Statistical properties of Continuous Composite Outcomes: Implications for clinical trial design

Jesse D. Troy a, b, *, Ryan A. Simmons b

a The Marcus Center for Cellular Cures, Duke University School of Medicine, USA
b Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 424 Erwin Road Suite, 1102 Hock Plaza Box 2721, Durham, NC, 27710, USA

A R T I C L E  I N F O

Keywords:
Clinical trials
Treatment outcome
Statistics

A B S T R A C T

Statistical efficiency can be gained in clinical trials by using composites of time-to-event outcomes when the individual component outcomes have low event rates. However, the utility of continuous composite outcome measures is not as clear. Efficiency can be either gained or lost by using a continuous composite outcome measure depending on several factors, including the strength of correlation between the component outcomes and the size of the treatment effect on each component. In this article we review these concepts from the standpoint of planning a new trial. Statistical properties of composites formed from normally distributed continuous outcomes are discussed. An example dataset is used to demonstrate concepts and complete mathematical details are provided. Finally, a conceptual model for clinical trial design with continuous composites is proposed that could be used as a guide to evaluate the utility of a continuous composite outcome in a future trial based on existing knowledge in the therapeutic area.

1. Introduction

Clinical trials typically specify multiple outcome measures in a hierarchical fashion, often designating a single primary outcome and several secondary outcome measures. The statistical design of the trial is usually justified in terms of power for detecting clinically relevant differences on the primary outcome. The statistical plan may also address power and correction for Type I error in the analysis of secondary outcomes. In applications where multiple co-primary outcomes are of clinical interest it is sometimes accepted to combine these into a single, composite outcome [1]. Doing so may afford gains in statistical efficiency, especially in the case of rarely occurring time-to-event outcomes, while avoiding the inflation of Type I error associated with testing each outcome separately [2]. For example, this practice has become common in cardiovascular trials that use the Major Adverse Cardiac Event (MACE) outcome, which is a composite of death from any cause, myocardial infarction, coronary revascularization, stroke or cardiac-related hospitalization [3]. In this setting, intervention is likely to impact multiple components of the composite. Furthermore, the event rate during the trial follow-up period is expected to be low for each of the MACE components. Given that statistical power in survival analysis is related to the event rate [4], the combination of these component events into a single time-to-event composite makes trial conduct feasible.

However, efficiency gains are not realized from all types of composite outcomes. For example, while a binary composite outcome might facilitate simultaneous study of several clinically relevant endpoints, this benefit could easily be outweighed if the composite event rate is near 50% in the control arm at which point trial power would be negatively impacted [5]. Therefore, it is necessary to carefully consider the utility of the proposed composite outcome when developing the statistical plan for the trial. While time-to-event and binary composites have become familiar to applied statisticians, we find little has been written regarding the statistical properties of continuous composites. Such continuous composite measures have been used in clinical trials in several therapeutic areas including, for example, rheumatoid arthritis [6] and multiple sclerosis [7]. Previous authors have established the theoretical properties of continuous, normally distributed composites and have evaluated the consistency of empirical experience with the theoretical properties [8]. However, this work has not been incorporated into a cohesive framework for planning de novo trials.

This article expands on previous work by considering how to use knowledge regarding the statistical properties of continuous compos-
2. Motivating example and definition of the continuous composite outcome

The potential use of a continuous composite outcome is motivated by the following hypothetical example, the setting for which is the early part (Phase IIA or IIB) of the product development life cycle. Specifically, we suppose that an exploratory, Phase II controlled trial will be conducted to evaluate the effect of an intervention on two disease-related measures of health status reflecting different biological pathways through which the investigational therapy is believed to work. We will assume the activity of the investigational therapy on these two biological pathways is measurable using validated instruments that yield normally distributed outcome measures. For purposes of this example, a biological pathway could be considered anything as granular as a cell signaling pathway or a higher-level concept such as the progression of factors leading to poor control of blood glucose levels in diabetic patients.

Prior to conducting this hypothetical trial, it is uncertain on which outcome the treatment will have the largest effect. Concrete examples of such a scenario might include things like treatments that could impact both gross and fine motor control in patients with Parkinson’s disease, or blood cholesterol and triglyceride levels in patients with a familial hypercholesterolemia. In this hypothetical example, given the pre-trial uncertainty regarding potential treatment effect size on each outcome, the investigators propose creating a composite outcome that reflects the joint effect of the treatment on both biological pathways. The investigators reason that this would ensure their trial does not miss a potential efficacy signal. To evaluate the utility of this proposed approach we formalize the problem as follows.

Let $i = (1, 2)$ index two treatment groups in a randomized controlled trial with $i = 1$ indicating the treatment and $i = 2$ indicating the control. Then, let $X_i \sim N(\mu_{X_i}, \sigma_{X_i}^2)$ be a measure of health status reflecting activity of a biological pathway in patients assigned to group $i$, and let $Y_i \sim N(\mu_{Y_i}, \sigma_{Y_i}^2)$ be another such measure reflecting biological activity in a different, but possibly overlapping, biological pathway. Furthermore, suppose outcomes $X$ and $Y$ have correlation

$\rho_1 = \text{corr}(X_i, Y_i)$. Finally, we define a simple composite as the sum of the two outcomes: $Z_i = X_i + Y_i$. It should be noted that while other composites are possible, e.g., a mean or some weighted combination of the two components, that the statistical properties described in the following exposition would still hold. The composite based on a sum is presented for simplicity.

The above definition allows each component, $X$ and $Y$, to have its own variance, and for that variance to differ by treatment group assignment. This is often the case, for example, because each outcome measure uses a different scale and has different statistical properties. Therefore, to facilitate interpretation we can standardize the two variables for combination into a composite by dividing the observations by their standard deviation, thus achieving a common variance as follows:

$\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma_{Y_1}^2 = \sigma_{Y_2}^2 = 1$. Finally, to further simplify the example we will assume that as a consequence of randomization there is a consistent association between the two measures of health status $X$ and $Y$ regardless of treatment group. Thus, we assume a common correlation: $\text{corr}(X_1, Y_1) = \text{corr}(X_2, Y_2) = \rho$.

3. Treatment effect size for the continuous composite outcome

Using the definitions above we can now examine the standardized effect size for the composite outcome measure. The standardized effect size is the ratio of the mean difference to the standard deviation (Cohen’s $d$). Note, that there are 3 treatment effect sizes, one for each of the component outcome measures $X$ and $Y$, and one for the composite $Z$. Given we have assumed a common standard deviation of 1, this simplifies the effect size formulation for $X$ and $Y$ as follows.

$$ES(X) = \frac{\mu_{X_1} - \mu_{X_2}}{\sigma} = \frac{(\mu_{X_1} - \mu_{X_2})}{\sigma}$$

$$ES(Y) = \frac{\mu_{Y_1} - \mu_{Y_2}}{\sigma} = \frac{(\mu_{Y_1} - \mu_{Y_2})}{\sigma}$$

Finally, in this example we will assume the treatment effects are both positive (the treatment does no harm) and that the treatment may possibly exert a smaller effect on $Y$ than it does on $X$.

$$ES(Y) \geq 0 \text{ and } ES(X) \geq 0$$

$$ES(Y) \leq ES(X)$$

Using this information, it can be shown that the treatment effect size on the composite outcome $Z$ is defined as follows [8]. The derivation is given in the Supplementary Materials.

$$ES(Z) = ES(X + Y) = \frac{ES(X) + ES(Y)}{\sqrt{1 + \rho}}$$

The formula shows that the treatment effect size for the composite is a function of the component effect sizes and the correlation between the two component outcome measures. Notably, the composite effect size is largest when there is no correlation between the components $X$ and $Y$ ($\rho = 0$) with stronger correlation having a diminishing effect on the composite effect size. The formula also shows that even when $\rho = 0$ the composite effect size is never as large as the sum of the component effect sizes. These relationships are investigated using an example dataset in the next section.

4. Application using an example dataset

Fig. 1 shows the relationship between two correlated outcomes, $X$ and $Y$, in an example dataset of $N = 120$ observations (60 per group).

![Fig. 1. Example dataset: Correlation between outcome measures X and Y.](image-url)
with \( \text{Corr}(X,Y) = 0.8 \). Fig. 2 shows the treatment effect sizes for \( X \), \( Y \) and \( Z \) in this example dataset. The treatment has a moderate sized effect on \( X \), with \( ES(X) = 0.5 \), and a small sized effect on \( Y \), with \( ES(Y) = 0.3 \). The treatment effect for the composite is \( ES(Z) = 0.41 \) and can be calculated from the formula given above. In this example, the composite treatment effect size is larger than the smallest component effect size. However, it is not as large as the largest component effect size. Therefore, in this example, with 60 patients per group this study has greater power (61%) to detect the treatment effect on the composite \( Z \) than it does to detect the treatment effect on \( Y \) alone (37% power). However, the study does not have as much power to detect the composite treatment effect as it does to detect the treatment effect on \( X \) alone (78% power). As we will see in the next section, this scenario is not universal. In fact, the magnitude of \( ES(Z) \) depends greatly on several factors, some of which may be difficult to know at trial design time.

5. Implications for trial design

Fig. 3 diagrams the relationship between \( ES(Z) \) and the correlation between \( X \) and \( Y \) for various values of the ratio of \( ES(Y)/ES(X) \). To understand this ratio, simply imagine that \( ES(X) = 1 \) and then the ratio describes the size of \( ES(Y) \). There are several important things to note from this graph. First, the vertical axis range is limited numerically but covers a wide range of qualitative values for the composite effect size, including moderate \( (ES(Z) = 0.5) \) to very large \( (ES(Z) \geq 1) \) effect sizes. Second, stronger correlation has the effect of diminishing \( ES(Z) \) regardless of how large the treatment effect is on \( X \) or \( Y \). Third, and perhaps most importantly, when the treatment effects on the components \( X \) and \( Y \) are highly divergent, e.g., \( ES(Y)/ES(X) \) equal to 0 or 0.25, then \( ES(Z) \) is smaller than the largest component effect size, \( ES(X) \). Therefore, there are a range of scenarios under which using a composite outcome might be viewed as beneficial, as it relates to amplifying the treatment effect size relative to using a single outcome alone, but there are also many cases in which selection of a single outcome measure would be more beneficial than a composite in terms of

![Figure 2](https://example.com/figure2.png)

Fig. 2. Treatment effects on the component outcomes (\( X \) and \( Y \)) and the composite outcome (\( Z \)). Data were generated from bivariate normal distributions as follows. Group 1: \( N = 60 \), \( X \sim N(0.5,1) \), \( Y \sim N(0.3,1) \), \( \text{Corr}(X,Y) = 0.8 \); Group 2: \( N = 60 \), \( X \sim N(0,1) \), \( Y \sim N(0,1) \), \( \text{Corr}(X,Y) = 0.8 \).

![Figure 3](https://example.com/figure3.png)

Fig. 3. Relationship between composite effect size, component effect sizes and correlation.
treatment effect size. These implications can be understood concretely in terms of relative sample sizes required for a two-sample t-test under a range of assumptions about $ES(Y)/ES(X)$ and $\rho_{XY}$ as follows.

Table 1 shows how efficiency might be gained (or lost) by using a composite outcome measures by comparing the sample size required to detect the composite effect size to the sample size required to detect the largest of the component treatment effect sizes. For example, when the components are not correlated, $\rho_{XY} = 0$, and the treatment has the same sized effect on each component, $ES(Y)/ES(X) = 1$, then the sample size required to detect the composite effect size is nearly half (0.53) of what would be required to detect the largest treatment effect on a single component outcome measure. However, this efficiency gain erodes rapidly as the correlation increases. For example, at $\rho_{XY} = 1$ the same sample size is required to maintain 80% power even if the treatment effects on the components are the same. Note also that even when $\rho_{XY} = 0$, the statistical advantage of a composite outcome disappears when the $ES(Y)$ is less than half of $ES(X)$. In these cases the use of a composite outcome measure actually requires larger sample sizes to maintain 80% power relative to using only the component outcome that is more greatly affected by the treatment. The most extreme case of this appears when the treatment impacts only X and not Y (last row of Table 1) with the relative increase in sample size being approximately 2–4 times what would be required to detect the effect on X alone.

There are several take-home messages in these observations for the applied biostatistician. First, if the component outcome measures are perfectly correlated then the treatment must have the same sized effect on each component in order for the composite to perform as well as using a single outcome measure. Second, if the component outcomes are weakly correlated then some efficiency will be gained by using a composite outcome measure even when the treatment effect sizes on the composite are different. However, the composite outcome exhibits substantial efficiency loss as the ratio $ES(Y)/ES(X)$ approaches 0 (i.e. as the treatment effect becomes more and more concentrated in a single component). Finally, and most importantly, the use of a continuous composite outcome measure is associated with substantial efficiency loss when the treatment impacts only a single component.

6. A conceptual model for evaluating the utility of a continuous composite outcome

Given the aforementioned complexity inherent to continuous composite outcome measures, we offer the following conceptual model that the applied biostatistician might use to determine the utility of a continuous composite outcome measure at trial design time. Briefly, Fig. 4 shows a hypothetical treatment that is believed to modulate two separate biological pathways. The treatment acts on Component X of the disease state through Pathway 1, and exerts an effect on Component Y of the disease state through Pathway 2. There are normally distributed outcome measures for both X and Y, and there is some correlation assumed between these outcomes. Consider, for example, X and Y to be measures of ambulation and cognition in a trial testing a treatment for multiple sclerosis.

As shown in Fig. 4, the treatment has semi-overlapping temporal effects on the two components of the disease state, X and Y. This is illustrated by the fact that the treatment activates Pathway 1 early, from time t0 to t2 whereas the treatment activates Pathway 2 at a later point, starting at time t1 and extending through t3. Using the example from multiple sclerosis, this would imply the treatment has near-term impacts on ambulation with effects on cognition emerging later. Applying the information gained from the earlier discussion, it is known that if there is a strong correlation between outcome measures X and Y then it may be feasible to use a composite outcome measure provided the measures are timed such that the treatment effect is of similar magnitude on both components. In the case diagrammed in Fig. 4 this is clearly a difficult decision if X and Y must be measured simultaneously, e.g., as part of the same interview process. However, the situation is less problematic if X and Y can be measured independently at the optimal time for each measurement. But even in this case, based on the information from Table 1, given

![Fig. 4. Conceptual model for determining the utility of a continuous composite outcome measure.](image-url)
the high correlation between measures of X and Y it is not expected that the decision to use a composite would provide an efficacy gain in this case. Thus, composite outcome measures that assess treatment effects on multiple dimensions of a complex disease may be challenging to put into practice.

7. Conclusions

This article demonstrates that the use of a composite outcome measure formed from normally distribution components provides statistical advantages over using a single outcome measure only in specific cases, the majority of which require advanced knowledge regarding the biological mechanism of the therapy and flexibility in how the components of the composite outcome are measured. Given the magnitude of the potential disadvantage to using a continuous composite outcome—potentially increasing required sample size 4-fold under certain circumstances—it is recommended that continuous composite outcome measures not be used in cases where the efficacy profile of the therapy is not well understood. General examples of such scenarios are early phase clinical trials where there is limited experience with multiple outcome measures, or studies of products where the therapeutic mechanism is poorly understood. In these scenarios, the use of a continuous composite can have unintended negative consequences. Lastly, we note that even in circumstances where a continuous composite outcome is advantageous statistically, the use of such an outcome may introduce difficulty in interpreting results. Difficulties in interpretation are likely to be exacerbated when the components measure different underlying biological or clinical phenomena, or when the components are measured at different time points relative to baseline. Therefore, even in cases where a continuous composite outcome measure would lead to an efficacy gain, this should be weighed carefully against any potential loss in interpretation of trial results.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

8. Author declaration

1) We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

2) We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

3) We confirm that neither the entire paper nor any of its content has been submitted, published, or accepted by another journal. The paper will not be submitted elsewhere if accepted for publication in the Journal.

4) We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

5) We confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

6) We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

CRedit statement

Troy: Conceptualization, Methodology, Software, Formal analysis, Writing-Original Draft, Visualization. Simmons: Conceptualization, Methodology, Validation, Writing - Review & Editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concte.2020.100655.

References

[1] Ich Harmonised Tripartite Guideline, Statistical principles for clinical trials. International conference on harmonisation E9 expert working group, Stat. Med. 18 (15) (1999) 1905–1942.

[2] N. Freemantle, M. Calvert, J. Wood, J. Eastaugh, C. Griffin, Composite outcomes in randomized trials: greater precision but with greater uncertainty?, J. Am. Med. Assoc. 289 (19) (2003) 2554–2559.

[3] Y.T. Wang, L.A. Kalenbach, C.P. Cannon, et al., Effect of medication Co-payment vouchers on P2Y12 inhibitor use and major Adverse cardiovascular events among patients with myocardial infarction: the ARTEMIS randomized clinical trial, J. Am. Med. Assoc. 321 (1) (2019) 44–55.

[4] D.A. Schoenfeld, Sample-size formula for the proportional-hazards regression model, Biometrics 39 (2) (1983) 499–503.

[5] N. Freemantle, M. Calvert, Weighing the pros and cons for composite outcomes in clinical trials, J. Clin. Epidemiol. 60 (2007) 658–659.

[6] F. Ibrahim, B.D. Tom, D.L. Scott, A.T. Prevost, A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes, Trials 17 (1) (2016) 272.

[7] H. Inojosa, D. Schriefer, T. Ziemssen, Clinical outcome measures in multiple sclerosis: a review, Autoimmun. Rev. (2020) 102512.

[8] H. Liu-Seifert, S. Andersen, M. Cae, et al., Statistical properties of continuous composite scales and implications for drug development, J. Biopharm. Stat. 27 (6) (2017) 1104–1114.