

Composite outcomes – the devil is in the details

Luis Eduardo Santos Fontes¹, Matheus Guedes Monteiro², Rafael Leite Pacheco³, Rachel Riera⁴

¹Corresponding author. Medicine Faculty of Petrópolis. Petrópolis, Rio de Janeiro, Brazil. ORCID: 0000-0002-0928-7031. luis.fontes@ceo2.com.br

²Medicine Faculty of Petrópolis. Petrópolis, Rio de Janeiro, Brazil. ORCID: 0000-0001-9026-9588. matheus.guedesmonteiro@gmail.com

³Federal University of São Paulo. São Paulo, São Paulo, Brazil. ORCID: 0000-0001-7487-8471. rleitpacheco@hotmail.com

⁴Federal University of São Paulo. São Paulo, São Paulo, Brazil. ORCID: 0000-0002-9522-1871. rachelriera@hotmail.com

ABSTRACT | Randomized controlled trials (RCT's) are the gold standard study design for assessing the efficacy and safety of health interventions, intending to support clinicians to deliver evidence-based healthcare. In the past decades, there was a rise in published RCT's using composite outcomes. A composite outcome is a set of two or more component endpoints gathered into a single primary or secondary outcome. This strategy is used to enhance the rate of events in clinical trials leading to smaller, cheaper, and faster studies. In this narrative review of the literature, we aimed to revise the concepts behind composite outcomes and discuss their utility and concerns regarding clinical decision-making. We presented definitions, types, advantages and disadvantages of composite outcomes, as well as concerns about the interpretation of trials' results when they are used. Considering the importance of this topic, clinicians must be aware of some aspects regarding composite outcomes in order to interpret them correctly and make appropriate clinical decisions.

KEYWORDS: Composite outcomes. Randomized controlled trials. Clinical decision making.

Since the first rudimentary experiment conducted by James Lind in 1747¹ testing which interventions were effective against scurvy, clinical trials have become the gold standard for assessing the efficacy and safety of medical interventions and allowing clinicians to choose the best options for patient care. Well-designed randomized clinical trials (RCTs) have been used for underpinning the indication of treatments in clinical guidelines and providing evidence to support them. Although the initial clinical trials were simple and straightforward, they have grown exponentially in number and complexity². One of the critical steps of planning an RCT is the definition of clinical question by the use of PICO (patients, interventions, comparator/control, and outcomes) that will be assessed. A key set of outcomes must include clinically relevant outcomes (also called endpoints) and must answer appropriately to the clinical question that motivated the study. Poor decisions when choosing endpoints could lead to meaningless measurements and evaluations for decision-making, contributing to research waste. Chalmers et al. have argued that most published medical research targeted low priority clinical questions and failed to assess important clinical outcomes³.

There are some trials designed to answer a clinical question in which a single outcome is evaluated. For example, GISSI-I study tested the efficacy of intravenous streptokinase in patients with acute myocardial infarction for reducing the risk of death⁴. In early randomized controlled trials, most of the researchers assessed the efficacy of an intervention comparing death rates between groups

as the primary endpoint. However, improvement in the clinical management of some diseases have modified their natural history, and the use of death rates became less frequent. Whenever the outcome of interest is less frequent or when researchers want to detect smaller benefits of a given intervention, clinical trials become more challenging. These circumstances require large sample sizes and more time to follow-up the participants, increasing costs and duration of trials and consequently, threatening its feasibility. One alternative to minimize these issues is choosing two or more endpoints, gathering them into a single one, formally known as COMPOSITE OUTCOME (CO).

Composite outcomes are a combination of two or more distinct endpoints so-called “component endpoints.” Authors have been used this strategy to achieve statistical power by improving overall event rates. With this approach, clinical trials might be conducted in a cheaper and faster way, with smaller sample requirements⁵. Composite outcomes might diminish resource difficulties either related to patients or related to funding⁶. Other advantage includes situations when researchers want to explore outcomes that are of equal value, avoiding the necessity of arbitrary choices for one outcome against other⁷.

There are three sorts of CO: “index/score” composites, “event rate” composites and “time to event” composites⁸. Table 1 summarizes the main characteristics of these three types of CO.

Table 1. Types and characteristics of composite outcomes

Type of CO	Characteristics	Reference example
Index/score	Uses a rating scale to achieve a score or index. Usefulness depends on the validity and reliability of the rating scale.	Behnke et. al.(9)
Event rate	Look for an event rate after a certain period of treatment or follow-up. “Event” is defined as the occurrence of any one event from a set of endpoints. Usefulness depends on the clinical importance and meaning of each component.	Subtil et. al. (10)
Time to event	Look for the time for the first event . “Event” is again defined as the occurrence of any one event from a set of endpoints. Usefulness depends on the clinical importance and meaning of each component.	Kim et.al. (11)

Medical research using CO was more common in the last decades. Heart failure trials are good examples of the shifting from a single primary outcome into a mortality-morbidity CO¹². Fremantle et al. looked up through nine medical journals from 1997 to 2001 and found 167 trials using CO in different specialties. Cardiovascular trials were the most common, corresponding for 38% of the studies. Other specialties that have been publishing utilizing this strategy includes Pediatrics, Hematology, Oncology, and others⁷. One study examined how CO have been used, constructed and reported in cardiovascular trials and evaluated how each component endpoint has contributed to the composite estimate in those trials¹³. Authors found 304 trials in 14 leading general medical, cardiology, and cardiothoracic surgery journals between 2000 and 2007. In those trials, 73% of the articles reported a primary CO and 27% reported a secondary outcome using a composite. Overall, there was an average of three component endpoints in the composites. Death was the most common individual endpoint reported but had minimal contribution to the composite estimates in most of the studies retrieved.

Interpreting the results of trials using composite outcomes is challenging. The Devil is in the details discussed below, and the readers must be aware. Clinicians have to keep in mind what are the component endpoints being tested and make your own judgments on the relevance of each one for their clinical question so that they could make a reasonable clinical decision.

McCoy and colleagues⁵ argue that clinicians must examine three core topics when considering the use of CO to support clinical decision making, based on foundations of the User's Guide to the Medical Literature¹⁴:

1. Clinical importance of the component endpoints for patients and clinicians.
2. The frequency of each component endpoint.
3. Variations in relative risk reductions between component endpoints.

When component endpoints are similar regarding relevance and frequency, with low variations in the relative risks, clinicians could rely on the composite result and make a more appropriate clinical decision. However, in circumstances in which these aspects

(importance, frequency, variability of estimates) differ substantially throughout the components, decision making is more delicate.

A reasonable CO must have all component endpoints as important and useful for patients and clinicians. When all components are of great clinical importance, we can rely on that one component alone could support clinical decision making, even if the others do not show an effect⁸. However, large variability in importance between component endpoints raises difficulties for decision-making. In a trial with a composite of three components, one given component of small clinical importance might be settled with two relevant components. If this less important component shows an effect and the most important components do not, we should be skeptical in using the composite for clinical decision making. For example, one study has tested systemic glucocorticoids versus placebo for chronic obstructive pulmonary disease, assuming a composite of death/need for intubation/hospital readmission for exacerbations/intensification of drug therapy as a CO¹⁵. Failure of the intervention was higher in the placebo group (33%) than in the intervention group (23%). Authors concluded that "systemic glucocorticoids resulted in improvement in clinical outcomes." However, combining important outcomes such as death, with relatively trivial components like hospital admission can lead to misinterpretation. The higher rate of treatment failure in the placebo group could lead one to believe that the rate of all components (such as death) were also higher among participants who received placebo. But it was not the case, death rates were similar between groups, but the composite was influenced by intensification of drug therapy.

The frequency of each component outcome is crucial for a reasonable decision making. If the most critical component is less frequent than a component of minor value, the CO is less informative. Ideally, the gradient frequency amongst the components should be minimal. A good example to clarify this concept is the SENIORS trial. This randomized controlled trial was conducted to determine the effect of nebivolol on mortality and cardiovascular hospital admission in elderly patients with heart failure¹⁶. The primary outcome was a composite of all-cause mortality or cardiovascular hospital admission. Table 2 shows the frequency of events for each component.

Table 2. Primary outcomes in SENIORS trial [16]

	Nebivolol (n=1067)	Placebo (n=1061)
Primary CO	332 (31.1%)	375 (35.3%)
All-cause Mortality contributing for CO	76 (7,1%)	99 (9,3%)
CV hospital admissions contributing for CO	256 (24%)	276 (26%)

The primary outcome occurred in 332 patients (31.1%) receiving nebivolol compared with 375 (35.3%) receiving placebo [hazard ratio (HR) 0.86, 95% CI 0.74–0.99; $P < 0.039$]. Authors concluded that “Nebivolol, a beta-blocker with vasodilating properties, is an effective and well-tolerated treatment for heart failure in the elderly.” However, the frequency of hospital admission was far more frequent than death events (24% vs. 7,1% and 26% vs. 9,3% in intervention and control group, respectively). This gradient raises concerns about utility for clinical decision making. Differently, a study that investigated effects of ramipril vs placebo in cardiovascular deaths (HOPE trial), myocardial infarction and stroke, showed smaller gradients between the composite components (8,1 %, 12,3% and 4,9% in control group , and 6,1% , 9,9% and 3,4% in ramipril group)¹⁷. This small gradient of frequency between components gives support to the use of the composite.

Finally, clinicians must scrutinize the variability in the estimates of each component endpoint. Designing components with an expected biological similar response is advice for researchers conducting clinical trials, and examining this in a paper is an important step judging the usefulness of a composite outcome. When there is a strong biological rationale for response to an intervention across all components, variations in point estimates are not expected. Despite rationale, only similar estimates (and confidence intervals) could lead clinicians to rely on CO and apply them to patient care. In HOPE trial cited above, component endpoints reacted similarly. In a recent article published in the Journal of the American Medical Association - JAMA¹¹, researchers tested the long-term effects of escitalopram in a CO of all-cause mortality, myocardial infarction, and percutaneous coronary intervention. Among 300 randomized patients the CO occurred in 61 patients (40,9%) in escitalopram group and in 81

(53,6%) in placebo group (HR 0,69; 95% CI, 0,49-0,96; $P = 0,03$). However incidences for all-cause mortality were 20,8% vs 24,5% (HR 0,82; 95% CI, 0,51-1,33; $P = 0,43$), for cardiac death 10.7% vs 13.2% (HR 0,79; 95% CI 0,41-1,52; $P = 0,48$); for myocardial infarction 8,7% vs 15,2% (HR 0,54; 95% CI 0,27-0,96; $P = 0,04$) and for percutaneous coronary intervention 12,8% vs 19,9% (HR 0,58; 95% CI 0,33-1,04; $P = 0,07$). Although point estimates appear to show benefit in all components, only myocardial infarction showed a statistically significant result favoring the intervention, and the estimative for other outcomes, such as cardiac death and all-cause mortality were very imprecise, since their confidence interval include a range of values that might represent either an important benefit or an important harm. The author concluded that “among patients with depression following recent acute coronary syndrome, 24-week treatment with escitalopram compared with placebo resulted in a lower risk of major adverse cardiac events after a median of 8.1 years”. Poor reporting or misinterpretation of the composite is an argument highlighted by critics. A real possibility to an average clinician is presuming that the described composite benefit relates to all component endpoints. Overall, the composite might show benefit, but it may be influenced by that one positive endpoint, falsely seeming that there is evidence of a benefit on the others. Clinicians mustn't rely on the composite in this situation and should analyze each component individually

Conclusion

Composite outcomes have been increasingly used in clinical trials to enhance statistical power, achieving results with smaller sample sizes and assessing different mortality-morbidity outcomes. It is a

clever and plausible strategy when the component endpoints were selected in a meaningful way, allowing clinicians, patients, and policymakers to make a well-informed decision. Reporting standards are necessary to ensure a proper takeaway message from articles using the composite. The validity of a CO is related to a set of small variations in the frequency of each component, similar estimates and confidence intervals, and similar importance to patients and clinicians.

So, never leave composite outcomes away from your critical eye.

Author contributions

Fontes LES was responsible for conception and design of the study, analysis and interpretation of the data, drafting the paper and final approval of the version to be published. Riera R was responsible for conception and design of the study, revising it critically for important intellectual content and final approval of the version to be published. Monteiro MG was responsible for acquisition, analysis, and interpretation of data for the work, drafting the paper, revising it critically for important intellectual content; and final approval of the version to be published. Pacheco RL was responsible for substantial contributions to the study design; acquisition, analysis, and interpretation of data, drafting the response to peer review, revising the work critically for important intellectual content and final approval of the version to be published.

Competing interests

No financial, legal or political competing interests with third parties (government, commercial, private foundation, etc.) were disclosed for any aspect of the submitted work (including but not limited to grants, data monitoring board, study design, manuscript preparation, statistical analysis, etc.).

References

1. Sutton G. James Lind aboard Salisbury. [Internet]. JLL Bulletin: Commentaries on the history of treatment evaluation. 2004 [cited 2018 Sep 17]. Available from: <http://www.jameslindlibrary.org/articles/james-lind-aboard-salisbury/>
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Med*. 2010;7(9). doi: [10.1371/journal.pmed.1000326](https://doi.org/10.1371/journal.pmed.1000326)
3. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374:86-9. doi: [10.1016/S0140-6736\(09\)60329-9](https://doi.org/10.1016/S0140-6736(09)60329-9)

4. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet*. 1986;1(8478):397-402.

5. McCoy CE. Understanding the Use of Composite Endpoints in Clinical Trials. *West J Emerg Med*. 2018;19(4):631-4. doi: [10.5811/westjem.2018.4.38383](https://doi.org/10.5811/westjem.2018.4.38383)

6. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol*. 2007;196(2):1-6. doi: [10.1016/j.ajog.2006.10.903](https://doi.org/10.1016/j.ajog.2006.10.903)

7. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite Outcomes in Randomized Trials: Greater Precision but with Greater Uncertainty? *J Am Med Assoc*. 2003;289(19):2554-9. doi: [10.1001/jama.289.19.2554](https://doi.org/10.1001/jama.289.19.2554)

8. Chi GYH. Some issues with composite endpoints in clinical trials. *Fundam Clin Pharmacol*. 2005;19(6):609-19. doi: [10.1111/j.1472-8206.2005.00370.x](https://doi.org/10.1111/j.1472-8206.2005.00370.x)

9. Behnke K, Sogaard J, Martin S, Bauml J, Ravindran AV, Agren H et al. Mirtazapine orally disintegrating tablet versus sertraline: a prospective onset of action study. *J Clin Psychopharmacol*. 2003;23(5):358-64. doi: [10.1097/01.jcp.0000085408.08426.05](https://doi.org/10.1097/01.jcp.0000085408.08426.05)

10. Subtil D, Brabant G, Tilloy E, Devos P, Canis F, Fruchart A et al. Early clindamycin for bacterial vaginosis in pregnancy (PREMEVA): a multicentre, double-blind, randomised controlled trial. *Lancet*. 2018;392:2171-79. doi: [10.1016/S0140-6736\(18\)31617-9](https://doi.org/10.1016/S0140-6736(18)31617-9)

11. Kim JM, Stewart R, Lee YS, Lee HJ, Kim MC, Kim JW et al. Effect of escitalopram vs placebo treatment for depression on long-term cardiac outcomes in patients with acute coronary syndrome: A randomized clinical trial. *JAMA*. 2018;320(4):350-7. doi: [10.1001/jama.2018.9422](https://doi.org/10.1001/jama.2018.9422)

12. Shen L, Jhund PS, Mogensen UM, Køber L, Claggett B, Rogers JK et al. Re-Examination of the BEST Trial Using Composite Outcomes, Including Emergency Department Visits. *JACC Hear Fail*. 2017;5(8):591-9. doi: [10.1016/j.jchf.2017.04.005](https://doi.org/10.1016/j.jchf.2017.04.005)

13. 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.

14. Guyatt G. Users' guides to the medical literature: essentials of evidence-based clinical practice. 2nd ed. New York: McGraw-Hill Medical; 2008.

15. Niewoehner DE, Erbland ML, Deupree RH, Collins D, Gross NJ, Light RW et al. Effect of systemic glucocorticoids on exacerbations of chronic obstructive pulmonary disease. Department of Veterans Affairs Cooperative Study Group. *N Engl J Med*. 1999;340(25):1941-7. doi: [10.1056/NEJM199906243402502](https://doi.org/10.1056/NEJM199906243402502)

16. Flather MD, Shibata MC, Coats AJS, Van Veldhuisen DJ, Parkhomenko A, Borbola J et al. Randomized trial to determine the effect of nebivolol on mortality and cardiovascular hospital admission in elderly patients with heart failure (SENIORS). *Eur Heart J*. 2005;26(3):215-25. doi: [10.1093/eurheartj/ehi115](https://doi.org/10.1093/eurheartj/ehi115)

17. Sleight P. The HOPE Study (Heart Outcomes Prevention Evaluation). *JRAAS*. 2000;1:18-20. doi: [10.3317/jraas.2000.002](https://doi.org/10.3317/jraas.2000.002)