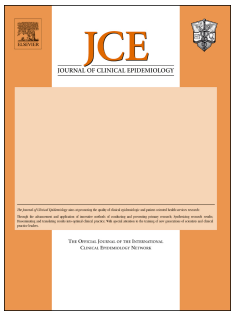
# Accepted Manuscript




Criteria for use of Composite Endpoints for Competing Risks – A Systematic Survey of the Literature with Recommendations

Veena Manja, Siwar AlBashir, Gordon Guyatt

PII: S0895-4356(16)30787-9

DOI: 10.1016/j.jclinepi.2016.12.001

Reference: JCE 9287

To appear in: *Journal of Clinical Epidemiology*

Received Date: 13 February 2016

Revised Date: 5 October 2016

Accepted Date: 1 December 2016

Please cite this article as: Manja V, AlBashir S, Guyatt G, Criteria for use of Composite Endpoints for Competing Risks – A Systematic Survey of the Literature with Recommendations, *Journal of Clinical Epidemiology* (2017), doi: 10.1016/j.jclinepi.2016.12.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Title: Criteria for use of Composite Endpoints for Competing Risks – A Systematic Survey of the Literature with Recommendations

Authors:

Veena Manja[1,2]

Siwar AlBashir[1,3],

Gordon Guyatt[1]

Institutions:

1. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton Ontario, Canada
2. Department of Internal Medicine, Division of Cardiology, Veterans Affairs Medical Center, Buffalo NY, USA
3. Department of Internal Medicine, Division of Gastroenterology, McMaster University, Hamilton Ontario, Canada

Corresponding author:

Veena Manja, MBBS, MSc
Staff Cardiologist,
Veterans Affairs Medical Center,
3495 Bailey Avenue, Buffalo NY 14215, USA
Phone: 001 716 862 7314
Fax: 001 716 862 8640
Email: veenamanja@hotmail.com

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

**Topic:** Criteria for use of composite endpoints for competing risks – A Systematic Survey of the Literature with recommendations

**Abstract:**

Background: Composite end points are frequently used in reports of clinical trials. One rationale for the use of composite endpoints is to account for competing risks. In the presence of competing risks, the event rate of a specific event depends on the rates of other competing events. One proposed solution is to include all important competing events in one composite endpoint. Clinical trialists require guidance regarding when this approach is appropriate.

Objectives: To identify publications describing criteria for use of composite endpoints for competing risk and to offer guidance regarding when a composite endpoint is appropriate on the basis of competing risks.

Data Sources: We searched MEDLINE, CINAHL, EMBASE, The Cochrane's Central & Systematic Review databases including the Health Technology Assessment database and the Cochrane's Methodology register from inception to April 2015, and candidate textbooks, to identify all articles providing guidance on this issue.

Study Selection and Data Extraction: Eligible publications explicitly addressed the issue of a composite outcome to address competing risks. Two reviewers independently screened the titles and abstracts for full text review; independently reviewed full text publications; and abstracted specific criteria authors offered for use of composite endpoints to address competing risks.

Results: Of 63645 titles and abstracts, 166 proved potentially relevant of which 43 publications were included in the final review. Most publications note competing risks as a reason for using composite endpoints without further elaboration. None of the articles or textbook chapters provide specific criteria

for use of composite endpoints for competing risk. Some advocate using composite endpoints to avoid bias due to competing risks and others suggest that composite endpoints seldom or never be used for this purpose. We recommend using composite endpoints for competing risks only if the competing risk is plausible and if it occurs with sufficiently high frequency to influence the interpretation of the effect of intervention on the endpoint of interest. These criteria will seldom be met. Review of heart failure trials published in the New England Journal of Medicine revealed that many of them use the composite endpoint of death or hospitalization; none of the trials, however, satisfied our criteria.

Conclusion: The existing literature fails to provide clear guidance regarding use of composite endpoint for competing risks. We recommend using composite endpoints for competing risks only if the competing risk is plausible and if it occurs sufficiently often.

Key words: 'Competing risks', 'composite end-points'

"What is new"

A systematic survey of the literature revealed limited guidance on when to use composite end-points in the presence of competing risks.

We provide guidance on this topic and propose that composite outcomes be used to overcome the problem of competing risks only when

(i) Competing risk is plausible (i.e. understanding of the biology suggests that the intervention might realistically increase more serious events, thus misleadingly reduce the less serious)

(ii) The more serious outcome occurs frequently enough that, if the intervention truly increases its frequency – appreciably decreasing the possibility of the less serious outcome occurring – the result would be a misleading decrease in the less serious event.

ACCEPTED MANUSCRIPT

**Introduction:**

Clinical trialists often specify composite endpoints composite endpoint as their primary outcome. A composite endpoint combines in a single endpoint all patients who experience at least one event included in the composite. For example, a commonly used composite endpoint in the field of cardiology is a composite of death, myocardial infarction (MI) or stroke; this would include all patients who experienced any of these events. Reasons for the use of composite endpoints include increasing statistical power by increasing the number of events, simplifying the interpretation for patients (it may be easier for patients to consider 1 risk estimate rather than several in considering risks and benefits of interventions for decision making), and accounting for competing risks.

Competing risks is a concern in randomized trials because of the possibility that an intervention may result in an apparent decrease in a less serious endpoint (e.g. MI) as a result of the intervention increasing a more serious endpoint (e.g. death). In other words, there is a competing risk if the intervention results in the death of individuals, some of whom, had they lived, would have experienced an MI.

One suggestion for dealing with the problem of competing endpoints is to construct a composite outcome that accounts for all competing risks in one outcome measure (e.g. a composite of MI and death). There are, however, concerns with the use of composite outcomes including challenges in interpretation, in particular making the impact of the intervention appear more important than it really is. Consider, for instance, if an intervention in fact has no impact on death, but does decrease the incidence of MI. Providing a single relative risk reduction for the composite may suggest to many that the intervention reduces both death and MI, and does so to the same degree, resulting in an overestimation of the impact or importance of the intervention on death and a possible

underestimation of the importance on MI. The greater the gradient in importance between components, the greater is the seriousness of such a misinterpretation. For example, the gradient between death and percutaneous coronary interventions is greater than the gradient between death and MI.

Recent publications have highlighted the frequency of the use of composite end points in published trials and have underscored concerns related to this practice (1-6). Consistent findings of these studies has been that clinical trialists very frequently – particularly in cardiology - choose composites as their primary outcomes, that components often include a large gradient of importance, that the less important endpoints typically occur more frequently than the more important endpoints, and that relative effects often differ substantially between components (with relative effects typically larger for less important outcomes).

These results suggest two fundamental problems with the use of composite endpoints. The first is an issue of interpretation: are clinicians to assume that relative effects on the composites apply to each of the components, and the absolute impact on components should be calculated accordingly, or make no such assumption and look at the composite without making any inferences about distribution of effects across components? Second, when the more important components contribute few outcomes and/or the effect is less in these components, there is high risk of spurious inferences from trials with composite endpoints, with treatment effects appearing more important than they actually are. Thus, confident interpretation of composite endpoints requires relatively small gradients of importance to patients and similar relative risk reductions across components (6).

The difficulties in interpretation, and risk of misinterpretation, can arise either if the composite is chosen to increase power or to address competing risks. The Diabetes REduction Assessment with ramipril and rosiglitazone Medication (DREAM) randomized trial (7) highlights the problem in the latter context. The

ACCEPTED MANUSCRIPT

DREAM trial implemented strategies to minimize risk of bias and enrolled 5269 participants with impaired glucose tolerance, assigned them to a hypoglycemic drug, rosiglitazone, or placebo, and documented the impact on a primary endpoint a composite of incident diabetes or death from any cause. The methods section of the paper justifies the composite using the competing endpoint criteria: 'death was included to account for the possibility that diabetes might develop at a different rate in individuals who die than in those who survive'.

Although rosiglitazone reduced the outcome of death or diabetes (306 events in the rosiglitazone group, 686 in placebo, hazard ratio [HR] 0.40 [95% confidence interval [CI] 0.35-0.46] p<0.0001), the drug had no effect on all-cause mortality (30 deaths with rosiglitazone, 33 with placebo HR 0.91 [95% CI, 0.55-1.49] p =0.7). Thus, a decrease in diabetes accounted for all the drug's impact on the composite. The authors nevertheless concluded that 'this large, prospective, blinded international clinical trial shows that 8 mg of rosiglitazone daily, together with lifestyle recommendations, substantially reduces the risk of diabetes or death by 60% in individuals at high risk for diabetes', potentially leading readers to infer that rosiglitazone decreased mortality – clearly a problematic inference (6, 8).

Given the problems of interpretation, it may be that composite endpoints should not be used gratuitously, and criteria for their parsimonious use should be available. We therefore undertook a systematic survey of the literature to identify publications that provide criteria for use of composite endpoints for competing risks. Considering the findings, we offer guidance for use of composite endpoints to address competing risks.

**Methods:**

**Eligibility Criteria:** Eligible articles explicitly addressed the appropriateness of use of a composite outcome to address competing risks in the context of medical interventions in any area.

ACCEPTED MANUSCRIPT

**Search Strategy:** We searched the following databases: MEDLINE, CINAHL, EMBASE, The Cochrane's Central & Systematic Review databases including the Health Technology Assessment database and the Cochrane's Methodology register from inception to April 2, 2015. The search strategy for MEDLINE involved various combinations of the following keywords, using the search field of keyword, abstracts, MeSH headings, exploded subject headings, publication type, text word and title: 'composite endpoint', 'composite outcome', 'combined endpoint', combined outcome', 'competing causes', 'competing endpoint' and 'competing risk'. We applied no language restrictions or date limits, used similar search strategies for the other databases, and considered articles cited in previous reviews, including cross-references and bibliographic citations of relevant publications. The website books.google.com provided the source for textbooks addressing this topic; we scanned relevant texts for chapters addressing competing risks and composite outcome.

**DATA COLLECTION AND ANALYSIS**:

**Study selection**: Two reviewers (VM and SA) reviewed the titles and abstracts of potentially relevant publications. If either reviewer considered the article possibly eligible, it was included for the full text review. The two reviewers assessed the full text articles of the selected citations for eligible studies and resolved disagreements by discussion. A $K \geq 0.65$ was chosen a priori to indicate adequate agreement.

**Data Extraction:** The two reviewers abstracted the title of the article, year and journal of publication, primary purpose of the paper, and methods proposed to address the problem of competing risks. The reviewers noted, in eligible publications, description of criteria for using composite endpoints for competing events and extracted suggested criteria verbatim.

**Data Analysis:** We reviewed and summarized the extracted data noting in particular proposed criteria addressing the use of composite endpoints to overcome the problem of competing risks.

ACCEPTED MANUSCRIPT

**Development of Guidance:** For a number of years, one of us (GG) has taught in a graduate methodology course on advanced RCT methods that includes a session devoted to composite endpoints. Yearly discussions of the issue led to the development of candidate criteria for use of a composite endpoint to deal with the problem of competing endpoints. The authors of the present article reviewed and discussed the criteria, arrived at preliminary definitive criteria, and applied these to a number of example articles. Textbooks (9) and review articles (10) frequently cite the example of death as a competing risk to hospitalization for worsening heart failure (HF) in RCTs of HF management strategies and recommend using composite endpoints to overcome the issues of competing risks. We identified landmark HF trials published in the New England Journal of Medicine (NEJM) and applied the criteria to these trials (11).

**Results:**

ACCEPTED MANUSCRIPT

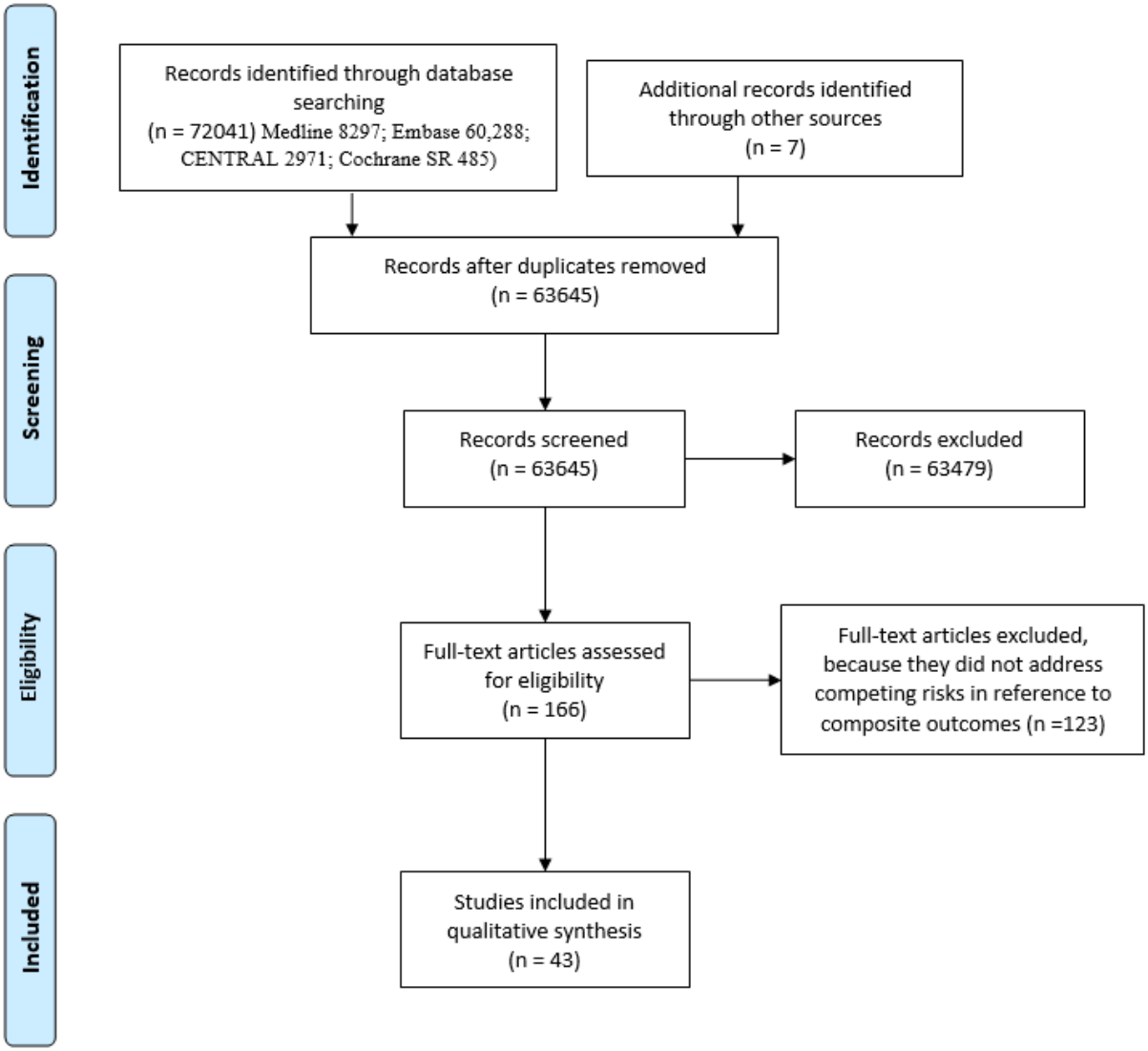



Figure 1: PRISMA flow diagram

As shown in figure 1, of 63645 citations, 166 were included in the final full text review of which 43 proved eligible for inclusion in this survey. A kappa statistic, 0.92, indicated near-perfect inter-rater agreement. The 43 eligible publications included 27 full articles, 9 letters, 5 textbooks and 2 editorials. Of these, 11 addressed statistical methods for addressing competing risks, 9 addressed non-statistical aspects of competing risk methodology, and the remainder were review articles. None of the journal articles or textbooks suggested clear and explicit criteria for when to use, and when not to use,

ACCEPTED MANUSCRIPT

composite endpoints to address the problem of competing risks. One textbook devoted to multiple analyses in clinical trials listed the 'principles for the use of combined endpoints' without mentioning competing risks (12), while another suggested that composite endpoints can help avoid survivor bias when there is competing risk without specifying criteria for composite endpoint use (9). Other textbooks addressed statistical formulations and considerations for competing risk (13) in the context of 'event-free survival' as the composite endpoint (using non-cancer mortality as the competing risk to cancer recurrence as an example) or noted that death from other causes can be a competing risk (14) without specifying criteria. The User's Guide to Medical Literature (15) notes 'another benefit to justify the use of composite endpoints is to avoid competing risks in outcome assessment' without specifying criteria for composite endpoint use to address competing risks.

Many authors suggest that using composite endpoints can avoid the bias from competing risk while others have suggested that this may be problematic and confuse rather than usefully inform. In their response to a letter regarding composite endpoints (16), DeMets and Califf argue in favor of using composite endpoints to overcome the problem of competing risk:,

> 'Although a treatment may reduce the risk of nonfatal events, such treatment comparisons become problematic when many deaths also occur. Patients who die without having the nonfatal event cannot be counted in the event-free group or in the group with the event, because they might have had the event had they lived longer. Thus, composite endpoints must include mortality to capture the overall effects of treatment. The statistical theory behind this approach is contained in literature on competing risk analysis'.

Similarly, Neaton and colleagues (10) note in 2005 that

> 'Another reason that composites are used is to avoid the problem of competing risks. For example, in a trial of patients with advanced heart failure, an end point of hospitalization for

ACCEPTED MANUSCRIPT

> heart failure would be criticized because it does not account for mortality. The censoring of deaths in such a trial (patients who died without hospitalization) is probably ‘‘informative’’ because a patient who is censored for death is likely not at the same risk of hospitalization (had they survived) as a patient who survived as long and is still at risk for hospitalization. If censoring because of death, or reasons for it, varied by treatment group, the estimate of the treatment effect would be biased’.

In 2008, Song and colleagues (17) wrote,

> ‘Another rationale for using the composite endpoint is to account for mortality. This data structure is the so-called competing risks: death can terminate the occurrence of other diseases, whereas other diseases’ occurrence do not preclude the occurrence of death. Combining deaths into the composite endpoint is a solution to avoid the competing risks problems due to death.’

Other authors suggest that composite endpoints in the presence of competing risks can be misleading. Lim and colleagues (5) note,

> ‘When death is part of a composite outcome (98% of surveyed trials), competing risks can influence the results because patients who die cannot have any further end points. The extent to which the results are affected increases as the magnitude of imbalance in the number of deaths in the 2 groups increases. One can account for death with time-to-event analysis (survival), but we recommend the simultaneous reporting of each individual end point so that readers can ascertain its individual contribution’

Several authors (18-20) have suggested scoring system or hierarchical ranking as a solution to the problem of including endpoints of varying importance to patients when composite endpoints have to be used for competing risks. Such an approach requires value judgments that few seem comfortable making: the solution has very seldom been implemented.

ACCEPTED MANUSCRIPT

In summary, although many authors have highlighted the problem of competing risks, have raised the possibility of using a composite endpoint in the presence of competing risks, and have addressed statistical considerations that arise in the context of composite endpoints, no author has previously suggested clear guidance for the appropriate use of a composite endpoint in the setting of competing risks in randomized trials.

**Guidance, and its application:**

As in any use of composite endpoints, authors must explicit report the frequency of all components in both the intervention and control groups. Competing risks will, however, be a problem in the context of randomized trials only if the intervention reduces the likelihood of the less serious outcome through an increase in the more serious outcome. Even if this is to some extent the case, the competing risk will not result in serious distortion if the frequency of the more serious event is low and is eclipsed by the much higher frequency of the less serious event.

Using a composite endpoint to overcome the problem of competing risks will be misleading if there is a large gradient in the importance of the component endpoints, the more important endpoints occur less frequently, and the impact of the intervention is restricted to the components of less importance. This risk of misleading composite endpoints dictates that a composite endpoint be used to address competing risks only if the likelihood that a competing risk situations exists, and that this competing risk will lead to appreciable distortions in results.

Therefore, we propose that composite outcomes be used to overcome the problem of competing risks only when the following are the case: (i) Competing risk is plausible (i.e. our understanding of the biology suggests that the intervention might realistically increase more serious events, thus misleadingly reduce the less serious). With respect to this first criterion, we note that it becomes stronger if the authors of the study using competing risks have formally documented this risk of the intervention,

including in the informed consent document, the study protocol, and the data monitoring plan. (ii) The relative frequency of occurrence of the more serious outcome compared to the less serious outcome is great enough that, if the intervention truly increases the risk of the more serious event, it would misleadingly appear to decrease the less serious event. With respect to how high the ratio of the more to the less serious outcome needs to be to raise concern, this will depend on the absolute frequency of the anticipated magnitude of the positive intervention effect on the less serious outcome; the possible magnitude of the negative effect on the more serious outcome; and the magnitude of the diminution of the effect on the less serious when one increases the more serious that one is willing to tolerate. In any given situation, simple simulations might help trialists address the decision.

In our search for examples of the appropriate use of composite endpoints to address competing risks, we found few examples of clinical questions that met our criteria. We did find examples in areas of cerebrovascular surgery versus medical therapy for asymptomatic carotid stenosis and oxygen therapy for very low birth weight infants.

Benavente and colleagues published a meta-analysis of RCTs comparing carotid endarterectomy (CEA) versus medical therapy for asymptomatic carotid stenosis (21). The goal of CEA is to reduce stroke distal to the stenosis on the operated (ipsilateral) vessel. Five trials included 2440 patients with > 50% carotid stenosis who were randomized to CEA or medical therapy alone. Patients treated with CEA had fewer ipsilateral strokes (3.2% versus 6.2% in the medical group, Odd Ratio [OR] 0.46 with 95% CI 0.32 to 0.66). The rate of perioperative stroke or death was, however, significantly higher in the CEA group (2.4% versus 0.4% in the medical group, OR 4.51; 95% CI 2.36 to 8.34). In 2 of the trials included in the meta-analysis, several strokes occurred during angiography - these events were appropriately included in the analysis resulting in the 0.4% event rate in the medical group. Although the CEP of all stroke and perioperative stroke or death was lower for the CEA group (7.4% versus 9.2%, OR 0.68; 95% CI 0.51 to

0.9), the magnitude of benefit (OR 0.68 versus 0.46), relative to that of ipsilateral stroke, was diminished.

In this instance, the risk of a competing event (surgery causes deaths in patients who would, had they been treated medically, experience strokes) is real, and the number of associated deaths relative to strokes is sufficient to exaggerate the magnitude of the impact of the intervention on the less serious outcomes. Thus, the use of a composite including death, stroke, and transient ischemic attack is appropriate.

Another example involves the outcomes resulting from oxygen supplementation in critically ill neonates. In the 1940s and 1950s liberal use of oxygen led to 'an epidemic of retinopathy of prematurity (ROP)'(22). This led to a restriction of oxygen use, and tolerance of hypoxia in the premature infant that in turn increased mortality; the trade-off between maintaining adequate oxygenation versus preventing ROP in the premature infant remained unaddressed(23).

Several RCTs evaluating the optimal pulse-oximetry target in premature infants used a primary composite outcome of death and ROP. Polin and Bateman explain in an editorial (24), 'In all the studies, given the high expected rate of death among premature infants, death was included as an outcome because it competed with ROP as a risk, not because a difference in mortality was expected as a result of differences in oxygenation'. In this instance, competing risk of death is plausible because the mortality in premature infants is high enough that the effect of the intervention on ROP may be misleading if death were not considered in the analysis.

The results of the SUPPORT study (25), illustrate this concept. As shown in table 1, there was no difference in the primary composite outcome of death before discharge or severe retinopathy in the lower versus the higher oxygen saturation target; however the rate of occurrence of ROP was significantly higher in the higher target group. The competing risk of death before discharge was

plausible and occurred with high enough frequency that if not considered, would have distorted the true effect of the intervention on ROP (infants who die cannot develop ROP).

Table 1:

| Outcome | Low oxygen saturation target (N=654) n/N (%) | High oxygen saturation target (N=662) n/N (%) | Adjusted Relative Risk[1] (95% confidence interval) |
|---|---|---|---|
| Primary composite outcome (severe ROP or death before discharge | 171/605 (28.3) | 198/616 (32.1) | 0.90 (0.76 – 1.06) |
| Severe ROP | 41/475 (8.6) | 91/509 (17.9) | 0.52 (0.37 – 0.73) |
| Death before discharge | 130/654 (19.9) | 107/662 (16.2) | 1.27 (1.01 – 1.60) |

[1] Values were adjusted for stratification factors (study center and gestational-age group) as well as for familial clustering. BPD denotes bronchopulmonary dysplasia.

Table 1 - legend: Proportion of infants with the primary composite outcome and individual components of the composite. All surviving infants were evaluated by ophthalmologists for severe ROP. The discrepancy in the denominators is due to infants with 'undetermined ROP status' (49 in the low target group and 46 in the high target group).

In other instances, use of composite endpoint to deal with competing risk will be unnecessary and inadvisable. An example that has been previously noted in this paper is the DREAM trial(7). In this trial, based on available literature, it is implausible that rosiglitazone will increase death. In the unlikely event that rosiglitazone did increase death rates in those destined to develop diabetes, the number of deaths relative to the number of patients developing diabetes would be too few to result in a distortion of results: in the placebo group 658 patients developed diabetes and 33 patients died. Thus, the risk of a misleading inference through the spurious claim that rosiglitazone reduced death and the development of diabetes provides a compelling reason to avoid choice of a composite outcome in this trial.

Textbooks (9) and review articles (10), frequently illustrate the issue of competing risks using the example of death and hospitalization for worsening heart failure. Trials testing management strategies for HF frequently use a composite endpoint of death or hospitalization for worsening heart failure. A

review of landmark heart failure trials published in the NEJM revealed that 6 out of 26 RCTs used this composite endpoint. Some of the early heart failure trials used mortality as the primary endpoint and others have used a composite endpoint including outcomes other than these two. Table 2 summarizes event rates in trials that used the combined endpoint of death and hospitalization due to worsening heart failure. Although in each case the deaths are frequent enough to have distorted hospitalization results (the intervention might have substantially reduced hospitalizations by increasing deaths), in each case the direction of the effect was the same for the death and hospitalization outcomes (nor would it, prior to the trial, have been considered plausible that effects would move in opposite directions). Thus, these trials would not meet our criteria for use of composite endpoints for competing risks.

Table 2:

| | Death or hospitalization for CHF | | | | Death | | | | Hospitalization for HF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trial | Intervention (%) | Control | HR / OR (CI) | p | Intervention | Control | HR / OR (CI) | p | Intervention | Control | HR / OR (CI) | p |
| PARADIGM HF(26) 2014 | 914 (21.8) | 1117 (26.5) | 0.8 (0.73-0.87) | <0.001 | 558 (13.3) | 693 (16.5) | 0.8 (0.71-0.89) | <0.001 | 537 (12.8) | 658 (15.6) | 0.79 (0.71-0.89) | <0.001 |
| Echo-CRT (27) 2013 | 116 (28.7) | 102 (25.2) | 1.2 (0.92-1.57) | 0.15 | 45 (11.1) | 26 (6.4) | 1.81 (1.11-2.93) | 0.02 | 99 (24.5) | 90 (22.2) | 1.16 (0.87-1.55) | 0.25 |
| ASCEND-HF (28) 2011 | 321 (9.4) | 345 (10.1) | - 0.7 (-2.1-0.7) | 0.31 | 126 (3.6) | 141 (4.0) | -0.4 (-1.3-0.5) | 0.36 | 204 (6.0) | 208 (6.1) | -0.1 (-1.2-1.0) | 0.81 |
| EMPHASIS-HF(29) 2011 | 249 (18.3) | 356 (25.9) | 0.66 (0.56-0.78) | <0.001 | 147 (10.8) | 185 (13.5) | 0.77 (0.62-0.96) | 0.02 | 164 (12.0) | 253 (18.4) | 0.61 (0.50-0.75) | <0.001 |
| RAFT(30) 2010 | 297 (33.2) | 364 (40.3) | 0.75 (0.64-0.87) | <0.001 | 186 (20.8) | 236 (26.1) | 0.75 (0.62-0.91) | 0.003 | 174 (19.5) | 236 (26.1) | 0.68 (0.56-0.83) | <0.001 |
| SOLVD (31) 1992 | 434 (20.6) | 518 (24.5) | 0.8 (0.69-0.92) | <0.001 | 313 (14.8) | 334 (15.8) | 0.93 (0,79-1.1) | 0.3 | 184 (8.7) | 273 (12.9) | 0.65 (0.53-0.79) | <0.001 |

Table 2: Summary of 6 RCTs for heart failure management published in the NEJM which reported a composite endpoint of death and hospitalization for worsening HF. In the SOLVD trial, the primary outcome was mortality; the composite endpoints were included in the secondary endpoints.

**Discussion:**

ACCEPTED MANUSCRIPT

We found that many commentaries in review articles and textbook chapters note the use of composite endpoints to deal with the problem of competing risk, and most comment favorably on the use of composites in this regard. In spite of suggesting the use of composite endpoints to overcome the problem of competing risks, none of the commentaries provided clear guidance for when to use composite endpoints for competing risks. Noting the absence of guidance, we have suggested novel criteria for when to use, and not use, composite endpoints to address the competing risk issue.

Strengths of this study include the comprehensive review of the relevant methods literature including both journal articles and textbooks, the duplicate judgments of eligibility and abstraction of data, the abstraction of direct quotations from the relevant articles. With respect to our guidance, it was developed over many years and benefited from discussions with successive classes of graduate students.

The primary limitation of our study is that we did not recruit a wide range of methodologists in the development of our criteria, nor did we seek consensus among a group of such individuals. We could have, for example, recruited a cadre of experts who have written about use of composite endpoints to address competing risks and conducted a Delphi process to arrive at novel guidance. It may be, however, that the criteria we have suggested are sufficiently robust and intuitive that such an arduous process is not necessary.

Although randomized trials frequently use composite endpoints as primary endpoints, the reason for use of composite endpoints is rarely stated in the publications. As has been extensively noted in the literature and summarized in this article, use of composite endpoints may lead to difficulty with interpretation of results and misleading inferences about the results. Agencies such as the European network of Health Technology Assessment (EUnetHTA) state in their policy statement that the use of composite endpoints as primary endpoints is not recommended if a suitable single primary endpoint is

available (4). The guidance proposed in this paper should encourage appropriate consideration before using composite endpoints for competing risks.

**Conclusions:**

Numerous publications have highlighted the potential problem of competing risks in clinical trials, and noted the possibility of using a composite endpoints to address the problem. A number of authors have expressed enthusiasm for the composite endpoint solution to the competing risk problem. Composite endpoints however, have major limitations: they invariably present interpretation challenges, and are often subject to making effects seem more important than they really are. Thus, composite endpoints should be used judiciously and sparingly. An extensive literature search did not reveal specific guidelines for when to use composite endpoints for competing risks. We propose guidance on this topic based on available evidence. Composite endpoints should be used for competing risks only if the competing risk is plausible and if it occurs with high enough frequency to distort the effect of the intervention on the less serious endpoint. It is likely that these criteria will very seldom be met.

**Acknowledgement:** We thank Ms. Neera Bhatnagar for her assistance in designing the literature search strategy.

**References:**

1. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA. 2003;289(19):2554-9. PubMed PMID: 12759327.
2. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Permanyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. Bmj. 2007;334(7597):786. doi: 10.1136/bmj.39136.682083.AE. PubMed PMID: 17403713; PMCID: 1852019.
3. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. Bmj. 2010;341. doi: 10.1136/bmj.c3920.
4. EUnetHTA. Endpoints used for relative effectiveness assessment of pharmaceuticals - Composite endpoints. Methodological guidelines for rapid relative effectiveness assessment (REA) of Pharmaceuticals developed in WP5 of EUnetHTA JA2013. p. 1-22.

5. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. Ann Intern Med. 2008;149(9):612-7. PubMed PMID: 18981486.
6. Tomlinson G, Detsky AS. Composite end points in randomized trials: there is no free lunch. JAMA. 2010;303(3):267-8. doi: 10.1001/jama.2009.2017. PubMed PMID: 20085955.
7. Investigators DT, Gerstein HC, Yusuf S, Bosch J, Pogue J, Sheridan P, Dinccag N, Hanefeld M, Hoogwerf B, Laakso M, Mohan V, Shaw J, Zinman B, Holman RR. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. Lancet. 2006;368(9541):1096-105. doi: 10.1016/S0140-6736(06)69420-8. PubMed PMID: 16997664.
8. Montori VM, Isley WL, Guyatt GH. Waking up from the DREAM of preventing diabetes with drugs. BMJ. 2007;334(7599):882-4. doi: 10.1136/bmj.39169.447488.94. PubMed PMID: 17463460; PMCID: PMC1857789.
9. Chin R, Lee BY. Principles and Practice of Clinical Trial Medicine: Elsevier Science; 2008.
10. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. J Card Fail. 2005;11(8):567-75. doi: 10.1016/j.cardfail.2005.08.350. PubMed PMID: 16230258.
11. Sacks CA, Jarcho JA, Curfman GD. Paradigm shifts in heart-failure therapy--a timeline. N Engl J Med. 2014;371(11):989-91. doi: 10.1056/NEJMp1410241. PubMed PMID: 25184412.
12. Moyé LA. Multiple Analyses in Clinical Trials: Fundamentals for Investigators: Springer New York; 2006.
13. Young WR, Chen DG. Clinical Trial Biostatistics and Biopharmaceutical Applications: CRC Press; 2014.
14. Domanski MJ, McKinlay S. Successful Randomized Trials: A Handbook for the 21st Century: Lippincott Williams & Wilkins; 2009.
15. Guyatt G. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3E: McGraw-Hill Education; 2014.
16. Kessler KM. Combining composite endpoints: counterintuitive or a mathematical impossibility? Circulation. 2003;107(9):e70; author reply e. PubMed PMID: 12628965.
17. Song R, Cook TD, Kosorok MR. What we want versus what we can get: a closer look at failure time endpoints for cardiovascular studies. J Biopharm Stat. 2008;18(2):370-81. doi: 10.1080/10543400701697224. PubMed PMID: 18327727.
18. Hallstrom AP, Litwin PE, Weaver WD. A method of assigning scores to the components of a composite outcome: an example from the MITI trial. Control Clin Trials. 1992;13(2):148-55. PubMed PMID: 1316829.
19. Braunwald E, Cannon CP, McCabe CH. Use of composite endpoints in thrombolysis trials of acute myocardial infarction. Am J Cardiol. 1993;72(19):3G-12G. PubMed PMID: 8279357.
20. Lubsen J, Kirwan BA. Combined endpoints: can we use them? Stat Med. 2002;21(19):2959-70. doi: 10.1002/sim.1300. PubMed PMID: 12325112.
21. Benavente O, Moher D, Pham B. Carotid endarterectomy for asymptomatic carotid stenosis: a meta-analysis. BMJ. 1998;317(7171):1477-80. PubMed PMID: 9831572; PMCID: PMC28726.
22. Robertson AF. Reflections on errors in neonatology: I. The "Hands-Off" years, 1920 to 1950. J Perinatol. 2003;23(1):48-55. doi: 10.1038/sj.jp.7210842. PubMed PMID: 12556927.
23. Bolton DP, Cross KW. Further observations on cost of preventing retrolental fibroplasia. Lancet. 1974;1(7855):445-8. PubMed PMID: 4131442.
24. Polin RA, Bateman D. Oxygen-saturation targets in preterm infants. N Engl J Med. 2013;368(22):2141-2. doi: 10.1056/NEJMe1305534. PubMed PMID: 23642082.
25. Network SSGotEKSNNR, Carlo WA, Finer NN, Walsh MC, Rich W, Gantz MG, Laptook AR, Yoder BA, Faix RG, Das A, Poole WK, Schibler K, Newman NS, Ambalavanan N, Frantz ID, 3rd, Piazza AJ, Sanchez

PJ, Morris BH, Laroia N, Phelps DL, Poindexter BB, Cotten CM, Van Meurs KP, Duara S, Narendran V, Sood BG, O'Shea TM, Bell EF, Ehrenkranz RA, Watterberg KL, Higgins RD. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med. 2010;362(21):1959-69. doi: 10.1056/NEJMoa0911781. PubMed PMID: 20472937; PMCID: PMC2891970.

26. McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, Zile MR, Investigators P-H, Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014;371(11):993-1004. doi: 10.1056/NEJMoa1409077. PubMed PMID: 25176015.

27. Ruschitzka F, Abraham WT, Singh JP, Bax JJ, Borer JS, Brugada J, Dickstein K, Ford I, Gorcsan J, 3rd, Gras D, Krum H, Sogaard P, Holzmeister J, Echo CRTSG. Cardiac-resynchronization therapy in heart failure with a narrow QRS complex. N Engl J Med. 2013;369(15):1395-405. doi: 10.1056/NEJMoa1306687. PubMed PMID: 23998714.

28. O'Connor CM, Starling RC, Hernandez AF, Armstrong PW, Dickstein K, Hasselblad V, Heizer GM, Komajda M, Massie BM, McMurray JJ, Nieminen MS, Reist CJ, Rouleau JL, Swedberg K, Adams KF, Jr., Anker SD, Atar D, Battler A, Botero R, Bohidar NR, Butler J, Clausell N, Corbalan R, Costanzo MR, Dahlstrom U, Deckelbaum LI, Diaz R, Dunlap ME, Ezekowitz JA, Feldman D, Felker GM, Fonarow GC, Gennevois D, Gottlieb SS, Hill JA, Hollander JE, Howlett JG, Hudson MP, Kociol RD, Krum H, Laucevicius A, Levy WC, Mendez GF, Metra M, Mittal S, Oh BH, Pereira NL, Ponikowski P, Tang WH, Tanomsup S, Teerlink JR, Triposkiadis F, Troughton RW, Voors AA, Whellan DJ, Zannad F, Califf RM. Effect of nesiritide in patients with acute decompensated heart failure. N Engl J Med. 2011;365(1):32-43. doi: 10.1056/NEJMoa1100171. PubMed PMID: 21732835.

29. Zannad F, McMurray JJ, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B, Group E-HS. Eplerenone in patients with systolic heart failure and mild symptoms. N Engl J Med. 2011;364(1):11-21. doi: 10.1056/NEJMoa1009492. PubMed PMID: 21073363.

30. Tang AS, Wells GA, Talajic M, Arnold MO, Sheldon R, Connolly S, Hohnloser SH, Nichol G, Birnie DH, Sapp JL, Yee R, Healey JS, Rouleau JL, Resynchronization-Defibrillation for Ambulatory Heart Failure Trial I. Cardiac-resynchronization therapy for mild-to-moderate heart failure. N Engl J Med. 2010;363(25):2385-95. doi: 10.1056/NEJMoa1009540. PubMed PMID: 21073365.

31. Effect of Enalapril on Mortality and the Development of Heart Failure in Asymptomatic Patients with Reduced Left Ventricular Ejection Fractions. New England Journal of Medicine. 1992;327(10):685-91. doi: doi:10.1056/NEJM199209033271003. PubMed PMID: 1463530.

32. Effect of Enalapril on Survival in Patients with Reduced Left Ventricular Ejection Fractions and Congestive Heart Failure. New England Journal of Medicine. 1991;325(5):293-302. doi: doi:10.1056/NEJM199108013250501. PubMed PMID: 2057034.

ACCEPTED MANUSCRIPT