $(\mathrm{OR}_1, \mathrm{OR}_2, \rho)$ play a more important role in the choice between the relevant and the composite endpoint.

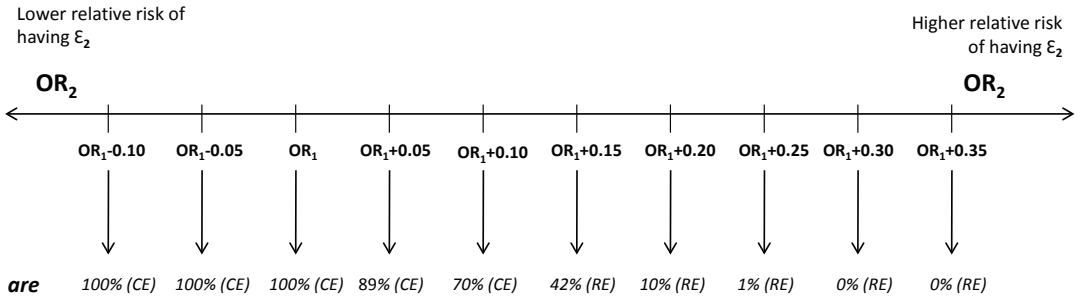


Figure 4 Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1 = 0.6$.

We have as well examined whether the behavior of the percentages $P_1(\mathrm{OR}_1)$ and $P_2(\mathrm{OR}_2)$ (see Figures 5 and 6) remains the same for different correlations ρ. We observe that the number of situations where the composite endpoint is preferred increases when ρ decreases and either OR_1 increases or OR_2 decreases (tables not provided).

6.3 Recommendations for the choice of the primary endpoint

We have splitted the recommendations into the following three cases: (I) when the correlation takes values between 0 and 1 ($0 < \rho < 1$); (II) when the relevant and the additional endpoint are independent ($\rho = 0$); and (III) when $\rho = 1$, implying that the relevant and the additional endpoint take the same value.

(I) Although the total number of scenarios that we have reproduced is very large and it has been useful to understand how the are behaves, when it comes to anticipate parameter values on which to base our decisions, accuracy cannot be as slim and is more realistic to render the recommendations to 3 or 4 categories of association, of strengths of the relative effect and of levels of frequency of the events. To this end, we have chosen four degrees of association: weak ($0 < \rho < 0.3$), medium-weak ($0.3 \leq \rho < 0.6$), medium-strong ($0.6 \leq \rho < 0.8$), strong ($0.8 \leq \rho < 1$); three categories for treatment effect: Large for Odds Ratios between 0.5 and 0.7, Medium for Odds Ratios between 0.7 and 0.9 and Low for Odds Ratios between 0.9 and 1; and four event rates in control group for the relevant and additional endpoints, low ($p \leq 0.025$), medium-low ($0.025 \leq p \leq 0.05$), medium-large ($0.05 \leq p \leq 0.075$), large ($p > 0.075$).

To derive recommendations, for each case we provide the percentage of cases in which the composite is preferred. On the basis of these percentages, we indicate whether the relevant or composite endpoint should be used. We are considering here that if the percentage of $are > 1$ is larger than 60%, the recommendation is to use of the composite endpoint; if the percentage is less than 40%, the recommendation is to use of the relevant endpoint; otherwise, if the percentage lies between 40% and 60%, the recommendation cannot be given. In this last case, we have reported that the recommendation is not conclusive and we have written CE/RE. There are not conclusions for all situations, therefore, the ensuing computation of the ARE is needed for the rest of particular situations.

Table 3 summarizes the recommendation in terms of the categories for $(\mathrm{OR}_1, \mathrm{OR}_2)$. Basically, the composite endpoint should be used when: i) treatment effect on the additional endpoint is large; ii) treatment effects on the relevant and additional endpoint are medium; iii) treatment effects on the relevant and additional endpoint are low and medium, respectively. On the other hand, the relevant endpoint is almost always preferred if the treatment effect on the additional endpoint is low and the treatment effect on the relevant is large or medium.

Recommendations taking into account the level of association together with the treatment effects on the relevant and on the additional endpoint, event rates in control group for both the relevant and the additional

Table 3 Recommendations in terms of treatment effects of the relevant and the additional endpoint, large ($0.5 \leq \text{OR} < 0.7$), medium ($0.7 \leq \text{OR} < 0.9$) or low ($0.9 \leq \text{OR} < 1$). Each cell indicates whether the relevant endpoint (RE) ($are \leq 1$) or composite endpoint (CE) ($are > 1$) should be used and, in parentheses, the percentage of cases in which composite is preferred based on the scenarios described in Table 2.

	Large treatment effect on ε_2	Medium treatment effect on ε_2	Low treatment effect ε_2
Large treatment effect on ε_1	CE (91.18%)	RE (23.06%)	RE (0%)
Medium treatment effect on ε_1	CE (100%)	CE (83.65%)	RE (6.52%)
Low treatment effect ε_1	CE (100%)	CE (100%)	CE (68.81%)

endpoint are summarized in Table 4. As earlier, we observe that the percentage of $are > 1$ decreases as the degree of association increases. This underlines the importance of the correlation to decide the primary endpoint. In particular, when the treatment effect either on the relevant or additional endpoint is medium, the value of the correlation might play a crucial role in the decision. Notice that the percentages of $are > 1$ in terms of the event rates are never larger than 75% or smaller than 50%, hence the frequency of the relevant and additional endpoints cannot characterize by themselves the decision on which primary endpoint to use.

Table 4 Recommendations in terms of degree of association between endpoints, weak ($0 < \rho < 0.3$), medium-weak ($0.3 \leq \rho < 0.6$), medium-strong ($0.6 \leq \rho < 0.8$), strong ($0.8 \leq \rho < 1$); treatment effects of the relevant and the additional endpoint, large ($0.5 \leq \text{OR} < 0.7$), medium ($0.7 \leq \text{OR} < 0.9$) or low ($0.9 \leq \text{OR} < 1$); event rates in control group for the relevant and additional endpoints, low ($p \leq 0.025$), medium-low ($0.025 \leq p \leq 0.05$), medium-large ($0.05 \leq p \leq 0.075$), large ($p > 0.075$). Each cell indicates whether the relevant endpoint (RE) ($are \leq 1$) or composite endpoint (CE) ($are > 1$) should be used and, in parentheses, the percentage of cases in which composite is preferred based on the scenarios described in Table 2.

	Correlation			
	Weak	Medium-weak	Medium-strong	Strong
Large treatment effect on ε_2	CE (99.72%)	CE (97.41%)	CE (92.87%)	CE (84.97%)
Medium treatment effect on ε_2	CE (74.96%)	CE (65.97%)	CE/RE (58.23%)	CE/RE (56.96%)
Low treatment effect ε_2	RE (23.61%)	RE (21.39%)	RE (20.99%)	RE (28.16%)
Large treatment effect on ε_1	CE/RE (49.80%)	CE/RE (42.29%)	RE (35.72%)	RE (38.00%)
Medium treatment effect on ε_1	CE (73.47%)	CE (68.72%)	CE (63.04%)	CE/RE (57.78%)
Low treatment effect ε_1	CE (92.16%)	CE (91.05%)	CE (89.87%)	CE (86.61%)
Low event rate for ε_1	CE (74.80%)	CE (73.89%)	CE (68.66%)	CE (65.36%)
Medium-low event rate for ε_1	CE (70.68%)	CE (66.41%)	CE (66.05%)	CE (62.95%)
Medium-large event rate for ε_1	CE (68.32%)	CE (62.78%)	CE/RE (59.01%)	CE (62.14%)
Large event rate for ε_1	CE (67.00%)	CE (60.52%)	CE/RE (53.29%)	CE/RE (50.22%)
Low event rate for ε_2	CE (66.63%)	CE/RE (57.96%)	CE/RE (53.87%)	CE/RE (52.92%)
Medium-low event rate for ε_2	CE (69.38%)	CE (64.16%)	CE/RE (55.77%)	CE/RE (54.99%)
Medium-large event rate for ε_2	CE (71.00%)	CE (67.34%)	CE (61.66%)	CE/RE (55.46%)
Large event rate for ε_2	CE (72.16%)	CE (69.20%)	CE (66.87%)	CE (66.61%)

(II) Whenever the relevant and additional endpoints are independent ($\rho = 0$), the composite endpoint would be intuitively preferred, however this is not always the case as Figure 3 shows. Following the rationale outlined above, Table 5 takes care of this situation. Note that the relevant endpoint is always preferred to the composite endpoint when the treatment effect on the relevant endpoint is large and the treatment effect on the additional endpoint is low. Besides, whenever the treatment effect on the relevant endpoint is medium and the treatment effect on the additional endpoint is low, the relevant endpoint should be the primary endpoint to lead the trial. Otherwise, if the treatment effect on the additional endpoint is large ($\mathrm{OR}_2 \leq 0.7$), the composite endpoint is always preferred.

Table 5 Recommendations in case of independence between the relevant and the additional endpoint ($\rho = 0$) in terms of treatment effects of the relevant and the additional endpoint, large ($0.5 \leq \mathrm{OR} < 0.7$), medium ($0.7 \leq \mathrm{OR} < 0.9$) or low ($0.9 \leq \mathrm{OR} < 1$). Each cell indicates whether the relevant endpoint (RE) ($are \leq 1$) or composite endpoint (CE) ($are > 1$) should be used and, in parentheses, the percentage of cases in which composite is preferred based on the scenarios described in Table 2.

	Large treatment effect on ε_2	Medium treatment effect on ε_2	Low treatment effect ε_2
Large treatment effect on ε_1	CE (100%)	CE/RE (48.84%)	RE (0%)
Medium treatment effect on ε_1	CE (100%)	CE (96.36%)	RE (15.12%)
Low treatment effect ε_1	CE (100%)	CE (100%)	CE (76.55%)

(III) The case of $\rho = 1$ was excluded from the settings of scenarios because in this case $are = 1$. The reason is that perfect linear dependence implies that the probabilities of the composite and the relevant endpoint are the same. As a result, it can be seen by inspection of (9) that the resulting are is equal to one. Hence, the decision rule sets up an equivalence between the relevant and composite endpoints in terms of efficiency.

7 Discussion

In this paper, we have proposed a method that allows an informed selection between a binary composite endpoint or one of its components as primary endpoint. Although composite endpoints are widely used as primary endpoints in clinical trials, as we have seen, they are not always the best option. The law governing the composite endpoint depends on the event rates, the magnitude of the treatment effects and the correlation between the components that form the composite. While the event rates and magnitude of the treatment effects can be reasonably well anticipated, this is not the case for the correlation between endpoints. Our methodology, and hence, the computation of the ARE has been established for different correlation values in each treatment group[new]. However, the scenarios to derive the guidelines have been restricted to the same correlation in both groups. The impact of this assumption as well as the scenarios with two different correlations remain as future work.

If at least we could anticipate the degree of association in terms of weak, medium or large, we could use Table 4 to decide which endpoint to use. The treatment effects of the relevant and the additional endpoints also have an important role for deciding the primary endpoint. As seen earlier in Table 3, when the additional endpoint presents a smaller treatment effect than the relevant endpoint, it could not be more efficient to base the trial on the composite instead of the relevant endpoint, since the effect of the therapy in these settings could be diluted by adding an endpoint.

In order to assess the appropriate choice of the primary endpoint, an interactive web-platform called CompARE (`https://cinna.upc.edu/compare`) has been designed to calculate the ARE method based on the information of the different endpoints together with anticipated values. This platform, initially developed for time-to-endpoints, has not been extended in other frameworks. We are currently working on extending the platform CompARE to address binary endpoints.

This paper has been restricted to composite endpoints defined by two components. The method could be used for composite endpoints formed by more than two components by identifying two subsets of possible components (S_R and S_A) and then comparing the composite versus one of its subsets, for instance, S_R. Furthermore, we could have used the difference (or ratio) of the event rates in both treatment groups instead of the odds ratios to capture the treatment effects. The are method in terms of event rates is being developed and is the focus of our future research.

The standard definition of the Asymptotic Relative Efficiency relates the efficiency of two statistic tests for the same set of hypothesis. In this case, it can be interpreted as the limiting ratio of sample sizes to give the same asymptotic power under sequences of contiguous alternatives (Noether (1954)). Gómez and Gómez-Mateu (2014) empirically proved that the interpretation of the ARE as the ratio of required sample sizes still holds when using two logrank tests to compare the hazard ratios under the relevant or the composite endpoint. It remains to be seen whether the are we have proposed for binary endpoints can as well be interpreted as a ratio of required sample sizes.

In this work the ARE method has been developed for discussing the use of a composite or one of its components as primary endpoint. We have assumed that both endpoints, ε_1 and ε_2, are important enough to be considered into the study and that one of the endpoints, ε_1, is more relevant than the other, ε_2. However, the ARE method does not take into account the relative relevance between ε_1 and ε_2. We understand this could be an important issue and remains open for future research.

Acknowledgements This work is partially supported by grants MTM2015-64465-C2-1-R (MINECO/FEDER) from the Ministerio de Economía y Competitividad (Spain), and 2014 SGR 464 (GRBIO) from the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain).

Conflict of Interest

The authors have declared no conflict of interest.

Appendix

A. Derivation of the Odds Ratio of a Binary Composite Endpoint

Theorem 7.1 *Let X_{ij1} and X_{ij2} denote the responses of two binary endpoints for the j-th patient in the i-th group of treatment ($i = 0, 1$, $j = 1, ..., n_i$). Denote by X_{ij*} the composite response defined as*

$$X_{ij*} = \begin{cases} 1, \text{ if } X_{ij1} + X_{ij2} \geq 1 \\ 0, \text{ if else } X_{ij1} + X_{ij2} = 0 \end{cases}$$

Denote by $p_1^{(i)} = \mathrm{P}(X_{ij1} = 1) = 1 - q_1^{(i)}$, $p_2^{(i)} = \mathrm{P}(X_{ij2} = 1) = 1 - q_2^{(i)}$ and $p_*^{(i)} = \mathrm{P}(X_{ij*} = 1) = 1 - q_*^{(i)}$ the probabilities of observing each endpoint in the i-th group, and by $\mathrm{O}_k^{(i)}$ the odds for each endpoint in the i-th group, that is, $\mathrm{O}_k^{(i)} = p_k^{(i)}/q_k^{(i)}$. Let $\mathrm{OR}_1, \mathrm{OR}_2$ be the odds ratio for both endpoints, that is, $\mathrm{OR}_k = \frac{p_k^{(1)}/q_k^{(1)}}{p_k^{(0)}/q_k^{(0)}}$, ($k = 1, 2$). Let $\rho^{(i)}$ the correlation between X_{ij1} and X_{ij2} in group i, ($i = 0, 1$).

The Odds Ratio for the composite endpoint, OR_*, is determined by six parameters $\mathrm{O}_1^{(0)}$, $\mathrm{O}_2^{(0)}$, OR_1, OR_2, $\rho^{(0)}$, $\rho^{(1)}$ and has the following expression:

$$\mathrm{OR}_* = \frac{\left((1 + \mathrm{OR}_1 O_1^{(0)})(1 + \mathrm{OR}_2 O_2^{(0)}) - 1 - \rho^{(1)}\sqrt{\mathrm{OR}_1 \mathrm{OR}_2 O_1^{(0)} O_2^{(0)}}\right)}{\left((1 + O_1^{(0)})(1 + O_2^{(0)}) - 1 - \rho^{(0)}\sqrt{O_1^{(0)} O_2^{(0)}}\right)} \frac{1 + \rho^{(0)}\sqrt{O_1^{(0)} O_2^{(0)}}}{1 + \rho^{(1)}\sqrt{\mathrm{OR}_1 \mathrm{OR}_2 O_1^{(0)} O_2^{(0)}}}$$

Proof.

$$\begin{aligned} \mathrm{OR}_* &= \left(\frac{1 - q_1^{(1)} q_2^{(1)} - \rho^{(1)}\sqrt{\frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}}}}{q_1^{(1)} q_2^{(1)} + \rho^{(1)}\sqrt{\frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}}}}\right) \cdot \left(\frac{1 - q_1^{(0)} q_2^{(0)} - \rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{q_1^{(0)} q_2^{(0)}}}}{q_1^{(0)} q_2^{(0)} + \rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{q_1^{(0)} q_2^{(0)}}}}\right)^{-1} \\ &= \left(\frac{\frac{1}{q_1^{(1)} q_2^{(1)}} - 1 - \rho^{(1)}\sqrt{\frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}}}}{1 + \rho^{(1)}\sqrt{\frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}}}}\right) \left(\frac{\frac{1}{q_1^{(0)} q_2^{(0)}} - 1 - \rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{q_1^{(0)} q_2^{(0)}}}}{1 + \rho^{(0)}\sqrt{\frac{p_1^{(0)} p_2^{(0)}}{q_1^{(0)} q_2^{(0)}}}}\right)^{-1} \end{aligned}$$

Taking into account:

$$\begin{aligned} \frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}} &= \frac{p_1^{(1)} p_2^{(1)}}{q_1^{(1)} q_2^{(1)}} \frac{q_1^{(0)}}{p_1^{(0)}} \frac{q_2^{(0)}}{p_2^{(0)}} O_1^{(0)} O_2^{(0)} = \mathrm{OR}_1 \mathrm{OR}_2 O_1^{(0)} O_2^{(0)} \\ \frac{p_1^{(0)} p_2^{(0)}}{q_1^{(0)} q_2^{(0)}} &= O_1^{(0)} O_2^{(0)} \\ \frac{1}{q_1^{(1)}} &= \frac{p_1^{(1)} + q_1^{(1)}}{q_1^{(1)}} = 1 + O_1^{(1)} = 1 + \mathrm{OR}_1 O_1^{(0)} \\ \frac{1}{q_1^{(1)} q_2^{(1)}} &= (1 + \mathrm{OR}_1 O_1^{(0)})(1 + \mathrm{OR}_2 O_2^{(0)}) \\ \frac{1}{q_1^{(0)} q_2^{(0)}} &= (1 + O_1^{(0)})(1 + O_2^{(0)}) \end{aligned}$$

Hence:

$$\mathrm{OR}_* = \frac{\frac{(1+\mathrm{OR}_1 O_1^{(0)})(1+\mathrm{OR}_2 O_2^{(0)}) - 1 - \rho^{(1)}\sqrt{\mathrm{OR}_1 \mathrm{OR}_2 O_1^{(0)} O_2^{(0)}}}{1+\rho^{(1)}\sqrt{\mathrm{OR}_1 \mathrm{OR}_2 O_1^{(0)} O_2^{(0)}}}}{\frac{(1+O_1^{(0)})(1+O_2^{(0)}) - 1 - \rho^{(0)}\sqrt{O_1^{(0)} O_2^{(0)}}}{1+\rho^{(0)}\sqrt{O_1^{(0)} O_2^{(0)}}}}$$

□

B. Results

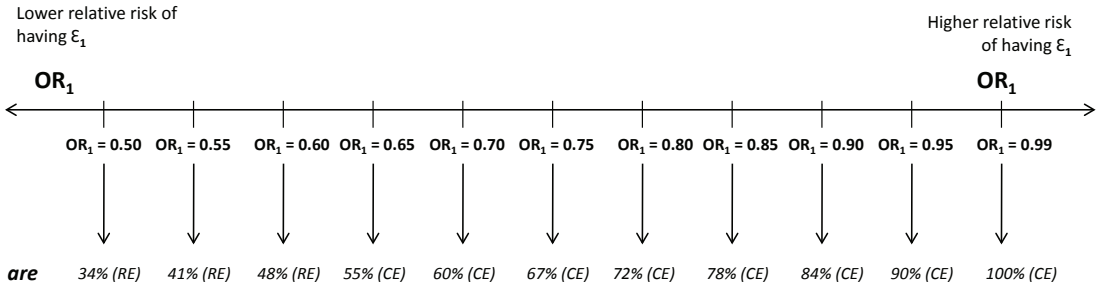


Figure 5 Percentage of scenarios in which the composite endpoint should be used depending on OR_1.

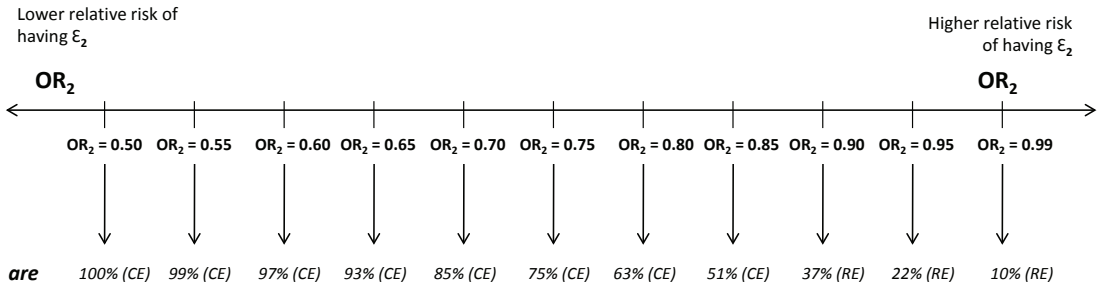


Figure 6 Percentage of scenarios in which the composite endpoint should be used depending on OR_2.

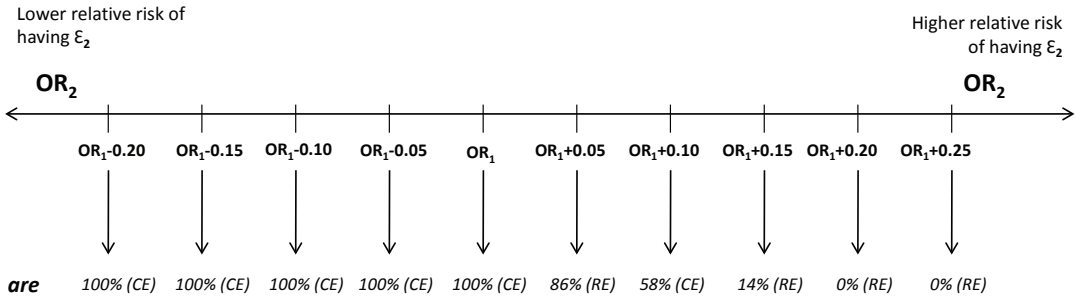


Figure 7 Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1 = 0.7$.

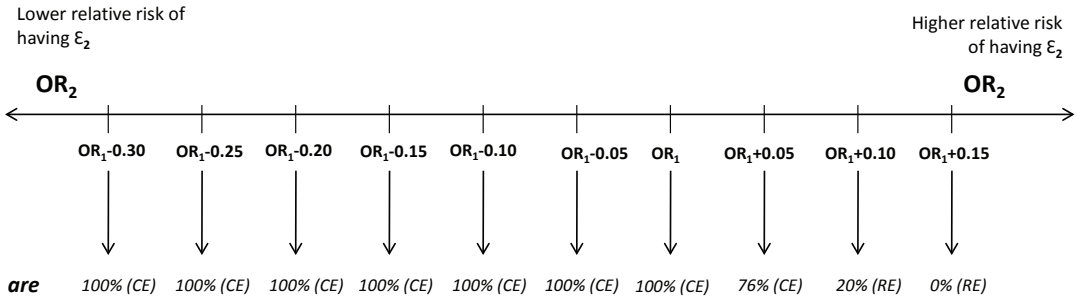


Figure 8 Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1 = 0.8$.

Table 6 Recommendations in terms of treatment effects of the relevant and the additional endpoint, large ($0.5 \leq \mathrm{OR} < 0.7$), medium ($0.7 \leq \mathrm{OR} < 0.9$) or low ($0.9 \leq \mathrm{OR} < 1$). Each cell indicates whether the relevant endpoint (RE) ($are \leq 1.1$) or composite endpoint (CE) ($are > 1.1$) should be used and, in parentheses, the percentage of cases in which composite is preferred based on the scenarios described in Table 2.

	Large treatment effect on ε_2	Medium treatment effect on ε_2	Low treatment effect ε_2
Large treatment effect on ε_1	CE (80.97%)	RE (15.65%)	RE (0.00%)
Medium treatment effect on ε_1	CE (99.84%)	CE (74.53%)	RE (4.23%)
Low treatment effect ε_1	CE (100.00%)	CE (99.99%)	CE (63.89%)

References

Bahadur R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. Stanford University Press. 158–168.

Bauer K. A., Eriksson B. I., Lassen M. R., Turpie A. G.; Steering Committee of the Pentasaccharide in Major Knee Surgery Study. (2001). Fondaparinux compared with Enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. *N Engl J Med*. **345**, 1305–1310.

Table 7 Recommendations in terms of degree of association between endpoints, weak ($0 < \rho < 0.3$), medium-weak ($0.3 \leq \rho < 0.6$), medium-strong ($0.6 \leq \rho < 0.8$), strong ($0.8 \leq \rho < 1$); treatment effects of the relevant and the additional endpoint, large ($0.5 \leq \text{OR} < 0.7$), medium ($0.7 \leq \text{OR} < 0.9$) or low ($0.9 \leq \text{OR} < 1$); event rates in control group for the relevant and additional endpoints, low ($p \leq 0.025$), medium-low ($0.025 \leq p \leq 0.05$), medium-large ($0.05 \leq p \leq 0.075$), large ($\leq p > 0.075$). Each cell indicates whether the relevant endpoint (RE) ($are \leq 1.1$) or composite endpoint (CE) ($are > 1.1$) should be used and, in parentheses, the percentage of cases in which composite is preferred based on the scenarios described in Table 2.

Correlation	Weak	Medium-weak	Medium-strong	Strong
Large treatment effect on ε_2	CE (97.29%)	CE (93.35%)	CE (87.94%)	CE (76.10%)
Medium treatment effect on ε_2	CE (68.08%)	CE (60.14%)	CE/RE (53.33%)	CE/RE (50.49%)
Low treatment effect ε_2	RE (21.53%)	RE (19.38%)	RE (18.93%)	RE (22.12%)
Large treatment effect on ε_1	RE (43.35%)	RE (35.82%)	RE (29.52%)	RE (28.45%)
Medium treatment effect on ε_1	CE (69.67%)	CE (64.59%)	CE (59.13%)	CE/RE (51.88%)
Low treatment effect ε_1	CE (91.33%)	CE (89.94%)	CE (88.09%)	CE (80.57%)
Low event rate for ε_1	CE (73.08%)	CE (71.92%)	CE (65.38%)	CE (58.18%)
Medium-low event rate for ε_1	CE (67.49%)	CE (63.06%)	CE (62.62%)	CE (56.39%)
Medium-large event rate for ε_1	CE (63.84%)	CE (58.08%)	CE/RE (54.85%)	CE (55.48%)
Large event rate for ε_1	CE (61.03%)	CE/RE (54.54%)	RE (48.06%)	RE (42.09%)
Low event rate for ε_2	CE (58.35%)	RE (49.03%)	RE (47.41%)	RE (43.03%)
Medium-low event rate for ε_2	CE (65.63%)	CE (59.95%)	CE/RE (50.24%)	RE (46.93%)
Medium-large event rate for ε_2	CE (68.45%)	CE (64.58%)	CE (58.19%)	RE (47.65%)
Large event rate for ε_2	CE (70.06%)	CE (66.95%)	CE (64.29%)	CE (61.53%)

Ferreira-González, Permanyer-Miralda G., Busse J.W., Bryant D.M., Montori V.M., Alonso-Coello P., Walter S.D., Guyatt G.H. (2007). Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. **60(7)**, 651–7.

Gómez G, Lagakos SW. (2013). Statistical considerations when using a composite endpoint for comparing treatment groups. *Stat Med*. Jul 1; 719–38.

Gómez, G. and Gómez-Mateu, M. (2014). The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses. *SORT*, **38(1)**, 73–88.

Legler J.M., Lefkopoulou M. and Ryan L.M. (1995). Efficiency and Power of Tests for Multiple Binary Outcomes. *Journal of the American Statistical Association*. **430(90)**, 680–693.

Lefkopoulou M. and Ryan L. (1993). Global tests for multiple binary outcomes. *Biometrics*. **1 49(4)**, 975–88.

Mascha E.J. and Sessler D.I. (2011). Statistical grand rounds: design and analysis of studies with binary- event composite endpoints: guidelines for anesthesia research. *Anesth Analg*. **112(6)**, 1461–71.

Noether, G. E. (1954). On a Theorem of Pitman. *The Annals of Mathematical Statistics.*

Prentice RL. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*. **44(4)**, 1033-48.

Pocock, S. J., Geller, N. L. and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, **43(3)**, 487–498.

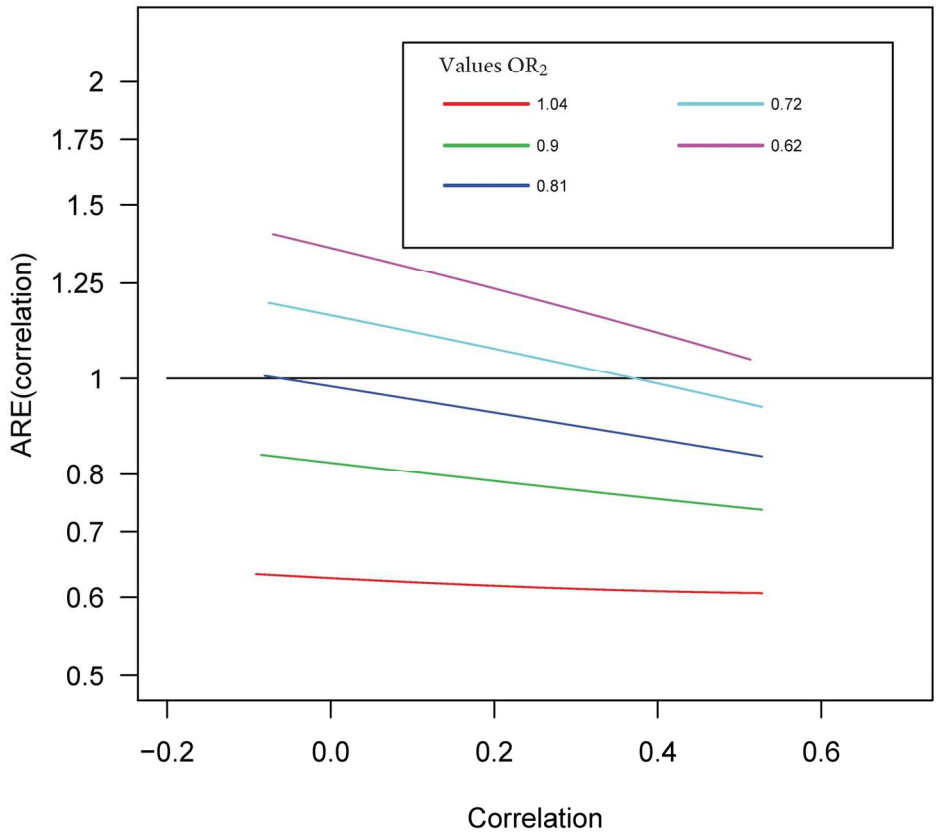
Rauch, G., Jahn-Eimermacher, A., Brannath, W. and Kieser, M. (2014). Opportunities and challenges of combined effect measures based on prioritized outcomes. *Statistics in Medicine*, **33(7)**, 1104–1120.

Rauch, G. and Kieser, M. (2013). An expected power approach for the assessment of composite endpoints and their components. *Computational Statistics and Data Analysis*, **60(1)**, 111–122.

Rauch, G., Wirths, M. and Kieser, M. (2014). Consistency-adjusted alpha allocation methods for a time-to-event analysis of composite endpoints. *Computational Statistics and Data Analysis*, **75**, 151–161.

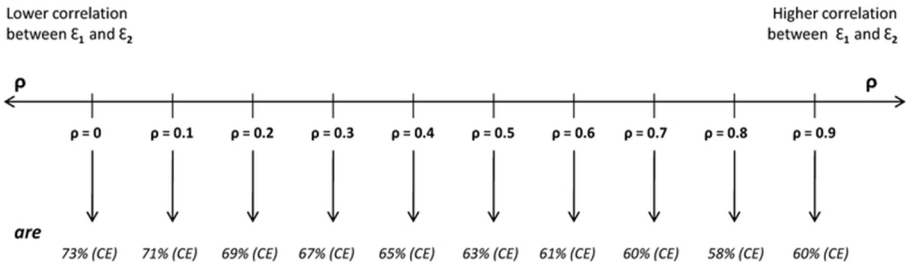
Sozu T., Sugimoto T. and Hamasaki T. (2010). Sample size determination in clinical trials with multiple co-primary binary endpoints. *Stat Med.* **29(21)**, 2169–79.

Stone GW, Ellis SG, Cannon L, Mann JT, Greenberg JD, Spriggs D, O'Shaughnessy CD, DeMaio S, Hall P, Popma JJ, Koglin J, Russell ME; TAXUS V Investigators. (2005). Comparison of a polymer-based paclitaxel-eluting stent with a bare metal stent in patients with complex coronary artery disease: a randomized controlled trial. *JAMA*, **294(10)**, 1215-23.



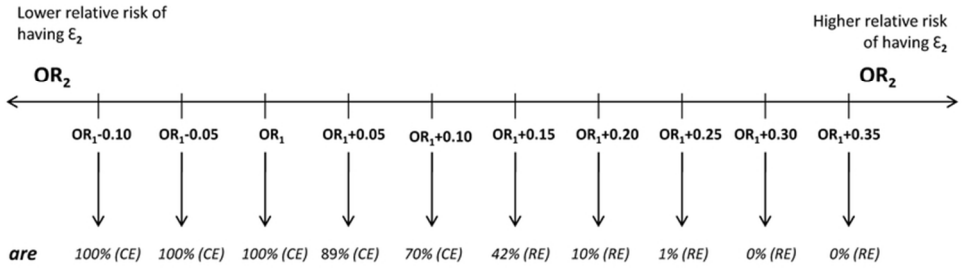
ARE of major adverse cardiac events (death from cardiac causes, myocardial infarction, or target-vessel revascularization) versus target-vessel revascularization for a range of correlation coefficient and different values of OR_2 for the parameters: $p_1^{(0)} = 0.173$, $p_2^{(0)}=0.055$ and $p_1^{(1)}=0.121$. The plot shows the curves of the are for each OR_2 depending on the assumed ρ.

152x152mm (300 x 300 DPI)



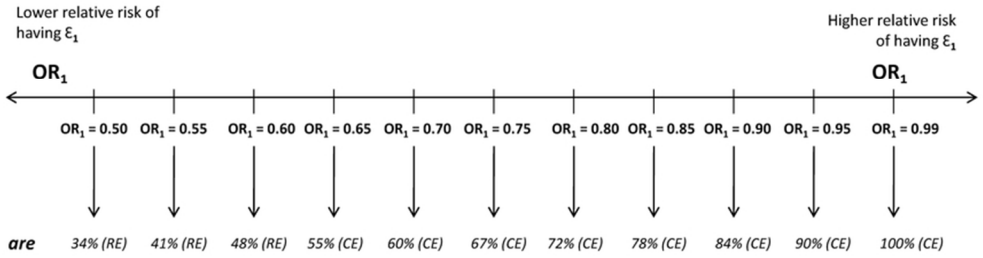
Percentage of scenarios in which the composite endpoint should be used depending on ρ.

72x20mm (300 x 300 DPI)



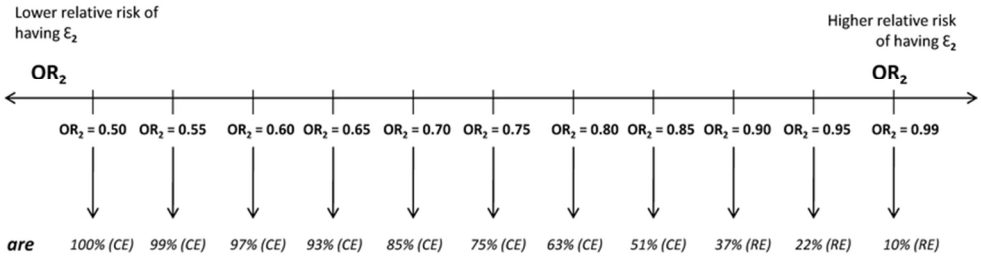
Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1=0.6$.

74x23mm (300 x 300 DPI)



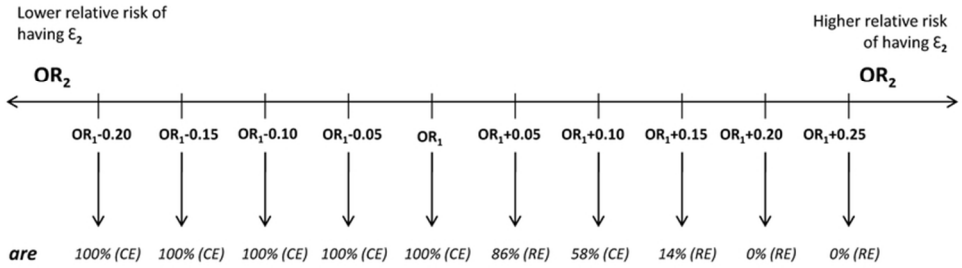
Percentage of scenarios in which the composite endpoint should be used depending on OR_1.

73x21mm (300 x 300 DPI)



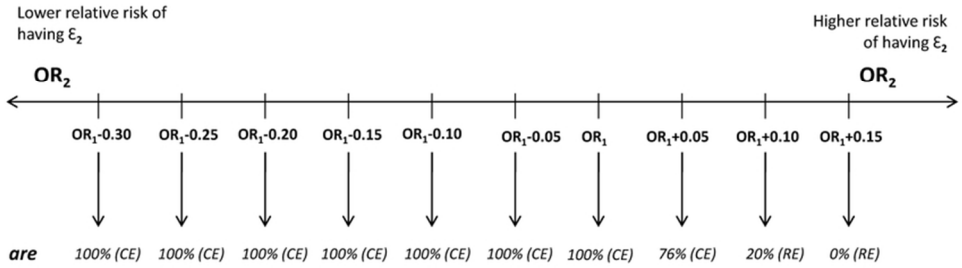
Percentage of scenarios in which the composite endpoint should be used depending on OR_2.

73x21mm (300 x 300 DPI)



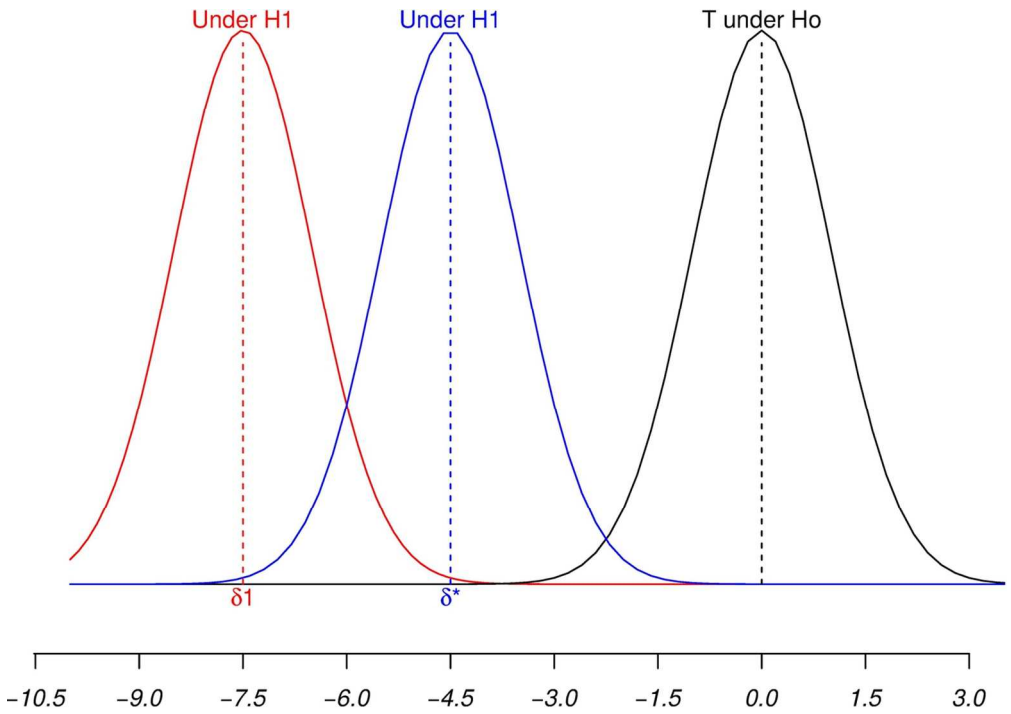
Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1=0.7$.

74x23mm (300 x 300 DPI)



Percentage of scenarios in which the composite endpoint should be used depending on OR_2 when $\mathrm{OR}_1=0.8$.

74x23mm (300 x 300 DPI)



Asymptotic behavior of the score test under the null hypothesis (most right curve) and under contiguous alternatives for each endpoint ε_1 (most left curve) and ε_* (second left).

123x86mm (300 x 300 DPI)