

Misspecification of Cox regression models with composite endpoints

Longyang Wu and Richard J. Cook^{*†}

Researchers routinely adopt composite endpoints in multicenter randomized trials designed to evaluate the effect of experimental interventions in cardiovascular disease, diabetes, and cancer. Despite their widespread use, relatively little attention has been paid to the statistical properties of estimators of treatment effect based on composite endpoints. We consider this here in the context of multivariate models for time to event data in which copula functions link marginal distributions with a proportional hazards structure. We then examine the asymptotic and empirical properties of the estimator of treatment effect arising from a Cox regression model for the time to the first event. We point out that even when the treatment effect is the same for the component events, the limiting value of the estimator based on the composite endpoint is usually inconsistent for this common value. We find that in this context the limiting value is determined by the degree of association between the events, the stochastic ordering of events, and the censoring distribution. Within the framework adopted, marginal methods for the analysis of multivariate failure time data yield consistent estimators of treatment effect and are therefore preferred. We illustrate the methods by application to a recent asthma study. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: copula; composite endpoint; Cox regression; model misspecification; randomized clinical trial

1. Introduction

Many diseases put individuals at elevated risk for a multitude of adverse clinical events, and researchers routinely design randomized clinical trials to evaluate the effectiveness of experimental interventions for the prevention of these events. Trials in cardiology, for example, record times of events such as non-fatal myocardial infarction, non-fatal cardiac arrest, and cardiovascular death [1]. In cerebrovascular disease, patients with carotid stenosis can be treated with medical therapy or surgery, and trials evaluating their relative effectiveness may record endpoints such as strokes ipsilateral to the surgical site, contralateral strokes, and death [2]. In oncology, researchers often design trials to study treatment effects on disease progression and death [3], but palliative trials of patients with skeletal metastases may be directed at preventing skeletal complications including vertebral and non-vertebral fractures, bone pain, and the need for surgery to repair bone [4]. In these and many other settings, although interest lies in preventing each of the respective events, it is generally infeasible to conduct studies to answer questions about each component.

When one type of event is of greater clinical importance than others, it can be chosen as the basis of the primary treatment comparison, and effects on other types of events can then be assessed through secondary analyses. When two or more events are of comparable importance, co-primary endpoints can be specified, but tests of hypotheses must typically control the experimental type I error rate through multiple comparison procedures [5–7]; these make decision analyses more complex. A seemingly simple alternative strategy is to adopt a so-called composite event [8,9] that is said to have occurred if any one of a set of component events occurs. The time of the composite event is therefore the minimum of the times of all component events.

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

^{*}Correspondence to: Richard J. Cook, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

[†]E-mail: rjcook@uwaterloo.ca

There are several additional reasons investigators may consider the use of composite endpoints in clinical trials. In studies involving a time-to-event analysis, the use of a composite endpoint will mean that more events will be observed than would be observed for any particular component. If the same clinically important effect is specified for the composite endpoint and one of its components, this increased event rate will translate into greater power for tests of treatment effects; at the design stage this translates to a reduction in the required number of subjects or duration of follow-up [9–11]. Composite endpoints are routinely adopted through the introduction of one or more less serious events, however, which presumably warrants revising the clinically important effect of interest. Moreover, we show later that with models featuring a high degree of structure, model assumptions may not even be compatible for the composite endpoint and one of its components.

In time-to-event analyses, interest may lie in the effect of an experimental treatment versus standard care on the risk of a non-fatal event. This is a common framework in trials of patients with advanced diseases where interest lies in improving quality of life through the prevention of complications. In such settings, individuals are at considerable risk of death and a competing risks problem arises. Investigators often deal with this by adopting a composite endpoint based on the time to the minimum of the non-fatal event of interest and death [12, 13]. This strategy leads to an ‘event-free survival’ analysis that is particularly common in cancer where progression-free survival is routinely adopted as a primary endpoint [14]. In palliative trials, however, a treatment may not be expected to have an effect of survival, and if a non-negligible proportion of individuals die before experiencing the clinical event of interest, this analysis can lead to a serious underestimation of the effect of the treatment [10, 15].

Recommendations are available in the literature on how to design trials, analyze resultant data, and report findings when composite endpoints are to be used [10–12, 16]. The main recommendations include that (i) individual components should have similar frequency of occurrence, (ii) the treatment should have a similar effect on all components, (iii) individual components should have similar importance to patients, (iv) data from all components should be collected until the end of trial, and (v) individual components should be analyzed and reported separately as secondary endpoints. The first three recommendations have face validity and seem geared towards helping ensure that conclusions regarding treatment effects on the composite endpoint have some relation to treatment effects on the component endpoints, thus helping in the interpretation of results. The collection of data on the occurrence of the component endpoints until the end of the trial facilitates separate assessment of treatment effects on each of the component endpoints. This means the consistency of findings across components can be empirically assessed.

The aforementioned issues have been actively debated in the medical literature [11, 16–19], but there has been relatively little formal statistical investigation of these points. In this paper, we discuss statistical considerations related to composite endpoint analyses and use the recommendations to guide the investigation. Because the Cox regression model is routinely adopted for the analysis of composite endpoints in clinical trials [12], we consider it here and point out important issues regarding model specification and interpretation. We formulate multivariate failure time models with proportional hazards for the marginal distributions that may be used to reflect the settings where composite endpoints are most reasonable according to the current guidelines. We study the asymptotic and empirical properties of estimators arising from a composite endpoint analysis. We also explore the utility of marginal methods based on multivariate failure time data [20]. We argue that the belief that composite endpoints provide an overall measure of the effect of treatment is overly simplistic, and a thoughtful interpretation of intervention effects based on composite endpoints alone is difficult. Their use as a primary basis for treatment comparison in clinical trials therefore warrants careful consideration.

The remainder of this paper is organized as follows. In Section 2, we construct bivariate failure time distributions for which the marginal distributions have proportional hazards between two treatment groups. We then derive the distribution for the time to the first event and show that it does not typically feature proportional hazards across the two treatment groups. We use large sample theory for misspecified models to derive the limiting value of the log hazard ratio from a naive Cox model, and empirical studies demonstrate finite sample properties which are in close alignment with the theory. An alternative approach to synthesizing data over component events is to conduct a global analysis on the basis of the marginal methods of Wei *et al.* [20]; we explore this in Section 3. An application to a recently completed asthma management study illustrates the various methods in Section 4, and we make the concluding remarks in Section 5.

2. Multivariate failure time distributions via copula functions

2.1. Construction of joint distributions based on copula functions

If $(U_1, U_2)'$ is a bivariate random variable with standard uniform margins on $[0, 1]$, a two-dimensional copula function can be defined as

$$C(u_1, u_2) = P(U_1 \geq u_1, U_2 \geq u_2), \quad (1)$$

[21]. If there exists a convex decreasing function $\mathcal{H}(u; \theta)$ such that $\mathcal{H} : (0, 1] \rightarrow [0, \infty)$ and $\mathcal{H}(1; \theta) = 0$, and if the copula function can be written as

$$C(u_1, u_2; \theta) = \mathcal{H}^{-1}(\mathcal{H}(u_1; \theta) + \mathcal{H}(u_2; \theta); \theta),$$

then copula belongs to the *Archimedean family*; the univariate function $\mathcal{H}(u; \theta)$ is called the *generator* for the copula [22]. Suppose $(U_{i1}, U_{i2})'$ and $(U_{j1}, U_{j2})'$ are two random variables drawn from the joint distribution (1). A common measure of the association between U_1 and U_2 is Kendall's τ , defined as

$$\tau_\theta = P \{ (U_{i1} - U_{j1})(U_{i2} - U_{j2}) > 0; \theta \} - P \{ (U_{i1} - U_{j1})(U_{i2} - U_{j2}) < 0; \theta \}.$$

where we write τ_θ to make the relation between θ and τ explicit.

For Archimedean copulas, this can be written as

$$\tau_\theta = 1 + 4 \int_0^1 \frac{\mathcal{H}(u; \theta)}{\mathcal{H}'(u; \theta)} du.$$

Copula functions have received considerable attention in the statistical literature in the past few years because they offer a convenient and attractive way of linking two marginal distributions to create a joint survival function [23]. Suppose T_1 and T_2 are a pair of non-negative random variables with respective survivor functions $\mathcal{F}_1(t_1|z; \alpha_1)$ and $\mathcal{F}_2(t_2|z; \alpha_2)$ given a covariate z . If we let $U_1 = \mathcal{F}_1(T_1|z; \alpha_1)$ and $U_2 = \mathcal{F}_2(T_2|z; \alpha_2)$ where α_k indexes the marginal distribution for $T_k|z$, then $U_k \sim \text{UNIF}(0, 1)$, $k = 1, 2$. We can define the bivariate 'survival' distribution function of (U_1, U_2) through a copula as in (1) and obtain a joint survivor function for $(T_1, T_2)'$ given Z as

$$\mathcal{F}_{12}(t_1, t_2|z; \Omega) = P(T_1 \geq t_1, T_2 \geq t_2|z; \Omega) = C_\theta(\mathcal{F}_1(t_1|z; \alpha_1), \mathcal{F}_2(t_2|z; \alpha_2); \theta), \quad (2)$$

where $\Omega = (\alpha', \theta)'$ with $\alpha = (\alpha'_1, \alpha'_2)'$. Because Kendall's τ is invariant to monotonic increasing or decreasing transformations [21], it can be interpreted as a measure of association of the transformed variables $(T_1, T_2)'$ given Z . The use of a copula function to define the joint distribution of $(T_1, T_2)|z$ is particularly appealing because one can specify the marginal distributions to have a proportional hazards form; this is not typically possible for joint distributions induced by random effects or intensity-based analyses.

If a composite endpoint analysis is planned, it would be based on modeling the random variable $T = \min(T_1, T_2)$, which has survival, density, and hazard function conditional on Z , given by

$$P(T \geq t|z) = \mathcal{F}(t|z; \Omega) = \mathcal{F}_{12}(t, t|z; \Omega), \quad (3)$$

$f(t|z) = -d\mathcal{F}(t|z; \Omega)/dt$, and $\lambda(t|z; \Omega) = -d \log \mathcal{F}(t|z; \Omega)/dt$, respectively. Suppose Z is a binary indicator where $Z = 1$ for individuals in a treatment group and $Z = 0$ otherwise. A key point is that the hazard ratio $\lambda(t|z = 1; \Omega)/\lambda(t|z = 0; \Omega)$ is not, in general, independent of time. As a result, even if the marginal distributions feature proportional hazards, the model for the composite endpoint will typically not. We study this point further in the next four settings for three different Archimedean copulas and the case of independent components.

2.1.1. *Composite endpoint analysis based on a Clayton copula.* The Clayton copula [24] is a member of the Archimedean family with generator $\mathcal{H}(u; \theta) = u^{-\theta} - 1$, $\mathcal{H}^{-1}(v; \theta) = (v + 1)^{-1/\theta}$ and copula function

$$C(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \quad (4)$$

with $\theta \geq -1$. Kendall's τ is then given by $\tau_\theta = \theta/(\theta + 2)$, which can be seen to vary over $[-1, 1]$.

Consider the joint distribution of $(T_1, T_2)|Z$ in which the marginal distribution for $T_k|Z$, $k = 1, 2$ features proportional hazards; so $\lambda_k(t|z) = \lambda_{k0}(t) \exp(\beta_k z)$ with $\Lambda_k(t|z) = \Lambda_{k0}(t) \exp(\beta_k z)$ where $\Lambda_{k0}(t) = \int_0^t \lambda_{k0}(s) ds$, $k = 1, 2$. If the joint survivor function $\mathcal{F}_{12}(t_1, t_2|z; \Omega)$ is determined by the Clayton copula through (2), by (3) the survivor function of the failure time $T = \min(T_1, T_2)$ given z is

$$\mathcal{F}(t|z; \Omega) = \left[\exp(\theta \Lambda_{10}(t) e^{\beta_1 z}) + \exp(\theta \Lambda_{20}(t) e^{\beta_2 z}) - 1\right]^{-1/\theta}. \quad (5)$$

Hence, the hazard ratio for the treatment versus control groups for the composite endpoint is

$$\frac{\lambda(t|z = 1; \Omega)}{\lambda(t|z = 0; \Omega)} = \frac{\left[\sum_{k=1}^2 \lambda_{k0}(t) \exp(\beta_k + \theta \Lambda_{k0}(t) e^{\beta_k})\right] / \left[\sum_{k=1}^2 \exp(\theta \Lambda_{k0}(t) e^{\beta_k}) - 1\right]}{\left[\sum_{k=1}^2 \lambda_{k0}(t) \exp(\theta \Lambda_{k0}(t))\right] / \left[\sum_{k=1}^2 \exp(\theta \Lambda_{k0}(t)) - 1\right]}, \quad (6)$$

which is not invariant with respect to time in general.

To gain some insight into this function, suppose the marginal distributions are exponential with common baseline hazards of $\lambda_{10}(t) = \lambda_{20}(t) = \lambda = \log 10$ so that the probability of a type k event occurring before $t = 1$ is 0.90 for a control subject (i.e., $P(T_k < 1|Z = 0) = 0.90$). Further suppose that a common hazard ratio of 0.50 holds for the two margins (i.e., $\exp(\beta_1) = \exp(\beta_2) = 0.50$). This setting is consistent with the recommendations that the component events occur with comparable frequency because $P(T_1 < T_2|Z) = 0.5$, and have comparable treatment effects ($\beta_1 = \beta_2$). Figure 1(a) contains a plot of the hazard ratio (6) over the time interval $[0, 1]$ for models with mild ($\tau_\theta = 0.2$), moderate ($\tau_\theta = 0.40$), and strong ($\tau_\theta = 0.60$) association. As can be seen, even when the treatment effects

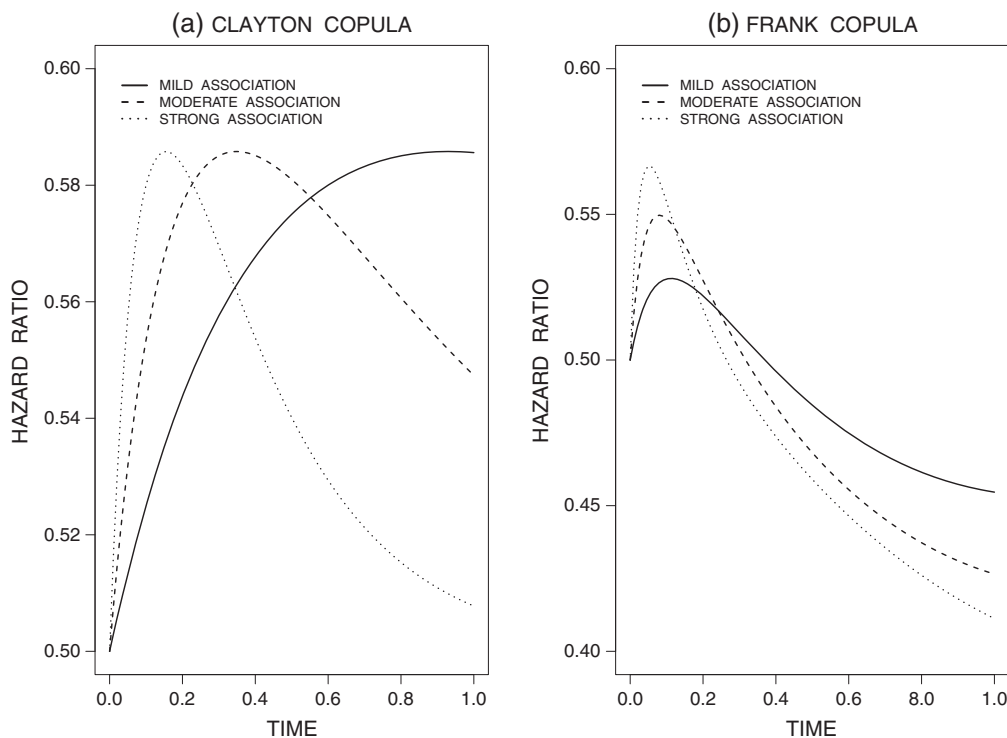


Figure 1. Plots of the hazard ratio over the time interval $[0, 1]$ for the composite endpoint model implied by the Clayton copula (panel (a)) and Frank copula (panel (b)) with marginal exponential distributions with $\lambda_1 = \lambda_2 = \log 10$ and $\exp(\beta_1) = \exp(\beta_2) = \exp(\beta) = 0.50$ and mild ($\tau_\theta = 0.20$), moderate ($\tau_\theta = 0.40$), and strong ($\tau_\theta = 0.60$) associations.

are the same for the two component endpoints, there can be non-negligible variation in the hazard ratio over time, and within this family of models, the nature of this variation depends on the strength of the association between the two failure times.

2.1.2. Composite endpoint analysis based on a Frank copula. The generator for the Frank copula [25] is $\mathcal{H}(u; \theta) = -\log((\exp(-\theta u) - 1)/(\exp(\theta) - 1))$, and the resulting copula function is

$$C(u_1, u_2; \theta) = -\theta^{-1} \log \left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right],$$

where $\theta \in \mathfrak{R}$; Kendall's τ is then $\tau_\theta = 1 - 4\theta^{-1} + 4\theta^{-2} \int_0^\theta t/(\exp(t) - 1)dt$. If we adopt the same marginal distributions as before, the survivor function for the composite endpoint is

$$\mathcal{F}(t|z) = -\frac{1}{\theta} \log \left[1 + \frac{(\exp(-\theta e^{-\Lambda_1(t)e^{\beta_1 z}}) - 1)(\exp(-\theta e^{-\Lambda_2(t)e^{\beta_2 z}}) - 1)}{e^{-\theta} - 1} \right],$$

but because $\lambda(t|z; \Omega) = -d \log \mathcal{F}(t)/dt$, the hazard ratio $\lambda(t|z = 1; \Omega)/\lambda(t|z = 0; \Omega)$ has a complicated form. Figure 1(b) contains a plot of this hazard ratio over $[0, 1]$, and as in the case of the Clayton copula, there is considerable variation in this ratio over time.

2.1.3. Composite endpoint analysis based on a Gumbel–Hougaard copula. The generator for the Gumbel–Hougaard [26] copula is $\mathcal{H}(u; \theta) = (-\log t)^\theta$ giving

$$C(u_1, u_2; \theta) = \exp \left(-((-\log u_1)^\theta + (-\log u_2)^\theta)^{\theta^{-1}} \right),$$

for $\theta \geq 1$; Kendall's τ is given by $\tau_\theta = (\theta - 1)/\theta$. The corresponding survivor function for the composite endpoint is

$$\mathcal{F}(t|z) = \exp \left(- \left[(\Lambda_1(t)e^{\beta_1 z})^\theta + (\Lambda_2(t)e^{\beta_2 z})^\theta \right]^{\theta^{-1}} \right),$$

and if $\beta_1 = \beta_2 = \beta$, the hazard is

$$\lambda(t|z) = \exp(\beta z) \left[\frac{(\Lambda_1(t)^\theta + \Lambda_2(t)^\theta)^{\theta^{-1}-1}}{\lambda_1(t)\Lambda_1(t)^{\theta-1} + \lambda_2(t)\Lambda_2(t)^{\theta-1}} \right].$$

Interestingly, the hazard ratio in this case is $\exp(\beta)$, which means that the proportional hazards model for the composite endpoint is compatible with a proportional hazards model for the margins. If the hazard ratio is in fact common for the component endpoints, then a consistent estimator will be obtained for this common effect on the basis of a Cox model for the composite endpoint.

2.1.4. Composite endpoint analysis with independent components. Here, we consider the setting where the component failure times are independent; a special case of $\tau_\theta = 0$ for the joint models in Sections 2.1.1–2.1.3. In this case, the hazard ratio for the composite endpoint analysis reduces to

$$\frac{\lambda(t|z = 1; \alpha)}{\lambda(t|z = 0; \alpha)} = \frac{\lambda_{10}(t) \exp(\beta_1) + \lambda_{20}(t) \exp(\beta_2)}{\lambda_{10}(t) + \lambda_{20}(t)}.$$

With nonhomogeneous hazards, it is apparent that the composite endpoint analysis is only compatible with a proportional hazards assumption if either (A.1) $\beta_1 = \beta_2 = \beta$ or (A.2) $\lambda_{10}(t) = \lambda_{20}(t)$. If $\beta_1 = \beta_2 = \beta$, then a consistent estimate of this common effect is obtained in a composite endpoint analysis. If $\beta_1 \neq \beta_2$ but the hazard functions are identical, the multiplicative effect is $(\exp(\beta_1) + \exp(\beta_2))/2$. If assumptions A.1 and A.2 do not hold, then the ratio is a complicated time varying function of the baseline hazards and respective treatment effects.

2.2. Misspecification of the Cox model with composite endpoints

The previous section demonstrated that the composite endpoint analysis is typically based on a misspecified Cox regression model if the marginal distributions satisfy the proportional hazards assumption. In this section, we investigate the frequency properties of estimators from a composite endpoint analysis when the component endpoints are associated through a copula function.

Let $T_i = \min(T_{i1}, T_{i2})$ denote the time of the composite endpoint for individual i in a sample of size m . Let $\{N_i(s), s < 0\}$ denote the counting process for subject i , which indicates the occurrence of the composite endpoint, so that $dN_i(s) = 1$ if $T_i = s$ and is zero otherwise. Suppose that it is planned to follow all subjects over the interval $(0, C^\dagger]$ but that subjects may be lost to follow-up or withdraw from the study prematurely. Let W_i represent the withdrawal time for subject i and $C_i = \min(W_i, C^\dagger)$ denote their right censoring time. Let $Y_i(s) = I(s \leq T_i)$ indicate whether subject i is at risk of the composite endpoint at time s , $Y_i^\dagger(s) = I(s \leq C_i)$ indicate whether they are under observation at time s , and $\bar{Y}_i(s) = Y_i^\dagger(s)Y_i(s)$ indicate whether they are event free and under observation. The observable counting process for the response is then based on $d\bar{N}_i(s) = \bar{Y}_i(s)dN_i(s)$ for subject i . The data for a sample of size m then consist of $\{(\bar{Y}_i(s), d\bar{N}_i(s)), 0 < s \leq Z_i, i = 1, \dots, m\}$, which if we let $\bar{Y}(s) = (\bar{Y}_1(s), \dots, \bar{Y}_m(s))'$, $d\bar{N}(s) = (d\bar{N}_1(s), \dots, d\bar{N}_m(s))'$ and $Z = (Z_1, \dots, Z_m)'$, we may write more compactly as $\{(\bar{Y}(s), d\bar{N}(s)), 0 < s, Z\}$.

The Cox model is widely used in the analysis of composite endpoints [27] to estimate the relative hazard where we assume the hazard function for $T_i|z_i$ to have the form

$$\psi(t|z_i) = \psi_0(t) \exp(\alpha z_i), \tag{7}$$

where $\psi_0(t)$ is a non-negative baseline hazard function corresponding to the control group and z_i is the treatment covariate for individual i , $i = 1, \dots, m$. The treatment effect α can be estimated using the maximum partial likelihood [28] by solving

$$U(\alpha) = \sum_{i=1}^m \int_0^\infty \bar{Y}_i(t) \left(z_i - \frac{S^{(1)}(\alpha, t)}{S^{(0)}(\alpha, t)} \right) dN_i(t) \tag{8}$$

where $S^{(k)}(\alpha, t) = \sum_{i=1}^m \bar{Y}_i(t) z_i^k \exp\{\alpha z_i\}$, $k = 0, 1$.

If $\{Y_i^\dagger(s), 0 < s\}$ is independent of $\{N_i(s), 0 < s\}$ given Z_i and if (7) is correctly specified, then (8) has expectation zero and the solution $\hat{\alpha}$ is consistent for the true value, α . In the independence case, this true value is β if the treatment effect is common (i.e., under A.1 $\beta = \beta_1 = \beta_2$) or $\alpha = \log(\exp(\beta_1) + \exp(\beta_2))/2$ if the baseline hazard functions are the same (i.e., under A.2). More generally, however, $\hat{\alpha}$ is consistent for α^* , the solution to expected score function $\mathcal{U}(\alpha) = E(U(\alpha))$ given by

$$\mathcal{U}(\alpha) = \int_0^\infty \left\{ E \left(\sum_{i=1}^m Z_i \bar{Y}_i(t) dN_i(t) \right) - \frac{E(S^{(1)}(\alpha, t))}{E(S^{(0)}(\alpha, t))} E \left(\sum_{i=1}^m \bar{Y}_i(t) dN_i(t) \right) \right\}, \tag{9}$$

where the expectation E is with respect to the true model for $\{(\bar{Y}(s), d\bar{N}(s)), 0 < s, Z\}$ [29–31]. By using the true model based on (3) and assuming independent censoring for the withdrawal time W_i with survival distribution $\mathcal{G}(w|z) = \mathcal{G}(w)$, these expectations can be obtained as follows:

$$\begin{aligned} E(S^{(1)}(\alpha, t)) &= m[\mathcal{G}(t)\mathcal{F}(t|Z; \Omega)] \exp(\alpha) P(Z = 1) \\ E(S^{(0)}(\alpha, t)) &= m[\mathcal{G}(t)\mathcal{F}(t|Z = 1; \Omega)] \exp(\alpha) P(Z = 1) + m[\mathcal{G}(t)\mathcal{F}(t|Z = 0; \Omega)] P(Z = 0). \end{aligned}$$

Likewise,

$$\begin{aligned} E \left(\sum_{i=1}^m \bar{Y}_i(t) dN_i(t) \right) &= m\mathcal{G}(t) \sum_{r=0}^1 \mathcal{F}(t|Z = r; \Omega) \lambda(t|Z = r) P(Z = r), \\ E \left(\sum_{i=1}^m Z_i \bar{Y}_i(t) dN_i(t) \right) &= m\mathcal{G}(t) [\mathcal{F}(t|Z = 1; \Omega) \lambda(t|Z = 1) P(Z = 1)]. \end{aligned}$$

To illustrate the bias resulting from a composite endpoint analysis, consider a randomized clinical trial in which subjects are to be followed over the interval $(0, C^\dagger]$ where $C^\dagger = 1$. Let $Z = 1$ for treated subjects and $Z = 0$ for control subjects and suppose $P(Z = 1) = 1 - P(Z = 0) = 0.5$. We set $\beta_1 = \beta_2 = \beta = \log 0.80$ to consider the case compatible with the current recommendations on the use of composite endpoints. We set λ_1 and λ_2 so that (i) $P(T_1 < T_2|Z = 0) = p_1$ equals a desired probability that the type 1 event occurs before the type 2 event among control subjects and that (ii) $P(C^\dagger < T) = \pi_A$ satisfies the administrative censoring rate for the composite endpoint among all subjects, where $\pi_A = 0.20$. Finally, suppose subjects may withdraw from the study early, and let W have an exponential distribution with rate ρ such that $P(C < T) = \pi$, where $P(C < T) = E_Z[P(W < T < C^\dagger|Z) + P(C^\dagger < T|Z)]$ and π is the overall censoring rate set to $\pi = 0.20, 0.40, 0.60,$ and 0.80 .

Figure 2 shows the limiting percent relative bias $(100(\alpha^* - \beta)/\beta)$ of the treatment coefficient from a composite endpoint analysis when the data are generated by a Clayton copula with mild ($\tau = 0.20$) and moderate ($\tau = 0.40$) association. We plotted this relative bias against $P(T_1 < T_2|Z = 0) = p_1$, and interestingly, the bias is greatest when $p_1 = 0.50$ but decreases as this probability approaches zero or one. In either of the extreme cases ($p_1 = 0$ or $p_1 = 1$), the composite endpoint coincides with the occurrence of a single endpoint, and a consistent estimate of the common treatment effect is obtained. Note that the bias $(\alpha^* - \beta)$ is positive, and hence, the limiting value of the treatment effect is more conservative than the true common value for each of the components. This means that the estimated value would, on average, under-represent the magnitude of the treatment effect on either component, a conclusion in line with the findings of [10, 15]. Moreover, we note that the common event rate and the common treatment effect are precisely the setting where composite endpoints are recommended for use [10–12, 16]. The plots also reveal the sensitivity of the limiting value to the degree of random censoring; the higher the censoring rate, the smaller the asymptotic bias. This highlights an important point that the limiting value of an estimator from a misspecified failure time model is highly sensitive to the censoring distribution even under independent censoring. By comparing the left and right panels in Figure 2, it

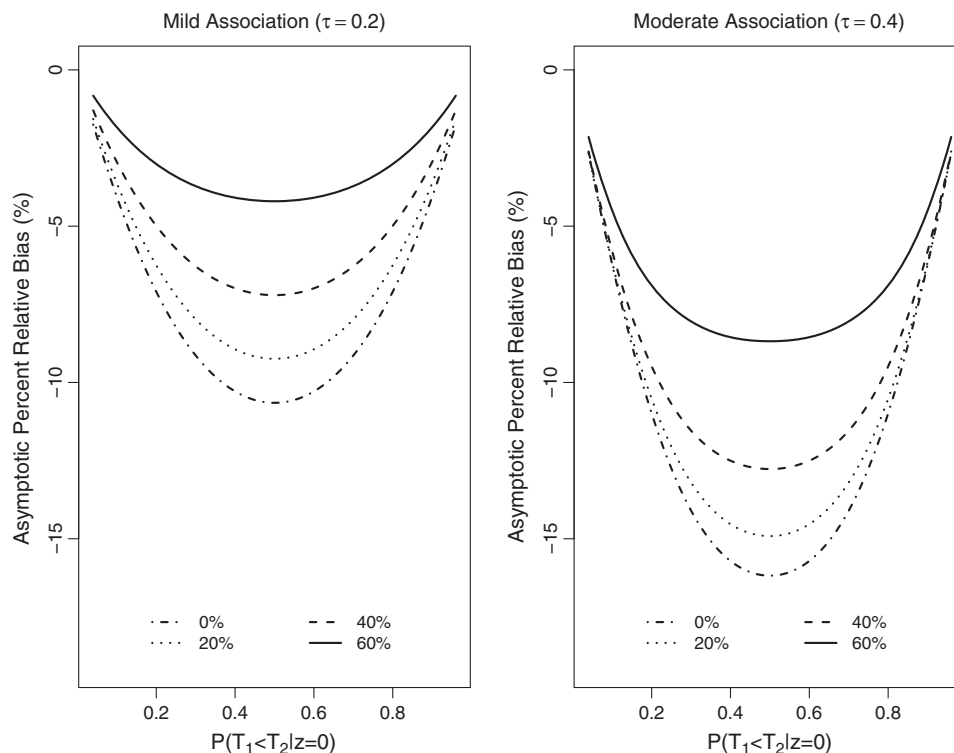


Figure 2. Asymptotic percent relative bias $(100(\alpha^* - \beta)/\beta)$ of Cox regression coefficient of treatment effect from composite endpoint analysis when bivariate failure times are generated by a Clayton copula; exponential margins, 20% administrative censoring ($\pi_A = 0.20$), 50:50 randomization, $\exp(\beta_1) = \exp(\beta_2) = 0.80$, and four different degrees of additional random censoring (none, 20%, 40%, and 60%).

is also apparent that the asymptotic bias is dependent on the degree of association between T_1 and T_2 ; the greater the association, the greater the asymptotic bias. This makes sense because when the event times are independent, consistent estimates should be obtained because assumptions A.1 and A.2 of Section 2.1.4 are satisfied.

Although of secondary interest, one can also show that $\hat{\psi}_0(t)$, $0 < t < C^\dagger$, is consistent for

$$\psi_0^*(t) = \frac{\sum_{r=0}^1 \mathcal{G}(t|Z=r)\mathcal{F}(t|Z=r)\lambda(t|Z=r)P(Z=r)}{\sum_{r=0}^1 \mathcal{G}(t|Z=r)\mathcal{F}(t|Z=r)\exp(\alpha^*I(r=1))P(Z=r)}$$

which when $P(Z=1) = 0.5$ and the censoring distribution is the same in the two groups reduces to

$$\psi_0^*(t) = [\mathcal{F}(t|Z=1)\lambda(t|Z=1) + \mathcal{F}(t|Z=0)\lambda(t|Z=0)]/[\mathcal{F}(t|Z=1)\exp(\alpha^*) + \mathcal{F}(t|Z=0)]$$

2.3. Simulation studies involving composite endpoints

2.3.1. Simulation design. Here, we simulate data from (2) to examine the empirical performance of estimators for finite samples. We assume that given Z , T_k has an exponential distribution with hazard $\lambda_k \exp(\beta_k Z)$, $k = 1, 2$, and model the association between T_1 and T_2 through a Clayton copula. We let $T = \min(T_1, T_2)$ denote the time of the composite endpoint as before. We suppose interest lies in following subjects over $(0, 1]$. As in the previous section, we determined the parameters λ_1 and λ_2 to satisfy the constraints $P(T_1 < T_2|Z=0) = p_1$, where $p_1 = 0.25$, $P(C^\dagger < T) = \pi_A$, and we set the administrative censoring rate to $\pi_A = 0.20$. We also incorporated random loss to follow-up with an exponential withdrawal time giving a net censoring rate of $\pi = 0.20, 0.40, 0.60$, and 0.80 subject to the constraint $\pi_A \leq \pi$.

For each parameter configuration, we derived the sample size for the composite endpoint analysis to achieve a prespecified power under the assumption that the Cox model in (7) holds. Therneau and Grambsch (2000) show that the required number of events is $D = 4(z_{1-\gamma_1} + z_{1-\gamma_2})^2/(\alpha^*)^2$, where z_q is the q th quantile of the standard Normal distribution, γ_1 is the type I error for a one-sided test, $1 - \gamma_2$ is the power, and α^* is the limiting value of treatment effect estimate obtained from (7). We focus on two-sided tests at the 5% significance level ($\gamma_1 = 0.05$) and sample sizes to achieve 80% power ($\gamma_2 = 0.20$). We calculated the required number of subjects as $m = D/P(T < C)$. In all simulation studies, we considered both equal treatment effects ($\beta_1 = \beta_2 = \beta = -.223$) and unequal treatment effects ($\beta_1 = -.223$ and $\beta_2 = 0$). For each parameter configuration, we generated 2000 replicates. We report the mean of the $\hat{\alpha}$ estimates, the empirical standard error (ESE), the average model-based standard error (ASE_1), and the average robust standard error (ASE_2). We also reported the empirical coverage probability (ECP%) of nominal 95% CIs for α^* based on robust standard errors and the empirical coverage probability of these intervals for β_1 (ECP%). The last column contains the empirical power (EP%) of a Wald test of the null hypothesis of no treatment effect.

2.3.2. Composite endpoints with dependent components. Table I contains the simulation results with dependent component times given by $\tau = 0.40$. The results for equal treatment effects are given in the top half of the table that we comment on first. The fourth column contains α^* , the limiting value of the estimator from the misspecified Cox model in (7). The fact that these values are all smaller in absolute value than the true common effects reveals the conservative nature of this limiting value, as already discussed in relation to Figure 2; the dependence of the limiting value on the degree of censoring is also apparent. This limiting value was used to derive the sample size (m) in the third column. The average estimator from the fitted Cox models reported in the fifth column closely approximates the limiting value. There is also close agreement between the empirical, average model-based, and average robust standard errors. The empirical coverage probabilities of the robust 95% CIs are very close to the nominal levels, and the empirical power is in good agreement with the nominal power of 80%. It is worth emphasizing that the empirical coverage probability is computed for the parameter α^* , not the common β ; for this latter parameter, the coverage rates are considerably lower.

In the bottom half of Table I, we reported the results for the case $\beta_1 \neq \beta_2$, where α^* is considerably smaller than β_1 . This smaller limiting value leads to considerably larger sample sizes to achieve

Table I. Frequency properties of estimators of treatment effect based on a composite endpoint with components arising from a Clayton copula: $p_1 = P(T_1 < T_2 | z = 0) = 0.25$, $\beta_1 = -0.223$, and $\tau = 0.4$

π_A	π	m	α^*	AVE($\hat{\alpha}$)	ESE	ASE ₁	ASE ₂	ECP*%	ECP%	EP%
<i>Common treatment effect: $\beta_2 = -0.223$</i>										
0.2	0.2	816	-0.195	-0.195	0.077	0.079	0.078	95.1	94.1	81.5
	0.4	1071	-0.196	-0.197	0.078	0.079	0.079	95.4	94.3	80.0
	0.6	1557	-0.199	-0.201	0.080	0.081	0.080	94.8	93.8	80.5
	0.8	2908	-0.206	-0.207	0.085	0.083	0.083	94.4	94.5	79.4
0.4	0.4	1076	-0.196	-0.197	0.079	0.079	0.079	95.1	93.1	80.4
	0.6	1557	-0.199	-0.201	0.081	0.080	0.080	94.7	93.6	79.8
	0.8	2907	-0.206	-0.208	0.084	0.083	0.083	95.5	95.0	78.8
0.6	0.6	1522	-0.202	-0.201	0.082	0.081	0.081	94.9	94.3	79.0
	0.8	2886	-0.207	-0.208	0.083	0.084	0.084	95.9	95.2	80.0
0.8	0.8	2779	-0.211	-0.208	0.087	0.085	0.085	94.8	94.1	78.5
<i>Different treatment effects: $\beta_2 = 0$</i>										
0.2	0.2	21743	-0.038	-0.038	0.015	0.015	0.015	94.9	0.0	78.4
	0.4	23103	-0.042	-0.042	0.017	0.017	0.017	94.9	0.0	79.4
	0.6	26037	-0.049	-0.049	0.019	0.020	0.020	95.5	0.0	79.5
	0.8	36581	-0.058	-0.058	0.024	0.023	0.023	94.2	0.0	79.3
0.4	0.4	19221	-0.046	-0.046	0.019	0.019	0.019	94.0	0.0	79.9
	0.6	24084	-0.051	-0.051	0.020	0.020	0.020	95.1	0.0	80.1
	0.8	36376	-0.058	-0.059	0.023	0.023	0.023	94.9	0.0	80.4
0.6	0.6	20656	-0.055	-0.055	0.022	0.022	0.022	94.9	0.0	81.8
	0.8	34960	-0.059	-0.060	0.024	0.024	0.024	95.0	0.0	80.5
0.8	0.8	30990	-0.063	-0.064	0.025	0.025	0.025	95.4	0.0	81.4

$\pi_A = P(C^\dagger < T)$ is the administrative censoring rate, $\pi = P(C^\dagger < T)$ is the net censoring rate, ESE is the empirical standard error, ASE₁ is the average model-based standard error, ASE₂ is the average robust standard error, ECP*% is the empirical coverage probability for α^* of nominal 95% CIs using the robust standard error, ECP% is the empirical coverage probability for β_1 of nominal 95% CIs using the robust standard error, and EP% is the empirical power of a Wald test of $H_0 : \alpha = 0$ based on the robust standard error.

the desired power. Again, however, we see close agreement between the average estimate and the limiting value, and very close agreement between the average model-based and average robust standard errors. The empirical coverage probability (for α^*) is also consistent with the nominal level, as is the empirical power.

2.3.3. Composite endpoints with independent components. Table II presents the simulation results with independent components (i.e., $\tau = 0$). The results in the top half of Table II reveal that the limiting value α^* is the same as the common value $\beta = \beta_1 = \beta_2$ as expected because assumption A.1 of Section 2 is satisfied. Again, the average point estimate is in close agreement with this common value, and the three standard errors are in close agreement. When the treatment has an effect on T_1 and not T_2 , α^* is again considerably smaller than β_1 . Note, however, even though this is a misspecified model, the limiting value does not depend on the censoring distribution. This much smaller value leads to larger sample size requirements than in the top half of the table. Because the first component T_1 happens less frequently than the second component T_2 (i.e., $P(T_1 < T_2 | Z = 0) = 0.25$), the limiting value from the misspecified Cox model is heavily attenuated in this setting. However, neither administrative nor random censoring appears to affect the limiting value of the estimator of treatment effect.

3. A multivariate semiparametric analysis

3.1. Limiting values for a Wei–Lin–Weissfeld analysis

In this section, we investigate the utility of the marginal approach of Wei *et al.* [20] for handling multivariate failure time data. This approach is based on formulating ordinary Cox models for each component

Table II. Frequency properties of estimators of treatment effect based on a composite endpoint with independent components: $p_1 = P(T_1 < T_2 | z = 0) = 0.25$, $\beta_1 = -0.223$.

π_A	π	m	α^*	AVE($\hat{\alpha}$)	ESE	ASE ₁	ASE ₂	ECP*%	ECP%	EP%
<i>Common treatment effect: $\beta_2 = -0.223$</i>										
0.2	0.2	644	-0.223	-0.224	0.090	0.090	0.090	95.6	95.6	79.5
	0.4	865	-0.223	-0.225	0.090	0.090	0.090	95.0	95.0	80.6
	0.6	1310	-0.223	-0.227	0.090	0.090	0.090	95.3	95.3	80.7
	0.8	2654	-0.223	-0.223	0.088	0.090	0.090	95.6	95.6	80.4
0.4	0.4	872	-0.223	-0.226	0.089	0.090	0.090	95.6	95.6	81.5
	0.6	1315	-0.223	-0.226	0.090	0.090	0.090	95.8	95.8	80.3
	0.8	2655	-0.223	-0.223	0.088	0.090	0.090	95.2	95.2	80.6
0.6	0.6	1323	-0.223	-0.223	0.091	0.090	0.090	95.1	95.1	79.9
	0.8	2660	-0.223	-0.223	0.088	0.090	0.090	95.3	95.3	80.4
0.8	0.8	2670	-0.223	-0.221	0.091	0.090	0.090	94.8	94.8	78.5
<i>Different treatment effects: $\beta_2 = 0$</i>										
0.2	0.2	11,750	-0.051	-0.052	0.021	0.021	0.021	94.4	0.0	80.6
	0.4	15,666	-0.051	-0.052	0.021	0.021	0.021	94.9	0.0	81.0
	0.6	23,499	-0.051	-0.052	0.021	0.021	0.021	94.7	0.0	80.3
	0.8	46,998	-0.051	-0.052	0.020	0.021	0.021	95.6	0.0	81.2
0.4	0.4	15,666	-0.051	-0.052	0.021	0.021	0.021	95.2	0.0	81.1
	0.6	23,499	-0.051	-0.052	0.021	0.021	0.021	95.3	0.0	80.1
	0.8	46,998	-0.051	-0.052	0.020	0.021	0.021	95.3	0.0	81.3
0.6	0.6	23,500	-0.051	-0.052	0.021	0.021	0.021	94.1	0.0	81.5
	0.8	46,998	-0.051	-0.052	0.020	0.021	0.021	95.6	0.0	81.4
0.8	0.8	46,999	-0.051	-0.051	0.021	0.021	0.021	94.7	0.0	80.6

$\pi_A = P(C^\dagger < T)$ is the administrative censoring rate, $\pi = P(C^\dagger < T)$ is the net censoring rate, ESE is the empirical standard error, ASE₁ is the average model-based standard error, ASE₂ is the average robust standard error, ECP*% is the empirical coverage probability for α^* of nominal 95% CIs using the robust standard error, ECP% is the empirical coverage probability for β_1 of nominal 95% CIs using the robust standard error, and EP% is the empirical power of a Wald test of $H_0 : \alpha = 0$ based on the robust standard error.

event to obtain component-specific estimates of treatment effect, and it is therefore compatible with the way the joint distributions were constructed using copula functions in Section 2. Estimation proceeds under a working independence assumption, as often adopted for analyses based on generalized estimating equations. We obtain a robust estimate of the covariance matrix, and then we obtain a global estimate of treatment effect by taking a weighted average of all component-specific estimates with weights chosen to minimize the variance of the global estimator. A key distinction between the global approach of Wei *et al.* [20] and the composite endpoint approach is that the former makes use of all observed events whereas the composite endpoint uses only information on the first event.

In the derivations that follow, the composite endpoint is comprised of K components, but we subsequently focus on the case $K = 2$. We let $dN_{ik}(s) = I(T_{ik} = s)$, $\{N_{ik}(s), 0 < s\}$ denote the counting process for type k events, and $\{N_i(s) = (N_{i1}(s), N_{i2}(s)), 0 < s\}$ denote the bivariate counting process for subject i , $i = 1, \dots, m$. Let $Y_{ik}(s) = I(s \leq T_{ik})$, $Y_i^\dagger(s) = I(s \leq C_i)$, and $\bar{Y}_{ik}(s) = Y_i^\dagger(s)Y_{ik}(s)$, $k = 1, \dots, K$, $i = 1, \dots, m$. The Cox model for the type k event is

$$\lambda_k(t|z_i) = \lambda_{k0}(t) \exp(\beta_k z_i),$$

where $\lambda_{k0}(t)$ is the baseline hazard function and β_k is the corresponding treatment effect. The k th component-specific score function for β_k is

$$U_k(\beta_k) = \sum_{i=1}^m \int_0^\infty \bar{Y}_{ik}(t) \left(Z_i - \frac{S_k^{(1)}(\beta_k, t)}{S_k^{(0)}(\beta_k, t)} \right) dN_{ik}(t), \quad (10)$$

where $S_k^{(r)}(\beta, u) = \sum_{i=1}^m \bar{Y}_{ik}(t) z_i^r \exp\{\beta_k z_i\}$, $r = 0, 1$.

Under the copula model of Section 2 with marginal distributions featuring proportional hazards, the solution to the score equation (10), $\hat{\beta}_k$, is consistent for the true treatment effect β_k . If we let

$\beta = (\beta_1, \dots, \beta_K)'$ and its estimate $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_K)'$, Wei *et al.* [20] show that $\sqrt{m}(\hat{\beta} - \beta)$ converges in distribution to a multivariate normal distribution with a zero-mean vector and variance-covariance matrix $\Sigma(\beta)$ and provides a consistent sandwich-type estimate for $\Sigma(\beta)$.

The global estimate of treatment effect proposed by Wei *et al.* [20] is a linear combination of all component-specific treatment effect estimates $\hat{\beta}_1, \dots, \hat{\beta}_K$ and can be obtained as

$$\hat{\bar{\beta}} = \mathbf{c}(\hat{\beta})' \hat{\beta}, \tag{11}$$

where the weight $\mathbf{c}(\hat{\beta}) = \hat{\Sigma}(\hat{\beta})^{-1} \hat{\mathbf{J}} [\hat{\mathbf{J}}' \hat{\Sigma}(\hat{\beta})^{-1} \hat{\mathbf{J}}]^{-1}$ is chosen to estimate the weight matrix to minimize the variance in the class of all linear estimators; $\hat{\Sigma}(\hat{\beta})$ is the estimate for the variance-covariance matrix of $\hat{\beta}$ and $\mathbf{J} = (1, \dots, 1)'$.

To compare the performances of the global approach and the composite endpoint approach, we obtain the limiting value of $\hat{\bar{\beta}}$ as

$$\bar{\beta}^* = \mathbf{c}(\beta)' \beta, \tag{12}$$

where $\mathbf{c}(\beta) = \Sigma^{-1}(\beta) \mathbf{J} [\mathbf{J}' \Sigma^{-1}(\beta) \mathbf{J}]^{-1}$. We therefore require the limiting value of the robust variance $\Sigma(\beta)$ to obtain the limiting value. The detailed derivations are deferred to the Appendix.

An alternative asymptotically equivalent approach to estimating the global effect and to deriving the limiting value involves specifying a single Cox regression model and fitting it using all events while ‘stratifying’ on the event type [32]. Although this has some appeal, we adopt the current framework on the basis of synthesizing estimates from separate Cox regression models because it makes explicit the fact that the global estimate and associated limiting value may be viewed as a weighted average of the component-specific estimates.

Table III. Empirical properties of the global estimates of treatment effect based on Wei–Lin–Weissfeld analysis: data were generated under a Clayton copula with $\tau = 0.40$, $\beta_1 = -0.223$.

π_A	π	m	$\bar{\beta}^*$	AVE($\hat{\beta}$)	ESE	ASE ₁	ASE ₂	ECP*%	ECP%	EP%
<i>Common treatment effect: $\beta_2 = -0.223$</i>										
0.2	0.2	621	-0.223	-0.223	0.084	0.072	0.086	95.9	95.9	83.6
	0.4	828	-0.223	-0.223	0.086	0.074	0.087	95.1	95.1	82.0
	0.6	1242	-0.223	-0.221	0.088	0.077	0.088	95.0	95.0	80.8
	0.8	2484	-0.223	-0.223	0.089	0.083	0.090	95.6	95.6	80.3
0.4	0.4	828	-0.223	-0.223	0.087	0.076	0.087	95.4	95.4	82.7
	0.6	1242	-0.223	-0.221	0.089	0.078	0.088	95.0	95.0	79.9
	0.8	2484	-0.223	-0.223	0.089	0.083	0.090	95.6	95.6	80.6
0.6	0.6	1242	-0.223	-0.223	0.090	0.081	0.089	95.1	95.1	79.7
	0.8	2484	-0.223	-0.222	0.089	0.083	0.090	95.2	95.2	80.5
0.8	0.8	2484	-0.223	-0.225	0.088	0.086	0.090	95.2	95.2	80.5
<i>Different treatment effects: $\beta_2 = 0$</i>										
0.2	0.2	7090	-0.066	-0.067	0.025	0.021	0.025	95.9	0.0	84.2
	0.4	9664	-0.065	-0.066	0.025	0.022	0.025	94.5	0.0	83.3
	0.6	14623	-0.065	-0.066	0.026	0.023	0.026	94.8	0.0	82.8
	0.8	28219	-0.066	-0.066	0.026	0.024	0.027	95.3	0.0	81.7
0.4	0.4	10203	-0.064	-0.065	0.025	0.022	0.025	95.1	0.0	83.6
	0.6	14897	-0.064	-0.066	0.025	0.023	0.025	94.6	0.0	83.2
	0.8	28316	-0.066	-0.066	0.026	0.024	0.027	95.2	0.0	80.6
0.6	0.6	14733	-0.065	-0.066	0.026	0.024	0.026	94.1	0.0	83.4
	0.8	28202	-0.066	-0.067	0.026	0.025	0.027	95.2	0.0	81.7
0.8	0.8	27355	-0.067	-0.069	0.026	0.026	0.027	95.4	0.0	82.2

$\pi_A = P(C^\dagger < T)$ is the administrative censoring rate, $\pi = P(C < T)$ is the net censoring rate, ESE is the empirical standard error, ASE₁ is the average model-based standard error, ASE₂ is the average robust standard error, ECP*% is the empirical coverage probability for $\bar{\beta}^*$ of nominal 95% CIs using the robust standard error, ECP% is the empirical coverage probability for β_1 of nominal 95% CIs using the robust standard error, and EP% is the empirical power of a Wald test of $H_0 : \bar{\beta} = 0$ based on the robust standard error.

3.2. Comparison of the global approach and the composite endpoint analysis

Table III reports the results from a global analysis of treatment effect based on the marginal analysis proposed by Wei *et al.* [20]. In this table, the sample sizes were computed on the basis of the formula for the composite endpoint analysis using the limiting value of the regression coefficient. As one would expect from (10), when the treatment effects are equal, then the marginal analysis yields consistent estimators for this common effect and the mean estimate across all simulated trials is very close to the limiting value. Moreover, the ESE and the average robust standard error were in very close agreement; the average model-based standard error is conservative because it is based on the working independence assumption being correct. The empirical coverage probabilities (based on the robust standard errors) were compatible with the nominal 95% level for β^* when $\beta_1 = \beta_2$. When $\beta_1 \neq \beta_2$, the empirical coverage for β_1 was zero, a reflection of the difference between β^* and β_1 . When $\beta_2 = 0$, the limiting value β^* was quite small, and hence, the sample sizes of the trial were much larger.

When $\beta_1 \neq \beta_2$, the composite endpoint and global analyses yield estimators that do not coincide with β_1 , β_2 , or each other. We next compare these limiting values. We consider the case in which two failure times are generated by a Clayton copula with exponential margins and a single treatment covariate modeled through proportional hazards with $\beta_1 = \log(0.80)$ and $\beta_2 = 0$. We consider mild and moderate association between the failure times with $\tau = 0.20$ and $\tau = 0.40$, respectively. Administrative censoring was set to 40% and additional random censoring from an exponential withdrawal time gave cases with 60% and 80% as well. The limiting values of the composite endpoint and global analyses were plotted against $P(T_1 < T_2 | Z = 0) = p_1$ in Figure 3. It is apparent that when p_1 approaches zero, the limiting value for both methods approaches 0. For the composite endpoint, this makes sense because the first event is most likely to be a type 2 event for which there is no treatment benefit. As p_1 approaches 1, the limiting value for the composite endpoint analysis approaches β_1 for analogous reasons. The limiting value from the global analyses tracks these limiting values quite well, but tend to correspond to larger estimates of treatment effect because the limiting value is larger in absolute value. Thus, even when the two components have equal frequencies and the proportional hazards assumption holds for each component, the global analysis, in the limit, will yield an estimate of treatment effect that is greater than that of the composite endpoint analysis. These relationships hold across both levels of association and over different degrees of censoring. Although we have restricted attention to the Clayton copula in these calculations and empirical studies, this investigation could be repeated under other copula models, and although the limiting values would differ, qualitatively similar findings would be expected.

4. Application to an asthma management study

We now apply both the composite endpoint analysis and the global approach to an asthma management study [33]. This is a two-phase, multicenter, randomized, parallel group effectiveness trial for comparing two treatment strategies for asthma management over a 2-year period. The control strategy is a ‘clinical strategy’, in which the treatment was guided on the basis of patient symptoms and spirometry readings. The experimental strategy is a so-called ‘sputum strategy’ (SS), whereby a cellular analysis of sputum samples was used to guide corticosteroid therapy use to keep eosinophils cell counts less than 2%. In phase I, a total of 107 patients were identified through the minimum treatment to maintain control. The aim of this asthma study was to investigate whether SS is more effective than clinical strategy on reducing the number and severity of exacerbations in phase II.

In our analysis, we focus on two types of exacerbations: mild exacerbations defined as requiring a daily maintenance dose of fluticasone of $< 250 \mu\text{g}$ and severe exacerbations defined here as requiring a minimum daily maintenance dose $\geq 250 \mu\text{g}$. The composite endpoint is defined as the time to the first of the two type of exacerbations. Figure 4 displays the empirical distribution function plots for the two component types of exacerbations and for the composite endpoint. It is apparent that the severe exacerbations occur much more frequently than mild exacerbations and thus represent the majority of the events contributing to the composite endpoint.

Table IV presents the results of the proportional hazards regression analysis in which the single binary covariate is the treatment indicator taking the value one for patients in the experimental (SS) group and zero otherwise. From these results, it is clear that the experimental SS strategy leads to a significantly lower hazard of severe exacerbations with a relative risk reduction of 47% (95% CI: 0.02, 0.71;

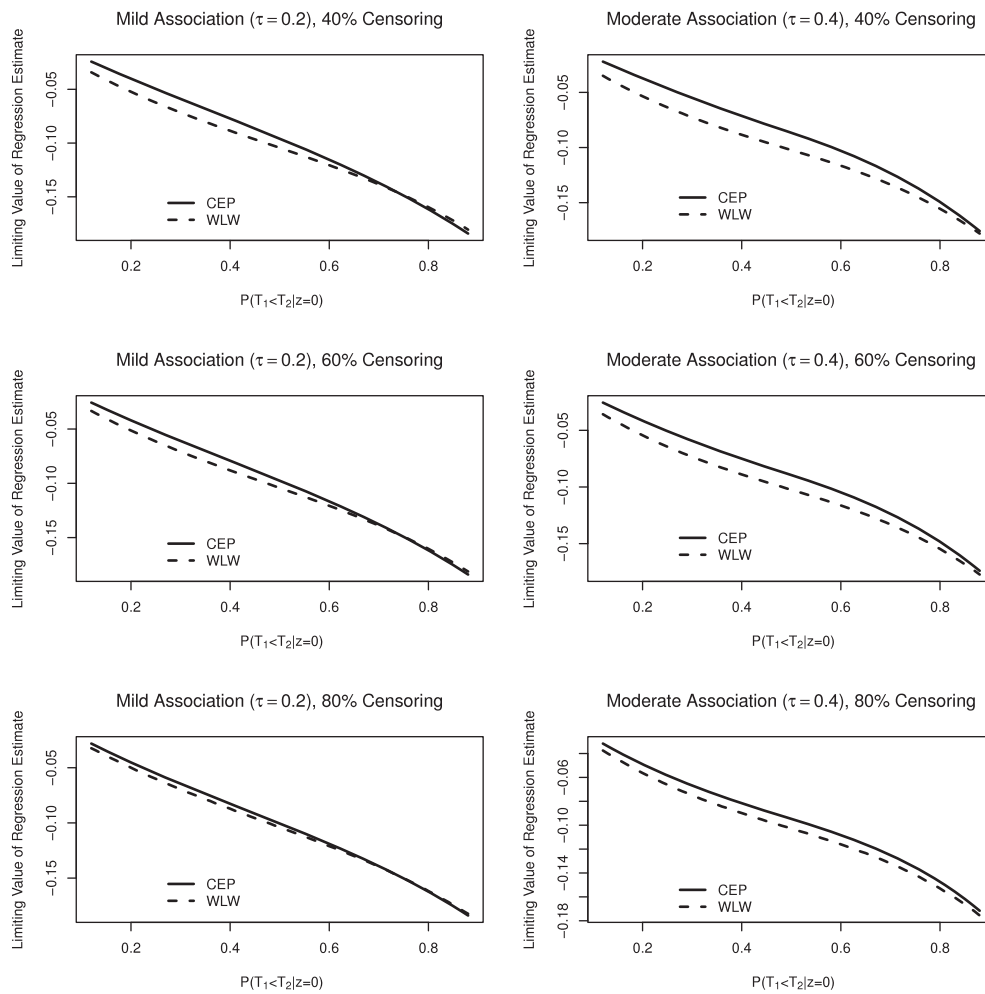


Figure 3. Plot of limiting values of regression estimates of treatment effect based on a composite endpoint analysis and a global Wei *et al.* [20] analysis with bivariate data generated with a Clayton copula; $\beta_1 = \log 0.80$, $\beta_2 = 0$.

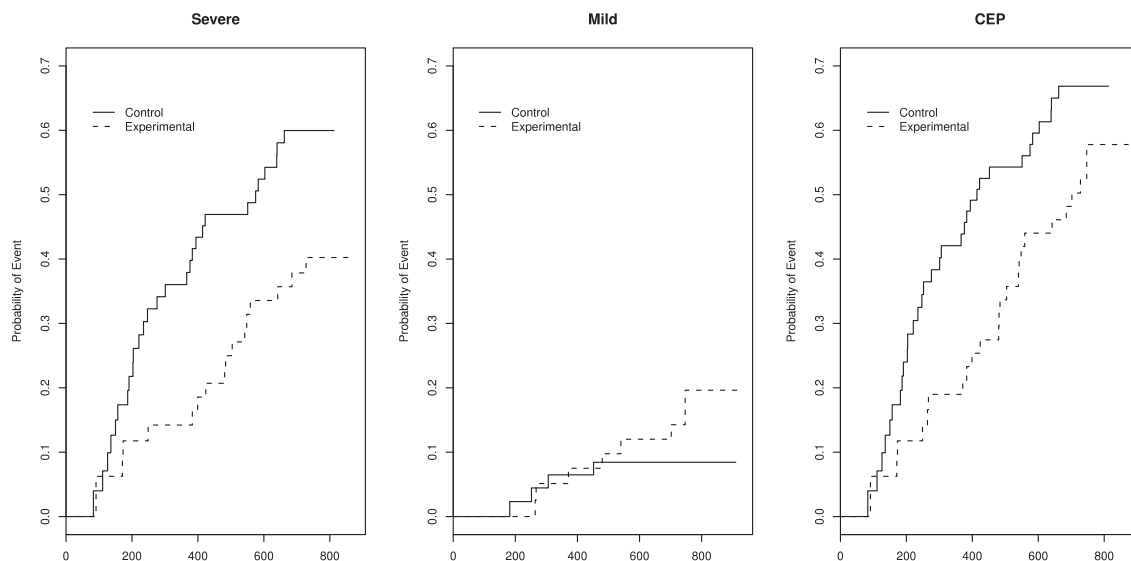


Figure 4. Empirical distribution functions for severe exacerbations, mild exacerbations, and the composite endpoint in asthma trial.

Table IV. Analysis results of the asthma management study.

Endpoint/Analysis	RR	95% CI	<i>p</i> -value	<i>p</i> *
Severe	0.53	(0.285, 0.977)	0.042	0.22
Mild	2.14	(0.624, 7.310)	0.227	0.11
Composite	0.665	(0.388, 1.138)	0.137	0.063
Global (WLW)	0.702	(0.405, 1.219)	0.209	

$p = 0.042$) but has little effect on the occurrence of mild exacerbations ($p = 0.227$). The result from the composite endpoint analysis is not statistically significant with $p = 0.137$. The Wei–Lin–Weissfeld [20] global analysis yields an estimate that is close to that obtained from the composite endpoint analysis, but it is apparent when examining the effects on the separate components that a global estimate is not an adequate summary of the data. The last column of Table IV gives the p -values for testing the proportional hazards assumption using univariate tests based on Schoenfeld residuals [32]. There is insufficient evidence to reject the null hypothesis of proportional hazards for each component, and the test yields a p -value just shy of statistical significance for the composite endpoint analysis at 0.063. Thus, although we have demonstrated that, in principle, if the proportional hazards assumption holds for the components of a composite endpoint, it generally does not hold for composite endpoint itself, the tests do not suggest problems with model fit for this particular data. Although the association may be well characterized by the Gumbel–Hougaard copula, the power of the tests for departures from proportional hazards may also be inadequate.

5. Discussion

Composite endpoints are widely adopted in clinical trials, and fitting a Cox proportional hazards model is the standard approach to estimating treatment effects on the basis of such endpoints. We have demonstrated that even when the treatment effects are the same for component endpoints under marginal Cox models, the Cox model for the composite endpoint is typically misspecified because the proportional hazards assumption does not in general hold. The estimator of treatment effect under such a misspecified Cox model for the composite endpoint has a slightly conservative limiting value, meaning that the benefit of treatment was under-estimated in the settings we examined. We found several factors that influence the limiting value including the strength of the association between the individual component events, stochastic ordering of the individual components, and the degree and nature of the censoring process; empirical studies corroborated these findings. Although we have not explored this here, it is clear from Section 2 that the specific copula function would also have an important effect. More generally, variation in the treatment effect across the individual components makes it even more difficult to interpret estimators.

Composite endpoints are often thought to offer a measure of the ‘overall effect’ of a treatment [9]. In fact, the opposite can be true if treatment effects are in opposing directions for different components, and one component tends to occur first. The event tending to occur first will have the greatest influence on the estimator of treatment effect based on the composite endpoint, masking the effect on the events that tend to occur later. In the case where the events occur with equal frequency and the treatment effect is the same for the components, the asymptotic calculations of Section 2 show that the estimate based on the composite endpoint suggests a smaller benefit that holds for the components. Although some might argue that this is therefore a conservative approach, it is essentially incorrect. The global approach of Wei *et al.* [20] can, however, provide evidence of this adverse effect in the component-wise analysis, and this will lead to an attenuation of the global effect in the weighted analysis.

Another rationale put forward for adopting composite endpoints is to model the event-free survival probability. For example, Sheehe [34] proposed that the event-free survival curve can be computed on the basis of Cox model estimates of hazard ratios from the composite endpoint containing mortality as a component. As we have demonstrated, estimators from Cox regression using composite endpoints can be attenuated by including components for which there is no treatment effect. Using estimates from the

composite endpoint analysis (event-free survival) may not provide a valid representation of the effect on the non-fatal event.

Two of the guidelines for the use of composite endpoints include the requirement that individual component events should be of roughly equal frequency, and the treatment effects should be comparable across all components. Our analytical and empirical investigation shows that these may not be sufficient conditions if interest lies in estimating these common effects in the sense that even when these conditions are satisfied, the association between the two events can lead to substantial bias in estimators based on composite endpoints. We support the recommendations that (i) data from all components should be followed until the end of the trial and (ii) individual components should be analyzed and reported separately. This alternative design and analysis facilitates a global approach [20] based on combining estimates from individual components, as well as assessment of whether this is appropriate. In the context of the copula-based joint model, we found that the global approach, in general, outperforms the composite endpoint analysis in terms of the properties of the resulting estimators and power or sample size requirements.

We have formulated a model with proportional hazards for each component event through the use of a copula function to reflect an idealized situation in alignment with the recommendations in the literature. We restricted attention to the situation with two component endpoints, but three or more components are often specified in practice. When multiple components are of interest, copula functions with an ‘exchangeable’ association structure can be readily adopted; more baseline marginal hazard functions and treatment effects would need to be specified. It is relatively straightforward to extend the derivations and empirical studies reported here for this setting but more challenging to cover a meaningful spectrum of settings, summarize results, and make recommendations.

Alternative frameworks could naturally be adopted for specifying models for correlated failure time data. One might, for example, consider intensity-based models where the risk of one type of event changes with the occurrence of another type of event. This could arise because of a biological mechanism in which the medical risk actually increases or if treating physicians alter the therapy being given. This formulation, although natural for characterizing the response process, is not compatible with proportional hazards for the marginal models. One might also consider frailty models for addressing the association between event times, but again, the marginal models will not have a proportional hazards form.

We have assumed independent censoring in this paper. Another way in which patients may be treated differently following the occurrence of a clinically important event is to be withdrawn from a study. The occurrence of one event may increase the risk an investigator may withdraw the patient from the study and result in response-dependent censoring. If the events are independent conditional on the treatment covariate, this will not pose a problem but otherwise will lead to biased estimates of the baseline hazard functions and treatment effects. Use of inverse probability of censoring weights will help reduce this bias, and this is currently under investigation.

Finally, we have focused on the frequency properties of estimators under a Cox regression models. Cox models are used routinely, but of course, the proportional hazards assumption may not be valid for either the marginal distributions or the composite endpoint. There is increasing interest in use of alternative regression models for the analysis of survival data including accelerated failure time models and additive models. Exploration of the behavior of estimators from such models warrants study.

Appendix: Derivation of the limiting value $\bar{\beta}$

Suppose the proportional hazards assumption holds for the marginal distribution of each component event and a copula model is used to characterize the association. Under a working independence assumption [20], the limiting value for $\hat{\beta}_k$ is β_k . Here, we consider a generic scalar covariate Z_i , but note that the functions simplify with Z_i a binary treatment indicator because $Z_i^r = Z_i$ in this case. Let $S_k^{(r)}(t) = m^{-1} \sum_{i=1}^m \tilde{Y}_{ik}(t) \lambda_{k0}(t) \exp(\beta_k Z_i) Z_i^r$, $s_k^{(r)}(t) = E(S_k^{(r)}(t))$, $S_k^r(\beta_k, t) = m^{-1} \sum_{i=1}^m Y_{ik}(t) \exp(\beta_k Z_i) Z_i^r$, and $s_k^{(r)}(\beta_k, t) = E(S_k^{(r)}(\beta_k, t))$, $r = 0, 1, 2$, where $E(\cdot)$ denotes expectation with respect to the true distribution. Let $\mathcal{A}(\beta) = \text{diag}\{A_k(\beta_k)\}$, $k = 1, \dots, K$, where the k th diagonal element of $\mathcal{A}(\beta)$ is

$$A_k(\beta_k) = \int_0^\infty \left\{ \frac{s_k^{(2)}(\beta_k, t)}{s_k^{(0)}(\beta_k, t)} - \frac{s_k^{(1)}(\beta_k, t)^{\otimes 2}}{s_k^{(0)}(\beta_k, t)} \right\} s_k^{(0)}(t) dt,$$

by the Theorem 4.2 of Andersen and Gill [35]. In the present setting, the true model is known, the required expectations can be obtained in closed form, and the integral can be evaluated using numerical methods.

If

$$M_{ik}(t) = \bar{N}_{ik}(t) - \int \bar{Y}_{ik}(t)\lambda_{0k}(t)e^{\beta_k Z_i} dt$$

is the martingale for events of type k ; let

$$w_{ik}(\beta_k) = \int_0^\infty \left\{ Z_i - \frac{s_k^{(1)}(\beta_k, t)}{s_k^{(0)}(\beta_k, t)} \right\} dM_{ik}(t),$$

and $\mathbf{w}_i(\beta) = (w_{i1}(\beta_1), \dots, w_{iK}(\beta_K))'$. Then if we define $\mathcal{B}(\beta) = E(\mathbf{w}_i(\beta)\mathbf{w}_i(\beta)')$, the asymptotic robust covariance matrix $\Sigma(\beta)$ takes the form $\mathcal{A}(\beta)^{-1}\mathcal{B}(\beta)\mathcal{A}(\beta)^{-1}$ [20]. This can be used to obtain the limiting value through (12).

The entries of $\mathcal{B}(\beta)$ are obtained as follows. The (j, j) element of $\mathcal{B}(\beta)$ is

$$\begin{aligned} E(w_{ij}^2(\beta_j, t)) &= E\langle w_{ij}(\beta_j), w_{ij}(\beta_j) \rangle \\ &= E \left(\int_0^\infty \left\{ Z_i - \frac{s_j^{(1)}(\beta_j, t)}{s_j^{(0)}(\beta_j, t)} \right\}^2 s_j^{(0)}(t) dt \right) \\ &= \int_0^\infty E \left(\left\{ Z_i - \frac{s_j^{(1)}(\beta_j, t)}{s_j^{(0)}(\beta_j, t)} \right\}^2 s_j^{(0)}(t) \right) dt \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the predictable covariation process and the last equality holds because of Fubini's theorem [36]. The (j, k) element of $\mathcal{B}(\beta)$ is then

$$\begin{aligned} E(w_{ij}(\beta_j)w_{ik}(\beta_k)) &= E\langle w_{ij}(\beta_j), w_{ik}(\beta_k) \rangle \\ &= E \iint_0^\infty \left(Z_i - \frac{s_j^{(1)}(\beta_j, t)}{s_j^{(0)}(\beta_j, t)} \right) \left(Z_i - \frac{s_k^{(1)}(\beta_k, t)}{s_k^{(0)}(\beta_k, t)} \right) \langle dM_j(t_j), dM_k(t_k) \rangle. \end{aligned}$$

Using the covariance function for correlated martingales of Prentice and Cai [37], $\langle dM_j(t_j), dM_k(t_k) \rangle$ can be obtained. In the case of bivariate data, $\langle dM_1(t_1), dM_2(t_2) \rangle$ is obtained simply as

$$\begin{aligned} \langle dM_1(t_1), dM_2(t_2) \rangle &= \mathcal{F}(dt_1, dt_2|z_i; \Omega)dt_1dt_2 + \mathcal{F}(t_1, dt_2|z_i; \Omega)\Lambda_1(dt_1|z_i; \Omega)dt_1dt_2 \\ &\quad + \mathcal{F}(dt_1, t_2|z_i; \Omega)\Lambda_2(dt_2|z_i; \Omega)dt_1dt_2 \\ &\quad + \mathcal{F}(dt_1, dt_2|z_i; \Omega)\Lambda_1(dt_1)\Lambda_2(dt_2|z_i; \Omega)dt_1dt_2, \end{aligned}$$

where $\Lambda_k(dt_k|z_i; \Omega) = d\Lambda_k(t_k|z_i; \Omega)/dt_k$; $\mathcal{F}(dt_1, dt_2|z_i; \Omega) = \partial^2\mathcal{F}(t_1, t_2|z_i; \Omega)/\partial t_1\partial t_2$, $\mathcal{F}(dt_1, t_2|z_i; \Omega) = \partial\mathcal{F}(t_1, t_2|z_i; \Omega)/\partial t_1$, and $\mathcal{F}(t_1, dt_2|z_i; \Omega) = \partial\mathcal{F}(t_1, t_2|z_i; \Omega)/\partial t_2$. More specifically, if the joint survivor function $\mathcal{F}(t_1, t_2|z_i; \Omega)$ is specified by the Clayton copula with margins of two exponential distributions, then $\langle dM_j(t_j), dM_k(t_k) \rangle$ can be obtained in closed form and $E(w_{ij}(\beta_j)w_{ik}(\beta_k))$ can be obtained through numerical integration. Thus, we obtain the limiting value of robust variance, then the limiting weights can be calculated using $\mathbf{c}(\beta) = \Sigma^{-1}(\beta)/\mathbf{J}'\Sigma^{-1}\mathbf{J}$ and the limiting value $\bar{\beta}$ using equation (12).

Acknowledgements

Longyang Wu was supported by a graduate scholarship from the Natural Sciences and Engineering Research Council and an Ontario Graduate Scholarship. This research was also supported by grants from the Canadian Institutes for Health Research and the Natural Sciences and Engineering Research Council of Canada to Richard Cook. Richard Cook is a Canada Research Chair in Statistical Methods for Health Research.

References

1. POISE Study Group. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008; **371**:1839–1847.
2. Barnett HJM, Taylor DW, Eliasziw M, Fox AJ, Ferguson GG, Haynes RB, Rankin RN, Clagett GP, Hachinski VC, Sackett DL, Thorpe KE, Meldrum HE for the North American Symptomatic Carotid Endarterectomy Trial Collaborators. Benefit of carotid arterectomy in patients with symptomatic moderate or severe stenosis. *The New England Journal of Medicine* 1998; **339**(14):1415–1425.
3. Carlson RH. Prostate cancer: composite endpoint trips up satraplatin trial. *Oncology Times* 2007; **24**:44–47.
4. Hortobagyi GN, Theriault RL, Porter L, Blayney D, Lipton A, Sinoff C, Wheeler H, Simeone JF, Seaman J, Knight RD, Heffernan M, Reitsma DJ, Kennedy I, Allan SG, Mellars K for the Protocol 19 Aredia Breast Cancer Study Group. Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *The New England Journal of Medicine* 1996; **335**(24):1785–1792.
5. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 1995; **57**(1):125–133.
6. Sankoh AJ, D'Agostino Sr RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Statistics in Medicine* 2003; **22**(20):3133–3150.
7. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. *Controlled Clinical Trials* 2000; **21**(6):527–539.
8. Ferreira-González I, Permyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Composite endpoints in clinical trials: the trees and the forest. *Journal of Clinical Epidemiology* 2007; **60**(7):660–661.
9. Cannon CP. Clinical perspectives on the use of composite endpoints. *Controlled clinical trials* 1997; **18**:517–529.
10. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *Journal of the American Medical Association* 2003; **289**(19):2554–2559.
11. Montori VM, Permyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, Alons J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schünemann HJ, Jaeschke R, Guyatt GH. Validity of composite end points in clinical trials. *British Medical Journal* 2005; **330**(7491):594–596.
12. Chi GYH. Some issues with composite endpoints in clinical trials. *Fundamental & Clinical Pharmacology* 2005; **19**:609–619.
13. Ferreira-González I, Permyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *Journal of Clinical Epidemiology* 2007; **60**(7):651–657.
14. Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? *Annals of Oncology* 2010; **21**(12):2324–2332.
15. DeMets DL, Califf RM. Lessons learned from recent cardiovascular clinical trials: part I. *Circulation* 2002; **106**:746–751.
16. Neaton JD, Gray G, Zuckerman BD, Konstam M. Key issues in end point selection for heart failure trials: composite end points. *Journal of Cardiac Failure* 2005; **11**(8):567–575.
17. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. *Annals of Internal Medicine* 2008; **149**:612–617.
18. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schünemann HJ, Permyer-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. *British Medical Journal* 2007; **334**(7597):786–792.
19. Bethel MA, Holman R, Haffner SM, Califf RM, Huntsman-Labed A, Hua TA, McMurray J. Determining the most appropriate components for a composite clinical trial outcome. *American Heart Journal* 2008; **156**(4):633–640.
20. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of American Statistical Association* 1989; **84**:1065–1073.
21. Genest C, MacKay J. The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician* 1986; **40**(4):280–283.
22. Nelsen RB. *An Introduction to Copulas*. Springer: New York, 2006.
23. Joe H. *Multivariate Models and Dependence Concepts*. Chapman and Hall: London, 1997.
24. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**:141–151.
25. Genest C. Frank's family of bivariate distributions. *Biometrika* 1987; **74**:549–550.
26. Gumbel E. Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de statistique de l'Université de Paris* 1960; **9**:171–173.
27. Cox DR. Regression models with lifetables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.
28. Cox DR. Partial likelihood. *Biometrika* 1975; **62**:269–276.
29. Cox DR. Tests of separate family of hypotheses. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* 1961:105–123.
30. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; **30**:1–25.
31. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika* 1986; **73**:363–369.
32. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer: New York, 2000.

33. Jayaram L, Pizzichini MM, Cook RJ, Boulet L-P, Lemi re C, Pizzichini E, Cartier A, Hussack P, Goldsmith CH, Laviolette M, Parameswaran K, Hargreave FE. Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations. *European Journal of Respiratory* 2006; **27**:483–494.
34. Sheehe PR. Composite end points in clinical trials. *Journal of the American Medical Association* 2010; **303**:1698.
35. Andersen PK, Gill RD. Cox regression model for counting processes: a large sample study. *The Annals of Statistics* 1982; **10**:1100–1120.
36. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. John Wiley and Sons: New York, 1991.
37. Prentice RL, Cai J. Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* 1992; **79**(3):495–512.