Estimating and Testing an Index of Bias Attributable to Composite Outcomes in Comparative Studies.

Fredi Alexander Diaz-Quijano, MD, Ph.D.

University of São Paulo, School of Public Health, Department of Epidemiology - Laboratório de Inferência Causal em Epidemiologia (LINCE-USP). São Paulo, Brazil. E-mail: frediazq@usp.br; Twitter: @DiazQuijanoFA

Abstract

Objective: To develop an index to evaluate the bias attributable to composite outcomes (BACO) in comparative clinical studies.

Study design and Setting: I defined the BACO index as the ratio of the logarithm of the association measure (e.g., relative risk) of the composite outcome to that of its most relevant component endpoint (e.g., mortality). Methods to calculate the confidence intervals and test the null hypotheses (BACO index = 1) were described and applied in systematically selected clinical trials. Two other preselected trials were included as “positive controls” for being examples of primary composite outcomes disregarded due to inconsistency with the treatment effect on mortality.

Results: The BACO index values different from one were classified according to whether the use of composite outcomes overestimated (BACO index >1), underestimated (BACO index between zero and <1), or inverted (BACO index <0) the association between exposure and prognosis. In three of 23 systematically selected trials and the two positive controls, the BACO indices were significantly lower than one (p <0.005).

Conclusion: BACO index can warn that the composite outcome association is stronger, weaker, or even opposite than that of its most critical component.

Keywords: Composite outcome; Mortality; Bias; Clinical trials; Delta method; Statistical inference

What is new?

- BACO index is the ratio of the logarithm of the association measure (e.g., RR) of the composite outcome to the logarithm of the association measure of the component endpoint that represents the study target (e.g., mortality).
- In comparative studies, the BACO index can be used to evaluate the correspondence between the effect on a composite outcome and that on its most critical component.
- This index could help to preset rules to make decisions for interpretation of clinical studies.
- A significant BACO should lead to the caution that the association of the composite is stronger (BACO index >1), weaker (BACO index between zero and <1), or even opposite (BACO index <0) than that of its most critical component.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
1. Introduction

Composite outcomes, particularly those that define an event when at least one of a group of component endpoints occurs, are increasingly used in clinical research [1–3]. In the last two decades, publications using terms referring to composite outcomes increased progressively in PubMed, from less than one hundred per year, before 2004, to more than one thousand per year since 2018. About a third of these publications are associated with the terms “clinical” and “trial” (Figure 1). These figures highlight the importance of this kind of outcome to evaluate interventions in clinical practice.

Figure 1. Annual counts of publications using terms related to Composite Outcome.*

* Output in Pubmed using the terms: “composite outcome” OR “composite endpoint” OR “composite end-point”.

The use of a composite outcome can have different methodological purposes [3,4]. When several potentially eligible endpoints exist, a composite outcome can avoid the need to choose a simple one and prevent problems associated with multiple comparisons [1,4]. Moreover, it can be a way to deal with competing risks [5]; or deliberately integrate events of different nature, such as indicators of the effectiveness and safety of an intervention [6].

In many other instances, composite outcomes are used to represent severity more broadly than with a single event (e.g., death). This practice facilitates a higher number of events for analysis, which increases the power of the study, reduces its costs, and provides a faster response to a research question [1,2,4]. However, the association measures are often misinterpreted as the effect of exposure on each of the elements of the composite outcome [2,7]. This is problematic when the overall effect on the composite outcome does not follow the same direction as its most critical components.

In some cases, effects on primary composite outcomes have been ruled out because they differ greatly from those on mortality [8–10]. This may occur because the composite outcome includes events generated by different mechanisms. For example, some components may be directly related to severity (heart attacks, death), while others may be more influenced by medical decisions and resource availability (hospitalization, catheterization). Therefore, some component endpoints can introduce bias that affects the estimation of the overall effect.
Despite their importance, little guidance is available on how composite outcomes should be interpreted, especially in situations of varied direction in the association across the event subtypes [4,6,11]. Moreover, there is a lack of statistical tools to support the decision to accept or rule out the use of the composite outcome. Furthermore, because composite outcomes are inherently correlated with their components [7], comparative estimates to quantify and test biases associated with the use of this type of endpoint are challenging to make.

In this paper, I proposed an index to evaluate the bias attributable to the composite outcome (BACO), which is simply the ratio of the logarithm of the association measure of the composite outcome to that of the component endpoint that represents the study target (e.g., mortality). For this purpose, I described how to calculate confidence intervals and perform hypothesis tests. Then, I applied these procedures in a group of clinical trials recently published in major medical journals.

2. Methods

2.1. BACO Index

I aimed to compare the association measure of a composite outcome with that obtained with a component endpoint, which indisputably represents the study target. Here, I chose the any-cause mortality as the target [12], and defined the BACO index as follows:

$$BACO\ index = \theta = \frac{\varphi_c}{\varphi_d}$$

Where $\varphi_c$ and $\varphi_d$ are the natural logarithms ($ln$) of ratio-based association measures for a composite outcome and death, respectively. In this paper, I mean relative risk (RR) as the association measure; however, the concept could be applied to other measures (e.g., hazard ratio or incidence rate ratio).

A BACO index equal to one indicates that there is no bias attributable to the use of a composite outcome. A BACO index higher than one indicates that the association with the composite outcome is stronger than that with death. A value between zero but less than one suggests that the association with the composite outcome is biased towards nullity. On the other hand, a negative value would result when the bias leads to an inversion of the association.

These interpretations can be applied regardless of the reference group in the comparative study. Because if the comparison groups were inverted, the signs of both the numerator and denominator would also be inverted.

Since the BACO index is a ratio between two correlated variables, calculating its variance, $\nu(\theta)$, requires considering the covariance between $\varphi_c$ and $\varphi_d$. Specifically, by applying the Taylor series estimator [13], we have that

$$\nu(\theta) = \theta^2 \left[ \frac{\nu(\varphi_c)}{\varphi_c^2} \cdot \frac{\nu(\varphi_d)}{\varphi_d^2} + 2 \cdot \frac{c(\varphi_c, \varphi_d)}{\varphi_c \varphi_d} \right]$$

Where $\nu(\varphi_c)$ and $\nu(\varphi_d)$ are the variances of $\varphi_c$ and $\varphi_d$, respectively, and $c(\varphi_c, \varphi_d)$ is their covariance. This covariance is usually calculated through complex matrix operations [13,14]. However, in the case of the BACO index, we analyze effects on two outcomes, one of which is completely embedded within the other. Therefore, $\varphi_c$ can be disaggregated as a sum of two effects: on death and on the other component endpoints, namely, $\varphi_c = \varphi_d + \varphi_{c-d}$. As detailed
in the Appendix A, with this approach, I deduced that \( c(\varphi_c, \varphi_d) = v(\varphi_c) \). Consequently, the formula of \( v(\theta) \) can be simplified to:

\[
v(\theta) = \theta^2 \left[ \frac{v(\varphi_c)}{\varphi_c^2} + \frac{v(\varphi_d)}{\varphi_d^2} - 2 \frac{v(\varphi_c \varphi_d)}{\varphi_c \varphi_d} \right]
\]

A 95% confidence interval for the BACO index is

\[
\theta \pm 1.96\sqrt{v(\theta)}
\]

And the statistic \( Z \) with normal distribution to test the null hypothesis (BACO index = 1) is

\[
Z = \frac{\theta - 1}{\sqrt{v(\theta)}}
\]

Thus, the BACO index can be estimated and tested directly from contingency tables (Appendix A). Alternatively, we can combine regression parameters, which may be useful to estimate a BACO index from adjusted association measures. For example, the \( \ln(\text{RR}) \) can be estimated as the coefficient of a Poisson regression, which is an alternative to avoid the problems of convergence of the Log-Binomial regression [15]. After, we can combine the results into one parameter vector and simultaneous (co)variance matrix of the sandwich/robust type, which is appropriate even if the estimates were obtained with the same or overlapping data [16]. Then, using the non-linear combination of parameters, the BACO indices can be calculated as the ratio of the regression coefficient of the composite outcome to that of mortality (see example in Appendix B). These procedures lead to results virtually identical to those obtained directly from contingency tables (Appendix C).

### 2.2. Estimation with simulated data

To illustrate different results of the BACO index, I simulated a comparative study in which a group of a thousand people was exposed to an experimental intervention and presented mortality of 4.8% during the follow-up. This group was compared with a reference group of another thousand people, which showed an 8% mortality. Consequently, the RR for death was 0.6 (4.8%/8%), in other words, the intervention had an efficacy of 40% for reducing mortality.

Now consider four composite outcomes, all of which included death within their components. The first composite outcome maintained the same RR as observed for mortality. The second outcome led to an overestimation of the treatment effect with a RR of 0.3 (efficacy of 70%). On the contrary, the third outcome led to an underestimation of the effect with a RR of 0.8 (efficacy of 20%). Finally, the fourth composite outcome led to an inversion of the association measure, suggesting that the intervention duplicates the risk of the event (Table 1).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Intervention (n=1,000)</th>
<th>Reference (n=1,000)</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>48</td>
<td>80</td>
<td>0.6</td>
</tr>
<tr>
<td>Composite outcome 1 (unbiased)</td>
<td>120</td>
<td>200</td>
<td>0.6</td>
</tr>
<tr>
<td>Composite outcome 2 (overestimated effect)</td>
<td>60</td>
<td>200</td>
<td>0.3</td>
</tr>
<tr>
<td>Composite outcome 3 (underestimated effect)</td>
<td>160</td>
<td>200</td>
<td>0.8</td>
</tr>
<tr>
<td>Composite outcome 4 (inverted effect)</td>
<td>320</td>
<td>160</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 1. Frequency of simulated outcomes in two hypothetic groups.
Figure 2 illustrates the point estimates and the corresponding 95% confidence intervals (95% CI). For the first composite outcome (unbiased), the BACO index was equal to one (95% CI: 0.46 to 1.54). For the second composite outcome, which overestimated the effect of the intervention on prognosis, the BACO index was 2.36 (95% CI: 1.14 to 3.58). The third outcome, which underestimated the effect, presented a BACO index of 0.44 (95% CI: 0.11 to 0.76). For the last simulated composite outcome, which had an inverted association measure, the BACO index was negative: -1.36 (95% CI: -2.48 to -0.24).

![Fig. 2  BACO indices of four simulated composite outcomes.](image)

### 2.3. Application

I searched PubMed to identify 2-group, parallel-design clinical trials published in the journal groups of JAMA, NEJM, and Lancet in 2019 (updated on 07/01/2020), using the following word combination: composite primary (endpoint OR outcome OR (“end-point”)) (mortality OR death) (randomised OR randomized) (trial) (JAMA OR NEJM OR Lancet). Next, I selected those studies whose primary outcome was a composite, binary, and included all-cause mortality within their components. Secondary subgroup analyses and studies with five or fewer fatal events were excluded. A study was also excluded whose outcome results were mainly based on imputations because it presented substantial losses during the follow-up [17].

I reviewed each article and built a database to reproduce individual information for the variables of intervention, composite outcome, and death. When necessary, I contacted the corresponding author to ask for data not provided in the article. For each trial, I calculated the RR for both the composite outcome (RRc) and death from any cause (RRd), using the corresponding intervention as the independent variable. After, I estimated the BACO index as a ratio of regression coefficients and performed Wald-type tests, based on the delta method, for the null hypothesis that the BACO index is equal to one. These analyses were conducted using Stata software (version 15.0, Stata Corp LP, College Station, TX, USA), and the commands employed are described in the Appendix B.

This work intended to reproduce the independent application of the BACO index in each of the studies. Therefore, I did not consider the number of trials to adjust the level of significance. However, I preset 0.005 as a level to define a statistically significant BACO [18].
A higher level (e.g., 0.05) was not chosen because I assumed that the composite outcomes were purposefully defined to be consistent with mortality, thus expecting a low pre-test probability that the BACO index is different from one. However, I used the term “suggestive” for p values between 0.005 and 0.05 [19].

Besides the selected studies from 2019, I also analyzed the clinical trials CAPRICORN and EXPEDITION [20,21]. CAPRICORN investigated carvedilol in patients with left-ventricular dysfunction after acute myocardial infarction [20]; and EXPEDITION evaluated intravenous caripode in high-risk coronary artery bypass graft surgery patients [21]. These two studies were included as “positive controls” because they represented examples of primary composite outcomes that were disregarded for being inconsistent with the treatment effect on mortality [9,10].

3. Results

From 82 articles, I selected 23 clinical trials, most of which were about cardiovascular diseases. Besides mortality, the composite outcomes integrated diverse components, often including cardio-cerebrovascular events, such as myocardial infarction or stroke, and other related to the use of health services, such as hospital admission and vascular interventions. The sample sizes of these trials ranged from 240 to 536,233. While the number of patients evolving composite outcomes ranged from 31 to 4,067, the number of deaths ranged from six to 1,140 (Table 2).

In six studies [22–27], the RRs of the composite outcomes were further from the null value than the corresponding RRs of death (Table 3). Consequently, their BACO indices were greater (although not statistically different) than one. One study had the same association measure for both the composite outcome and any-cause death (BACO index = 1) [28].

BACO index was lower than one in the other 16 studies [29,30,39–44,31–38], and the BACO was statistically significant in three of them. These included the study by Yasuda et al., in which a monotherapy with rivaroxaban was compared to a combination therapy with rivaroxaban plus a single antiplatelet agent [32]. The apparent effect of the monotherapy on the composite outcome (RRc: 0.74; 95% CI: 0.57 to 0.96) was lower than the effect on mortality (RRd: 0.56; 95% CI: 0.39 to 0.82), in patients with atrial fibrillation and coronary disease. Thus, the BACO index was 0.53 (95% CI: 0.21 to 0.86; p=0.0048).

The second study with statistically significant BACO was that by Lanz et al., which compared a self-expanding versus a balloon-expandable bioprosthesis for transcatheter aortic valve replacement in patients with symptomatic severe aortic stenosis [37]. The self-expanding bioprosthesis group presented an incidence of the composite outcome about 43% higher than observed in the balloon-expandable group (RRc: 1.43; 95% CI: 1.06 to 1.92). However, mortality in the first group was about three-times that of the second one (RRd: 2.96; 95% CI: 0.81 to 10.8). In this trial, the BACO index was 0.33 (95% CI: -0.1 to 0.76; p=0.002).

In the other study, Onland et al. evaluated the effect of systemic hydrocortisone compared with placebo on a composite outcome of death or bronchopulmonary dysplasia in very preterm infants [38]. The intervention was not associated with a significant change in the composite outcome incidence (RRc= 0.95; 95%CI: 0.84 to 1.08). However, the hydrocortisone group exhibited lower mortality compared to placebo (RRd: 0.65; 95% CI: 0.42 to 0.99). BACO index in this case was 0.11 (95% CI: -0.17 to 0.39; p<0.001).
Table 2. General description of the study population and outcomes of clinical trials selected.

<table>
<thead>
<tr>
<th>First Author</th>
<th>Characteristics of the study population</th>
<th>Composite elements other than all-cause mortality</th>
<th>n</th>
<th>Composite/Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holm NR</td>
<td>Left main coronary artery disease requiring revascularization.</td>
<td>Non-procedural myocardial infarction, repeat revascularization, or stroke.</td>
<td>1,201</td>
<td>275/104</td>
</tr>
<tr>
<td>Katheria A</td>
<td>Preterm infants (born at 23-31 weeks’ gestation).</td>
<td>Severe intraventricular hemorrhage.</td>
<td>474</td>
<td>49/32</td>
</tr>
<tr>
<td>Schüpke S</td>
<td>ACS for whom invasive evaluation was planned.</td>
<td>Myocardial infarction or stroke at one year.</td>
<td>4,018</td>
<td>321/163</td>
</tr>
<tr>
<td>Brott TG</td>
<td>Moderate or severe atherosclerotic symptomatic stenosis at the carotid bifurcation.</td>
<td>Stroke within 120 days after randomization or subsequent ipsilateral stroke up to 10 years after randomization.</td>
<td>4,754</td>
<td>447/39</td>
</tr>
<tr>
<td>Tomaniak M</td>
<td>Acute coronary syndromes (ACS) beyond one month after a percutaneous coronary intervention (PCI).</td>
<td>New Q-wave myocardial infarction.</td>
<td>7,487</td>
<td>130/89</td>
</tr>
<tr>
<td>Wu Y</td>
<td>ACS in resource-constrained hospitals.</td>
<td>Reinfarction/myocardial infarction or nonfatal stroke.</td>
<td>29,346</td>
<td>1,214/1,140</td>
</tr>
<tr>
<td>Hahn JY</td>
<td>Patients undergoing PCI.</td>
<td>Myocardial infarction, or stroke, at 12 months after the index procedure.</td>
<td>2,993</td>
<td>78/39</td>
</tr>
<tr>
<td>Packer DL</td>
<td>Symptomatic AF, aged ≥65 years; or younger than 65 years with one or more risk factors for stroke.</td>
<td>Disabling stroke, serious bleeding, or cardiac arrest.</td>
<td>2,204</td>
<td>190/125</td>
</tr>
<tr>
<td>Macdougall IC</td>
<td>Adults undergoing maintenance hemodialysis.</td>
<td>Nonfatal myocardial infarction, nonfatal stroke, or hospitalization for heart failure.</td>
<td>2,141</td>
<td>658/515</td>
</tr>
<tr>
<td>Mack MJ</td>
<td>Severe aortic stenosis and low surgical risk.</td>
<td>Stroke or rehospitalization at one year.</td>
<td>1,000</td>
<td>110/16</td>
</tr>
<tr>
<td>Yasuda S</td>
<td>Aged ≥75 years, diagnosis of atrial fibrillation (AF) and stable coronary artery disease (patients with AF who had undergone PCI or coronary-artery bypass grafting (CABG) &gt;1 year earlier or who had angiographically confirmed coronary artery disease not requiring revascularization).</td>
<td>Stroke, systemic embolism, myocardial infarction, unstable angina requiring revascularization.</td>
<td>2,215</td>
<td>210/114</td>
</tr>
<tr>
<td>Schuetz P</td>
<td>Nutritional risk and with an expected length of hospital stay of &gt;4 days.</td>
<td>Admission to intensive care, non-elective hospital readmission, major complications, or decline in functional status at 30 days</td>
<td>2,088</td>
<td>504/173</td>
</tr>
<tr>
<td>Stone GW</td>
<td>Left main coronary artery disease of low or intermediate anatomical complexity.</td>
<td>Stroke or myocardial infarction.</td>
<td>1,905</td>
<td>379/208</td>
</tr>
<tr>
<td>Zenati MA</td>
<td>Patients undergoing CABG.</td>
<td>Nonfatal myocardial infarction or repeat revascularization.</td>
<td>1,150</td>
<td>169/83</td>
</tr>
<tr>
<td>Iversen K</td>
<td>Adults in stable condition who had endocarditis on the left side of the heart.</td>
<td>Unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen, from randomization until six months after antibiotic treatment was completed.</td>
<td>400</td>
<td>42/20</td>
</tr>
<tr>
<td>First Author</td>
<td>Characteristics of the study population</td>
<td>Composite elements other than all-cause mortality</td>
<td>n</td>
<td>Composite/Deaths</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----</td>
<td>-----------------</td>
</tr>
<tr>
<td>Lanz J</td>
<td>Aged ≥75 years, undergoing transfemoral</td>
<td>Any stroke, life-threatening or disabling bleeding, major</td>
<td>739</td>
<td>147/12</td>
</tr>
<tr>
<td></td>
<td>Transcatheter aortic valve replacement (TAVR) for treatment of symptomatic severe aortic stenosis, and who were deemed to be at increased surgical risk.</td>
<td>vascular complications, coronary artery obstruction requiring intervention, acute kidney injury (stage 2 or 3), rehospitalization for valve-related symptoms or congestive heart failure, valve-related dysfunction requiring repeat procedure, moderate or severe prosthetic valve regurgitation, or prosthetic valve stenosis within 30 days of the procedure.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onland W</td>
<td>Preterm infants (gestational age &lt;30 weeks or birth weight &lt;1,250 g) who were ventilator dependent.</td>
<td>Bronchopulmonary dysplasia (BPD) assessed at 36 weeks postmenstrual age.</td>
<td>372</td>
<td>268/73</td>
</tr>
<tr>
<td>Nagel E</td>
<td>Typical angina and either ≥2 cardiovascular risk factors or a positive exercise treadmill test.</td>
<td>Nonfatal myocardial infarction, or target-vessel revascularization within one year.</td>
<td>851</td>
<td>31/6</td>
</tr>
<tr>
<td>Ho KM</td>
<td>Severely injured patients who had a contraindication to anticoagulant agents.</td>
<td>Symptomatic pulmonary embolism at 90 days after enrollment.</td>
<td>240</td>
<td>34/27</td>
</tr>
<tr>
<td>Van Spall HGC</td>
<td>Adult patients hospitalized for heart failure (HF).</td>
<td>All-cause readmission or emergency department (ED) visit at three months.</td>
<td>2,494</td>
<td>1,243/247</td>
</tr>
<tr>
<td>Kozhuharov N</td>
<td>Patients hospitalized for acute heart failure.</td>
<td>Rehospitalization for AHF at 180 days.</td>
<td>788</td>
<td>228/116</td>
</tr>
<tr>
<td>Nguyen HQ</td>
<td>40 years or older who had any acute care use related to chronic obstructive pulmonary disease in the previous 12 months.</td>
<td>All-cause hospitalizations, observation stays, emergency department visits.</td>
<td>2,707</td>
<td>1,747/234</td>
</tr>
<tr>
<td>Vousden N</td>
<td>Users of maternity care.</td>
<td>Eclampsia or emergency hysterectomy.</td>
<td>536,233</td>
<td>4,067/998</td>
</tr>
<tr>
<td>Positive controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dargie HJ (CAPRICORN)</td>
<td>A proven acute myocardial infarction and a left-ventricular ejection fraction of &lt;=40%.</td>
<td>Hospital admission for cardiovascular problems.</td>
<td>1,959</td>
<td>705/267</td>
</tr>
<tr>
<td>Mentzer R. Jr (EXPEDITION)</td>
<td>High-risk CABG surgery patients.</td>
<td>Myocardial infarction (MI).</td>
<td>5,761</td>
<td>1,064/106</td>
</tr>
</tbody>
</table>
### Table 3. Relative risks of composite and death, and BACO index in clinical trials.

<table>
<thead>
<tr>
<th>First author</th>
<th>RR&lt;sub&gt;c&lt;/sub&gt; (95% CI)</th>
<th>RR&lt;sub&gt;d&lt;/sub&gt; (95% CI)</th>
<th>BACO index (95% CI)</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holm NR</td>
<td>1.51 (1.22 to 1.87)</td>
<td>1.09 (0.75 to 1.57)</td>
<td>4.85 (-14.7 to 24.4)</td>
<td>0.70</td>
</tr>
<tr>
<td>Katheria A</td>
<td>1.46 (0.85 to 2.51)</td>
<td>1.14 (0.58 to 2.23)</td>
<td>2.84 (-8.44 to 14.1)</td>
<td>0.75</td>
</tr>
<tr>
<td>Schüpke S</td>
<td>1.34 (1.08 to 1.66)</td>
<td>1.23 (0.91 to 1.66)</td>
<td>1.41 (-0.12 to 2.95)</td>
<td>0.60</td>
</tr>
<tr>
<td>Brott TG</td>
<td>0.71 (0.59 to 0.85)</td>
<td>0.78 (0.42 to 1.47)</td>
<td>1.41 (-2.08 to 4.9)</td>
<td>0.82</td>
</tr>
<tr>
<td>Tomanik M</td>
<td>0.73 (0.52 to 1.03)</td>
<td>0.74 (0.49 to 1.13)</td>
<td>1.05 (0.22 to 1.89)</td>
<td>0.90</td>
</tr>
<tr>
<td>Wu Y</td>
<td>0.87 (0.78 to 0.97)</td>
<td>0.87 (0.78 to 0.97)</td>
<td>1 (0.8 to 1.21)</td>
<td>0.99</td>
</tr>
<tr>
<td>Hahn JY</td>
<td>1.17 (0.75 to 1.81)</td>
<td>1.17 (0.63 to 2.19)</td>
<td>1 (-1.85 to 3.85)</td>
<td>1</td>
</tr>
<tr>
<td>Packer DL</td>
<td>0.87 (0.66 to 1.14)</td>
<td>0.86 (0.61 to 1.2)</td>
<td>0.89 (-0.31 to 2.08)</td>
<td>0.85</td>
</tr>
<tr>
<td>Macdougall IC</td>
<td>0.91 (0.8 to 1.03)</td>
<td>0.88 (0.75 to 1.02)</td>
<td>0.74 (0.22 to 1.26)</td>
<td>0.32</td>
</tr>
<tr>
<td>Mack MJ</td>
<td>0.61 (0.04 to 0.88)</td>
<td>0.45 (0.16 to 1.28)</td>
<td>0.62 (-0.16 to 1.4)</td>
<td>0.34</td>
</tr>
<tr>
<td>Yasuda S</td>
<td>0.74 (0.57 to 0.96)</td>
<td>0.56 (0.39 to 0.82)</td>
<td>0.53 (0.21 to 0.86)</td>
<td>0.0048</td>
</tr>
<tr>
<td>Schuetz P</td>
<td>0.84 (0.72 to 0.98)</td>
<td>0.72 (0.54 to 0.96)</td>
<td>0.52 (0.07 to 0.98)</td>
<td>0.04</td>
</tr>
<tr>
<td>Stone GW</td>
<td>1.16 (0.97 to 1.4)</td>
<td>1.35 (1.04 to 1.75)</td>
<td>0.51 (0.07 to 0.94)</td>
<td>0.03</td>
</tr>
<tr>
<td>Zenati MA</td>
<td>1.12 (0.84 to 1.48)</td>
<td>1.25 (0.82 to 1.89)</td>
<td>0.5 (-0.45 to 1.44)</td>
<td>0.30</td>
</tr>
<tr>
<td>Iversen K</td>
<td>1.35 (0.75 to 2.4)</td>
<td>1.88 (0.76 to 4.6)</td>
<td>0.47 (-0.24 to 1.18)</td>
<td>0.15</td>
</tr>
<tr>
<td>Lanz J</td>
<td>1.43 (1.06 to 1.92)</td>
<td>2.96 (0.81 to 10.8)</td>
<td>0.33 (-0.1 to 0.76)</td>
<td>0.002</td>
</tr>
<tr>
<td>Onland W</td>
<td>0.95 (0.84 to 1.08)</td>
<td>0.65 (0.42 to 0.99)</td>
<td>0.11 (-0.17 to 0.39)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Nagel E</td>
<td>0.96 (0.48 to 1.91)</td>
<td>2.04 (0.38 to 11.1)</td>
<td>-0.06 (-1.1 to 0.98)</td>
<td>0.04</td>
</tr>
<tr>
<td>Ho KM</td>
<td>0.97 (0.52 to 1.8)</td>
<td>1.41 (0.68 to 2.9)</td>
<td>-0.1 (-2.11 to 1.91)</td>
<td>0.28</td>
</tr>
<tr>
<td>Van Spall HGC</td>
<td>0.98 (0.91 to 1.06)</td>
<td>1.03 (0.81 to 1.3)</td>
<td>-0.63 (-7.64 to 6.38)</td>
<td>0.65</td>
</tr>
<tr>
<td>Kozhuharov N</td>
<td>1.1 (0.88 to 1.37)</td>
<td>0.94 (0.67 to 1.31)</td>
<td>-1.48 (-12 to 9.05)</td>
<td>0.64</td>
</tr>
<tr>
<td>Nguyen HQ</td>
<td>1.02 (0.96 to 1.07)</td>
<td>0.99 (0.78 to 1.27)</td>
<td>-2.27 (-88.3 to 83.7)</td>
<td>0.94</td>
</tr>
<tr>
<td>Vousden N</td>
<td>0.92 (0.86 to 0.97)</td>
<td>1.04 (0.92 to 1.18)</td>
<td>-2.35 (-11.2 to 6.51)</td>
<td>0.46</td>
</tr>
</tbody>
</table>

**Positive Controls:**
- Dargie HJ (CAPRICORN) 0.94 (0.84 to 1.06) 0.78 (0.62 to 0.97) 0.24 (-0.15 to 0.64) <0.001
- Mentzer R. Jr (EXPEDITION) 0.82 (0.73 to 0.91) 1.53 (1.04 to 2.26) -0.48 (-1.03 to 0.08) <0.001

* p-value for the null hypothesis of the BACO index is equal to one.

RR<sub>c</sub>: Relative Risk for the composite outcome; RR<sub>d</sub>: Relative Risk for any-cause death; BACO index: Ln(RR<sub>c</sub>)/Ln(RR<sub>d</sub>)

Three other works were suggestive of a significant BACO, in which the RR of the composite outcome was closer to null value than the RR of mortality [33,34,39]. These studies included the clinical trials by Schuetz et al. (BACO index: 0.52; 95%CI: 0.07 to 0.98; p=0.04); Stone et al. (BACO index: 0.52; 95% CI: 0.07 to 0.98, p=0.03); and Nagel et al., which exhibited a negative point estimate of the BACO index (-0.06; 95% CI: -1.1 to 0.98, p=0.04).

In the positive controls, the CAPRICORN showed that carvedilol, compared with placebo, was not significantly associated with the composite outcome (RR = 0.94; 95% CI: 0.84 to 1.06) but was associated with a 22% reduction in mortality (BACO index: 0.24; 95% CI: -0.15 to 0.64). On the other hand, in the EXPEDITION study, the use of cariporide was associated with an 18% lower incidence of the composite outcome, but with a 53% higher mortality, compared with placebo (BACO index: -0.48; 95% CI: -1.03 to 0.08). The BACO indices of these two studies were significantly lower than one (p <0.001).
4. Discussion

Composite outcomes can make the interpretation of results challenging [1,8]. Differential effects on their less critical but more frequent components may result in a misleading impression about the impact of a treatment [2,45]. Therefore, it has been recommended that if there is a great variation between the effects on the components, the composite outcome should be abandoned [8].

However, assessing this is difficult, considering the asymmetric distribution of association measures and random variations. Some authors have evaluated the differences between the associations of both the composite outcome and mortality based on disagreement in statistical significance [1]. This type of comparison is biased because of the fatal outcome is a sub-element of the composite and will always have fewer events [4]. Therefore, there will be less power to evaluate an association with mortality.

The BACO index summarizes the relationship between the associations of the composite outcome and its most critical endpoint. Being based on the logarithms, the comparison is more consistent with the association measure distributions. On the other hand, the integration into a single index allows a unique statistical test for the null hypothesis.

Other authors stated they had planned to calculate a ratio between the efficacy for the composite outcome and that for mortality [7]. However, they considered that it would be problematic because the observations were not independent, as the death contributes to the composite. In that sense, the methodology proposed to obtain the BACO index solves this problem by considering overlapping observations [16,46,47].

Despite this, a limitation for an index based on the ratio of efficacy measures is that denominators close to zero lead to unstable or seemingly inflated results [7,48]. Hence, a BACO index should not be computed together with another whose reference effect is different. In other words, comparisons between BACO indices only make sense to contrast two or more composite outcomes when they consider the same reference value (e.g., of RR for mortality). In other circumstances, it is prudent to interpret the BACO indices only by classifying them into descriptive categories (overestimation, underestimation, and effect inversion) and considering the null hypothesis test.

In practice, the BACO index proved to be a simple measure to validate the composite outcome in clinical trials. Three of the analyzed studies had an index significantly lower than one, suggesting that the composite outcome underestimated the association between the intervention and the prognosis. Additionally, three other studies were suggestive of a similar trend, i.e., the composite outcomes seemed to dilute the associations that were stronger for mortality.

In the positive controls, the BACO index was significantly different from one. These studies have been well recognized as examples of bias associated with the composite outcome. After many discussions about the results of the CAPRICORN and EXPEDITION trials, their composite outcomes were disregard and the decisions based on the effects on all-cause mortality [9,10]. This would be consistent with the results of the BACO index.

Facing the need for objective guidelines, the BACO index may be a statistical tool to help interpret composite outcomes. Moreover, its application could be adapted according to the research context. For example, the level of significance could be adjusted depending on the desired sensitivity to identify a BACO. Furthermore, we can calculate the index from adjusted regression coefficients in observational studies that require controlling confusion.
This work is based on the expectation that the composite outcome must be in the same direction as its most relevant component. This is important when what is sought is that the composite outcome offers more events to increase power. In other cases, the composite outcome may not necessarily have a similar association magnitude as mortality. For example, when an intervention is expected to improve the quality of life without affecting total survival. However, even in these cases, a significant BACO could lead to reconsidering the necessity of a composite outcome or disaggregating the estimations for the component endpoints adequately.

5. Conclusions

The BACO index calculation could be incorporated into the analysis plan of clinical studies. Thus, based on a predefined rule, researchers could make impartial decisions about maintaining or replacing a composite outcome as the primary endpoint. Even if the researchers decide to base their conclusions on the composite outcome, a significant BACO should lead to caution that the association of the composite is stronger, weaker, or even opposite than that of its most critical component.

Acknowledgments

I thank my children, Alejandro and Ana Maria, for helping me organize the bibliographic material. I also thank Dr. Kai Arzheimer, for his guidance on Stata commands; and Alene Alder-Rangel, for her help with the English edition of the article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. FADQ was granted a fellowship for research productivity from the Brazilian National Council for Scientific and Technological Development – CNPq, process/contract identification: 312656/2019-0. Conflicts of interest: None.

References

12


Appendix A. On the formula to estimate the variance of the BACO index.

Let \( \varphi_c \) and \( \varphi_d \) be natural logarithms (\( \ln \)) of the relative risks for the composite (\( RR_c \)) and the target outcome (\( RR_d \)), respectively. According the following contingency table:

<table>
<thead>
<tr>
<th>Table A.1. Absolute frequency distribution of the composite outcome and its most critical component (target) in two treatment groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target outcome (e.g. death)</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>Treatment group</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

We have that:

\[
RR_c = \frac{n_{t11}/n_{t..}}{n_{c1}/n_{c..}}
\]

\[
RR_d = \frac{n_{t11}/n_{t..}}{n_{c11}/n_{c..}}
\]

\[
BACO\ index = \theta = \frac{\varphi_c}{\varphi_d} = \frac{\ln (RR_c)}{\ln (RR_d)}
\]

Consider the formula of the Taylor series estimator of the variance of an estimator of a ratio, \( v(\hat{R}) \), of random variables (\( Y \) and \( X \), being \( \hat{R} = \hat{Y}/\hat{X} \)), according to which:

\[
v(\hat{R}) = \hat{R}^2 \left[ \frac{v(\hat{Y})}{\hat{Y}^2} + \frac{v(\hat{X})}{\hat{X}^2} - 2 \frac{c(\hat{Y}, \hat{X})}{\hat{Y}\hat{X}} \right]
\]

Where \( v(\hat{Y}) \) and \( v(\hat{X}) \) are variances of the estimators of \( Y \) and \( X \), respectively, and \( c(\hat{Y}, \hat{X}) \) is the covariance between them. Similarly, to calculate the variance BACO index, \( v(\theta) \), we have:

\[
v(\theta) = \theta^2 \left[ \frac{v(\varphi_c)}{\varphi_c^2} + \frac{v(\varphi_d)}{\varphi_d^2} - 2 \frac{c(\varphi_c, \varphi_d)}{\varphi_c\varphi_d} \right]
\]

Where \( v(\varphi_c) \) and \( v(\varphi_d) \) are the variances of \( \varphi_c \) and \( \varphi_d \), respectively, and \( c(\varphi_c, \varphi_d) \) is their covariance.

Now, let a term \( \varphi_{c-d} \) such as:

\[
\varphi_{c-d} = \varphi_c - \varphi_d = \ln \left( \frac{RR_c}{RR_d} \right) = \ln \left( \frac{n_{t11}/n_{t11}}{n_{c1}/n_{c11}} \right) = \ln \left( \frac{n_{c11}/n_{c1}}{n_{t11}/n_{t1}} \right)
\]

Consequently, variance estimators of \( \varphi_c \), \( \varphi_d \), and \( \varphi_{c-d} \) would be respectively:

\[
v(\varphi_c) = \frac{1}{n_{t1}} + \frac{1}{n_{c1}} - \frac{1}{n_{t..}} - \frac{1}{n_{c..}}
\]
\[
\nu(\varphi_d) = \frac{1}{n_{t11}} + \frac{1}{n_{c11}} - \frac{1}{n_{t.c}} - \frac{1}{n_{c.c}} \\
\nu(\varphi_{c-d}) = \frac{1}{n_{c11}} + \frac{1}{n_{t11}} - \frac{1}{n_{c,1}} - \frac{1}{n_{t,1}}
\]

By considering the additive properties of variance and covariance, we have that

\[
\nu(\varphi_c) = \nu(\varphi_d + \varphi_{c-d}) \\
\nu(\varphi_c) = \nu(\varphi_d) + \nu(\varphi_{c-d}) + 2c(\varphi_d, \varphi_{c-d}) \\
c(\varphi_d, \varphi_{c-d}) = \frac{\nu(\varphi_c) - \nu(\varphi_d) - \nu(\varphi_{c-d})}{2} \\
c(\varphi_c, \varphi_d) = \nu(\varphi_d) + c(\varphi_d, \varphi_{c-d}) \\
c(\varphi_c, \varphi_d) = \nu(\varphi_d) + \left( \frac{\nu(\varphi_c) - \nu(\varphi_d) - \nu(\varphi_{c-d})}{2} \right)
\]

Moreover, because

\[
\nu(\varphi_d) - \nu(\varphi_{c-d}) = \left( \frac{1}{n_{t11}} + \frac{1}{n_{c11}} - \frac{1}{n_{t,c}} - \frac{1}{n_{c,c}} \right) - \left( \frac{1}{n_{c11}} + \frac{1}{n_{t11}} - \frac{1}{n_{c,1}} - \frac{1}{n_{t,1}} \right) \\
\nu(\varphi_d) - \nu(\varphi_{c-d}) = \frac{1}{n_{t,1}} + \frac{1}{n_{c,1}} - \frac{1}{n_{t,c}} - \frac{1}{n_{c,c}} = \nu(\varphi_c)
\]

Then,

\[
c(\varphi_c, \varphi_d) = \frac{\nu(\varphi_c) + \nu(\varphi_c)}{2} = \nu(\varphi_c)
\]

Consequently, the estimator of the variance of the BACO index can be

\[
\nu(\theta) = \theta^2 \left[ \frac{\nu(\varphi_c)}{\varphi_c^2} + \frac{\nu(\varphi_d)}{\varphi_d^2} - 2 \frac{\nu(\varphi_c)}{\varphi_c \varphi_d} \right]
\]

A 95% confidence interval for the BACO index is

\[
\theta \pm 1.96\sqrt{\nu(\theta)}
\]

And the statistic Z with normal distribution to test the null hypothesis (BACO index = 1) is

\[
Z = \frac{\theta - 1}{\sqrt{\nu(\theta)}}
\]
Appendix B. Stata code to estimate and test the BACO index by combining regression parameters.

Suppose we have a dataset with exposure or treatment variable “t”, composite outcome variable “c”, and the most critical endpoint variable “d” (i.e., death). For this example, take the data of the CAPRICORN study, which compared 975 patients allocated to experimental intervention vs. 984 patients in the control group [20]. The composite outcome occurred in 340 and 365 patients, including 116 and 151 deaths, respectively. Consequently, the RRs were 0.94 (95% CI: 0.84 – 1.06) for the composite outcome and 0.78 (95% CI: 0.62 – 0.97) for death. Before calculating the BACO index based on the relative risks, we should store and integrate the regression coefficients as described in the code that follows:

```
poisson c t
est store coefcomp
poisson d t
est store coefdeath
suest coefcomp coefdeath
```

In this example, the coefficients generated by the regression models were stored under the names “coefcomp” and “coefdeath”. Then, by using “suest” (seemingly unrelated estimation) command, we combine the stored parameter estimates and associated (co)variance matrices. Now we can use “nlcom” as described in the code that follows. The output (presented in the box) will include the estimate of the BACO index, its standard error, and a 95% confidence interval.

```
nlcom [coefcomp_c]_b[t]/([coefdeath_d]_b[t])
```

|   | Coef. | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|---|-------|-----------|------|------|---------------------|
| _nl_1 | .2426837 | .2023317 | 1.20 | 0.230 | -.1538791    .6392466 |

Please note that the previous p-value (0.23) refers to the null hypothesis of the BACO index equal to zero (therefore, it should not be considered). To calculate the p-value for the null hypothesis of the BACO index equals one, we can use the “testnl” command:

```
testnl ([coefcomp_c]_b[t]/([coefdeath_d]_b[t]))=1
```

(1)  ([coefcomp_c]_b[t]/([coefdeath_d]_b[t]))=1

    chi2(1) = 14.01
    Prob > chi2 = 0.0002
Appendix C. Comparison of methods to calculate the standard error of the BACO index.

Consider the four hypothetic composite outcomes described in the simulation (section 2.2). Table C.1 presents the standard error of each BACO index, $\sqrt{\nu(\theta)}$, calculated both “directly” from the contingency tables, using formulas described in Appendix A, and by the nonlinear combination of regression parameters (NLCRP), using the tools described in Appendix B.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RR</th>
<th>BACO index</th>
<th>Standard Error Directly</th>
<th>Standard Error NLCRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite 1 (unbiased)</td>
<td>0.6</td>
<td>1.00</td>
<td>0.2768</td>
<td>0.2769</td>
</tr>
<tr>
<td>Composite 2 (overestimated effect)</td>
<td>0.3</td>
<td>2.36</td>
<td>0.6222</td>
<td>0.6224</td>
</tr>
<tr>
<td>Composite 3 (underestimated effect)</td>
<td>0.8</td>
<td>0.44</td>
<td>0.1655</td>
<td>0.1655</td>
</tr>
<tr>
<td>Composite 4 (inverted effect)</td>
<td>2.0</td>
<td>-1.36</td>
<td>0.5710</td>
<td>0.5712</td>
</tr>
</tbody>
</table>

In these examples, the two methods were virtually equivalent, and differences between them were observed only from the fourth decimal place (at the order of ten thousandths).