On assessing surrogacy in a single trial setting using a semi-competing risks paradigm

Debashis Ghosh
Departments of Statistics and Public Health Sciences, The Pennsylvania State University, 514A Wartik Building, University Park, PA,16802 U.S.A., ghoshd@psu.edu

Summary

There has been a recent emphasis on the identification of biomarkers and other biologic measures that may be potentially used as surrogate endpoints in clinical trials. We focus on the setting of data from a single clinical trial. In this paper, we consider a framework in which the surrogate must occur before the true endpoint. This suggests viewing the surrogate and true endpoints as semi-competing risks data; this approach is new to the literature on surrogate endpoints and leads to an asymmetrical treatment of the surrogate and true endpoints. However, such a data structure also conceptually complicates many of the previously considered measures of surrogacy in the literature. We propose novel estimation and inferential procedures for the relative effect and adjusted association quantities proposed by Buyse and Molenberghs (1998, *Biometrics*, 1014 – 1029). The proposed methodology is illustrated with application to simulated data, as well as to data from a leukemia study.

Keywords

Bivariate survival data; Copula model; Dependent Censoring; Multivariate failure time data; Prentice criterion

1. Introduction

Recently, surrogate endpoints have become of great interest in clinical research (Biomarkers Working Group, 2001). These are measures that can be collected in a shorter time period and/ or using fewer subjects than those normally considered in a classical clinical trial (e.g., survival). Surrogate endpoints are typically proposed based on biological considerations with a certain progression model of disease. One example of a surrogate is CD4 count levels in AIDS; the CD4 count can potentially serve as a surrogate endpoint for death. Another example from cancer studies is using tumor shrinkage as a surrogate endpoint for survival or disease-free survival. Also, high-throughput fields such as genomics and proteomics are generating molecular profiles that scientists might wish to use as surrogate endpoints in the future.

A seminal paper in the evaluation of surrogate endpoints was the work of Prentice (1989), in which he specifies conditions under which a test of a treatment effect on the surrogate endpoint provides valid inference for testing for the effect of treatment on the true endpoint. Because the so-called Prentice criterion is a difficult one to validate, many authors have focused on alternative measures of assessing surrogacy; a comprehensive discussion of them can be found in the recent monograph by Burzykowski et al. (2005).

We focus on the situation where the two endpoints are both times to event. We consider the scenario when the surrogate endpoint occurs before the true endpoint. This complicates the analysis of such data so that we have to make use of techniques from a field of survival analysis known as semi-competing risks data (Fine et al. 2001; Ghosh, 2006) for assessment of

surrogacy. The application of semi-competing risks approaches to the surrogate endpoints problem is novel; in fact, in chapter 11 of Burzykowski et al. (2005), the authors write that methods that allow for asymmetrical dependence of surrogate and true endpoints are lacking. They discuss the possibility of using semi-competing risks data methods but do not explore it any further. We focus on the setting of a single clinical trial and discuss issues particular to it. While the meta-analysis paradigm has become very important to the assessment of surrogacy (Daniels and Hughes, 1997; Gail et al., 2000; Buyse et al., 2000), it is important to start with the single trial setting so as to clearly understand the issues involved. The structure of this paper is as follows. In Section 2, we review the data structures observed in the paper, discuss semi-competing risks data and their application to the analysis of surrogate endpoints. We then review several measures of surrogacy and describe probabilistic models in Section 3. For the analysis of the failure time endpoints, we use the linear regression model (Cox and Oakes, 1984, §5.2), whose use in the assessment of surrogacy has been recently advocated by Ghosh (2008). However, he did not deal with semi-competing risk issue there. Because of the nature of the data structure, both formulation and estimation of existing surrogacy measures becomes challenging. Estimation procedures for the surrogacy measures proposed by Buyse and Molenberghs (1998), along with attendant asymptotic results, are provided in Section 4. The finite-sample properties of the proposed methods are assessed with application to simulated and real studies in Section 5. We conclude with some discussion in Section 6.

2. Preliminaries and background

2.1. Data Structures and Semi-Competing Risks Rationale

We start by making the following definitions. Let $a \wedge b$ denote the minimum of two numbers a and b. Define $I(A)$ to be the indicator function for the event A. Let S denote time to the surrogate endpoint, T the time to the true endpoint and C time to independent censoring. Let Z be the indicator for treatment group (0 = control; 1 = treatment). Assume that (S, T) and C are conditionally independent given Z. We observe n observations $(X_i, \delta_i^X, Y_i, \delta_i^Y, Z_i)$ ($i = 1, \ldots, n$), a random sample from $(X, \delta^X, Y, \delta^Y; Z)$, where $X = S \wedge T \wedge C$, $\delta^X = I(S \leq T \wedge C)$, $Y = T \wedge C$ and $\delta^Y = I(T \leq C)$.

Note that we have censored S by the minimum of T and C and not just by C. Within the semi-competing risks setup, we are interested in the following distributions:

1. The distribution of the true endpoint in the absence of independent censoring, potentially adjusting for covariates;

2. The distribution of the surrogate endpoint in the absence of **both** the true endpoint and independent censoring, potentially adjusting for covariates.

Because we wish to adjust for the dependent censoring in (2) by the true endpoint, this will complicate the analysis.

In the analysis of surrogate endpoints, there has been interest in measures of dependence between S and T, adjusting for Z. This has typically been formulated using copula models (Nelsen, 1999; Burzykowski et al., 2001). One point of note is that these measures are typically symmetric in S and T (Burzykowski et al. 2005, p. 194). One could argue that a more reasonable strategy is to consider asymmetric measures of dependence between S and T. This highlights the true endpoint as being the "gold standard" and treats the surrogate as a proxy that should in principle be observed more quickly than the true endpoint. Under this type of scenario, it seems reasonable to entertain probabilistic models for S and T on the wedge. The wedge refers to the region where $S \leq T$. This has not been done in the previous literature on surrogate endpoints. It is important to note that we do not assume that $P(S \leq T) = 1$; we instead seek to

make inferential statements about surrogacy on the wedge. This is the same view adopted by Fine et al. (2001) and Peng and Fine (2006).

We also note in passing that an alternative practical approach is to create a composite endpoint that represents the minimum of *S* and *T*. One example of such an endpoint is disease-free survival, defined as the minimum of time to disease recurrence and time to death. While this has recently been used as a surrogate endpoint in colorectal cancer (Sargent et al., 2005), there are several issues with the use of composite endpoints. First, there are difficulties in interpreting the composite endpoint in terms of the individual events themselves. Second, if the goal is to study the dependence between *S* and *T*, then it is impossible to do so if the composite endpoint is used.

3. Proposed Methodology for Assessing Surrogacy

3.1. Measures of Surrogacy

For assessing surrogacy, several authors have pointed out the problems with the use of the Prentice criteria in a purely testing format (Begg and Leung, 2000; Berger, 2004). Instead, much of the surrogate endpoint literature has been devoted to developing estimation and inference procedures for alternative measures of surrogacy.

Before defining the measures, it is important to formulate appropriate probabilistic models for the true and surrogate endpoints as functions of *Z*. In this paper, we formulate models based on the accelerated failure time model (Cox and Oakes, 1984, §5.2). For most existing surrogacy measures, we would need to fit the following regression models:

$$\log T = \beta Z + \varepsilon_1, \tag{1}$$

$$\log S = \alpha Z + \varepsilon_2, \tag{2}$$

and

$$\log T = \zeta \log S + \gamma Z + \varepsilon_3, \tag{3}$$

where ε_1; ε_2 and ε_3 are error terms and β, α and (ζ, γ) are unknown regression coefficients to be estimated. Note that for the Prentice (1989) criteria to be true, we need β and α to not equal zero but for $\gamma = 0$, in addition to $\log T$ and $\log S$ being correlated.

We use the AFT model rather than the proportional hazards model (Cox, 1972) in this manuscript. By assuming linear models (1) – (3), it is possible for all these models to be simultaneously compatible. By contrast, this does not happen with the proportional hazards model (Lin et al., 1997). In addition, the interpretation of the AFT model is akin to that of linear regression and is simpler than that for a hazard function; this has been discussed by Cox, among other authors (Reid, 1994, p. 450).

The first measure of surrogacy considered in the literature is proportion of treatment effect (PTE) (Friedman et al., 1992). PTE is defined as

$$\text{PTE} = \frac{\beta - \gamma}{\beta}. \tag{4}$$

To estimate PTE, we would need to fit models (1) and (3), obtain estimators $\hat{\beta}$ and $\hat{\gamma}$ from these models and plug them into (4).

Wang and Taylor (2002) and Ghosh (2008) explored the properties of a PTE-type quantity termed the F measure. The idea is to compare the distribution of the surrogate endpoints between the two treatment groups in which the distributions are weighted by the conditional distribution of the true endpoint, given treatment and the surrogate endpoint. An F value of zero corresponds to the surrogate endpoint being useless, while that of one implies it is perfect. Wang and Taylor (2002) show that their F measure reduces to the PTE in the case of uncensored data for a bivariate normal distribution conditional on covariates. Ghosh (2008) showed how to calculate a version of the F measure with bivariate censored data.

Buyse and Molenberghs (1998) argued for two alternative measures of surrogacy, the relative effect and the adjusted association. Their argument was based on the premise that surrogates are meaningful quantities to consider if the effect of treatment on the surrogate endpoint can be used to predict treatment effects on the true endpoint. The relative effect (RE) is defined, using the notation of models (1) and (2) as

$$\text{RE} = \frac{\beta}{\alpha}. \tag{5}$$

If the variance of log T and log S are the same, then a RE value of one corresponds to the surrogate being useful, while that of zero corresponds to the surrogate being useless. Note that with the semi-competing risk data structure, log $S \le$ log T, so the variances of the two random variables will generally not be the same.

The adjusted association (AA) is the correlation between the true and surrogate endpoint, adjusting for the treatment effect. Calculation of the adjusted association requires estimations of models (1) and (2), where (ε_1; ε_2) have a joint distribution parametrized by an association parameter. The RE and AA measures fit into the framework of Buyse and Molenberghs (1998) in which the treatment effect on the surrogate endpoint is used to predict the treatment effect on the surrogate endpoint. If the RE measure will be equal to one, then the surrogate is a perfect population-level surrogate (Buyse et al., 2000). By contrast, what the adjusted association measures is the correlation between the two endpoints at the individual level.

In the literature on semi-competing risks data, a popular class of models for modelling dependence is copula models on the wedge, which we describe in the next section.

3.2. Copula models on the wedge

Let $H(s, t) \equiv Pr(S > s, T > t)$ denote the joint survival distribution of (S, T). Then a copula model decomposes H into the marginal survival distributions:

$$H(s, t) = C_\theta\{F_S(s), F_T(t)\}, \tag{6}$$

where C is the copula function, θ is a dependence parameter, and F_S and F_T are the marginal survivor functions of S and T. The point of model (6) is to decouple the joint distribution of (S, T) into univariate components. In principle, the estimation of the marginal distributions can be decoupled from that of the dependence between S and T. While that does not happen here, we still think the copula formulation a useful one for formulating the appropriate probabilistic model.

One popular example of a copula is the Clayton-Oakes model (Clayton, 1978; Oakes, 1986). It is defined as follows. Let the cross-ratio function (Oakes, 1989) between S and T be defined by

$$\theta(t, s) = \frac{\lambda_T(s \mid S = t)}{\lambda_T(s \mid S \geq t)},$$

(7)

where

$$\lambda_T(t|A) = \lim_{\Delta t \to 0} \frac{d}{dt} \Pr(T < t + \Delta t | T \geq t, s \in A),$$

and A is a subset of the interval $(0,\infty)$. The Clayton-Oakes model makes the assumption that $\theta(s, t)$ in (7) is constant for all s and t, i. e. $\theta(t, s) = \theta$. In terms of model (6), the Clayton-Oakes model corresponds to $C_\theta(u, \upsilon) \equiv \{u^{-\theta} + \upsilon^{-\theta} - 1\}^{-1/\theta}$ with $\theta > 0$.

In the semi-competing risks literature, several authors have postulated Clayton-Oakes model only for the upper wedge of the joint distribution of (S, T). Day et al. (1997) considered this model and proposed a test of the independence of S and T. Fine et al. (2001) provided a closed form estimator of θ using modified weighted concordance estimating functions from Oakes (1982, 1986) along with an asymptotic variance estimator. Wang (2003) proposed an estimation procedure in this model, applicable more generally to copula models, based on a Doob-Meyer decomposition of appropriate counting processes.

For our setting, we wish to have models for the joint distribution of (S, T) on the wedge. By defining the model in this manner, several facts obtain. First, goodness of fit of the model can be assessed because the joint distribution of S and T when $S \leq T$ is identifiable from the observed data. This is proved in Fine et al. (2001). Second, even though this feature of the joint distribution is identifiable on the wedge, the marginal distribution of S is still not easily estimable using the observed data. In fact, while $H(s, 0)$ is typically the marginal survivor function for S in the unconstrained case, this will not be the case in the situation where $S \leq T$. Finally, because we have imposed the constraint that $S \leq T$, they are by definition dependent in this model. Even though Day et al. (1997) and Fine et al. (2001) propose tests for "independence" for the Clayton-Oakes model on the wedge, what they really refer to is a test for quasi-independence, in the sense of Tsai (1990). Conceptually, quasi-independence means that on all squares contained in the region $S \leq T$, S and T are independent.

In the surrogate endpoints problem, it does seem natural to consider models on the wedge, as it is necessary for S and T to be correlated in order for S to serve as a useful surrogate endpoint. Typically, there is strong scientific and empirical evidence to show that S and T are correlated. If this is not the case, then it would be make absolutely no sense to be considering S as a surrogate for T. It is important to note that this reasoning is applicable to a single trial. In the setting where data from multiple trials are considered, it is possible for the individual-level correlation between S and T to be weak but for the trial-level correlation to be strong (Burzykowski et al., 2005, Ch. 7). This would require a meta-analytical treatment of the AA and RE measures and is beyond the scope of our manuscript.

3.3 Limitation of PTE-type measures of surrogacy with semi-competing risks data

In this section, we discuss conceptual limitations to using the PTE and F measures of surrogacy with semi-competing risks data. With such a data structure, we noted earlier that $P(S > s, T > t/Z)$ is identifiable from the observed data given $s \leq t$. What is not nonparametrically estimable

is the conditional distribution of *T* given *S* and *Z*. This is because as noted by Fine et al. (2001), the marginal distribution of *S* given *Z* is not identifiable. Using [*A*] to denote the distribution of a random variable *A*, we have

$$
\begin{aligned}
[T|S, Z] &= \frac{[T, S, Z]}{[S, Z]} \\
&= \frac{[T, S \mid Z][Z]}{[S \mid Z][Z]} \\
&= \frac{[T, S \mid Z]}{[S \mid Z]}.
\end{aligned}
$$

While the conditional distribution of (*T, S*) conditional on covariates can be estimated non-parametrically on the wedge, [*S*/*Z*] cannot. Thus, in particular, the PTE and F measures with semi-competing risks data suffer from an inherent identifiability identifiability issue.

What is needed for estimation of [*T*/*S,Z*] is a model on the lower wedge (i.e., the region where *S > T*). If such a model is available, then the distribution of *S*/*Z* is estimable using the observed data. In the work of Fine et al. (2001), they propose using the gamma frailty model on the entire (*S, T*) in order to estimate the marginal distribution of *S*. However, one cannot assess goodness of fit for the model on the lower wedge using the observed data. Thus, derivation of the distribution of *S* given *Z* requires making unverifiable assumptions with semi-competing risks data. This result is analogous to that of Tsiatis (1975) for ordinary competing risks.

These arguments give further evidence against the use of the PTE and F measures for assessing surrogacy with semi-competing risks data. In the next section, we will instead focus on estimating the measures of Buyse and Molenberghs (1998).

3.4 Limitation of causal inference surrogacy with semi-competing risks data

Recently, there has been work exploring surrogate endpoints in the setting of models from causal inference (Frangakis and Rubin, 2002; Taylor et al., 2005). The focus is on estimating causal effects of the treatment within a certain subpopulation of the study subjects. It turns out that identification of causal effects and the feasibility of causal inference depends on whether or not potential outcomes can be defined on the region *S > T*. Suppose that *T* is time to death (survival). It has been noted by Robins (1995) that counterfactual outcomes for *S* when *S > T* are not well-defined at all. Zhang and Rubin (2003) refer to this as "truncation by death." Given this fact, one cannot identify subpopulations with the same potential outcomes for (*S, T*), which leads to the inability to define causal effects. Thus, all of the surrogacy measures that are nonparametrically identifiable using the observed data will have no causal interpretation if *T* is time to death. On the other hand, if *T* is time to disease recurrence and if *S* is time to biomarker positivity, then it is possible to define causal effects, as one can define the potential outcomes for (*S, T*) under the treatment groups for all individuals. While we do not pursue methods for causal inference here, we simply wish to point out that the interpretation of *S* and *T* on the lower wedge crucially affects the possibility of causal inference.

4. Estimation Procedures and Asymptotic Results

We will now develop methods of estimation and inference for the RE and AA quantities using semi-competing risks data. While these have been explored in the case of bivariate survival data by Burzykowski et al. (2001) and Ghosh (2008) for marginal proportional hazards and marginal AFT models, neither author explicitly dealt with semi-competing risks data.

4.1 Relative Effect

Because of the semi-competing risks structure to the surrogate and true endpoints, we are unable to decouple the estimation of the marginal regression model (2) with the dependence estimation. This complicates estimation of the RE and AA measures.

We first consider estimating the RE measure. While the estimation of the treatment on the true endpoint can be assessed using standard rank-based estimating function approaches, estimating the treatment effect on the surrogate endpoint is more complicated. For the estimate of the regression coefficient in (1), we use the following log-rank estimating function:

$$U_1(\beta) = \sum_{i=1}^{n} \delta_i^Y \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{Y}_j(\beta) \geq \tilde{Y}_i(\beta)\} Z_j}{\sum_{j=1}^{n} I\{\tilde{Y}_j(\beta) \geq \tilde{Y}_i(\beta)\}} \right],$$

(8)

where $\tilde{Y}_i(\beta) = Y_i \exp(-\beta' Z_i)$, $i = 1, \ldots, n$. Let $\hat{\beta}$ be a zero-crossing of $U_1(\beta)$. This is the estimating function proposed by Louis (1981).

For the estimation of α, we use the artificial censoring approach of Lin et al. (1996). The intuitive idea is to artificially trim the transformed surrogate endpoint time by a factor that allows for valid comparison between the two treatment groups. Define $\eta = (\alpha, \beta)$. This leads to the following estimating equation for estimation of η:

$$U_2(\eta) = \sum_{i=1}^{n} \tilde{\delta}_i^X(\eta) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j(\eta) \geq \tilde{X}_i(\eta)\} Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j(\eta) \geq \tilde{X}_j(\eta)\}} \right],$$

(9)

where $\tilde{X}_i(\eta) = \{S_i \exp(-\alpha Z_i) \wedge T_i \exp(-\beta Z_i - d) \wedge C_i \exp(-\beta Z_i - d)\}$,

$\tilde{\delta}_i^X(\eta) = I\{S_i \exp(-\alpha Z_i) \leq T_i \exp(-\beta Z_i - d) \wedge C_i \exp(-\beta Z_i - d)\}$

and $d = 0$ if $\alpha \leq \beta$ and $\beta - \alpha$ otherwise. Let $\hat{\alpha}$ be a zero-crossing of α from setting $U_2(\tilde{\eta}) = 0$, where $\tilde{\eta} = (\alpha, \hat{\beta})$. Once we have an estimate of β and α, the relative effect estimate is given by $\widehat{RE} = \widehat{\beta} / \widehat{\alpha}$. In the Appendix, we prove the consistency and asymptotic normality of $(\hat{\alpha}, \hat{\beta})$ and $\widehat{RE}$

Suppose we do not do the artificial censoring. An alternative approach would be to construct an estimating function for α similar to (8) based on the data (X_i, δ_i^X, Z_i), $i = 1, \ldots, n$. However, under model (2), it can be shown that the expected value of the estimating function at α_0 is not zero, which implies that estimates from this approach would be biased. This is why the artificial censoring in (9) is needed. Note that in (9), $\delta_i^X (i = 1, \ldots, n)$ is modified and converted to $\tilde{\delta}_i^X$. It is the case that for one of the treatment arms, $\delta^X = 1$ but $\tilde{\delta}^X = 0$; this is what is referred to as an artificial censoring event. The treatment group that is artificially censored depends on the relative magnitude of α and β If $\alpha \leq \beta$, then subjects in the $Z = 0$ are artificially censored, while subjects on treatment are potentially artificially censored if $\alpha > \beta$.

In more recent work, Peng and Fine (2006) have developed a pairwise analog of the artificial censoring approach of Lin et al. (1996). Their approach led to greater efficiency gains in the

simulation settings considered. Since the only covariate used in the regression models here is Z, the two methods are equivalent.

Note that the proposed estimating functions proposed are discontinuous functions of the regression coefficients, so this makes their standard error estimation difficult. We instead use a resampling-based method developed by Parzen et al. (1994) to derive confidence intervals for RE. We construct the pair of equations

$$U_1(\beta) = \sum_{i=1}^{n} \widehat{\psi}_{1i} G_i,$$

(10)

$$U_2(\eta) = \sum_{i=1}^{n} \widehat{\psi}_{2i} G_i,$$

(11)

where $(G_1, \ldots, G_n)$ are independent standard normal random variables,

$$\widehat{\psi}_{1i} = \delta_i^Y \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{Y}_j(\widehat{\beta}) \geq \tilde{Y}_i(\widehat{\beta})\} Z_j}{\sum_{j=1}^{n} I\{\tilde{Y}_j(\widehat{\beta}) \geq \tilde{Y}_i(\widehat{\beta})\}} \right] - d\widehat{\Lambda}_0^T(\tilde{Y}_i(\widehat{\beta})),$$

and

$$\widehat{\psi}_{2i} = \tilde{\delta}_i^X(\eta) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j(\widehat{\eta}) \geq \tilde{X}_i(\widehat{\eta})\} Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j(\widehat{\eta}) \geq \tilde{X}_i(\widehat{\eta})\}} \right] - d\widehat{\Lambda}_0^S(\tilde{X}_i(\widehat{\eta}))$$

and $\widehat{\Lambda}_0^T$ and $\widehat{\Lambda}_0^S$ are Nelson-Aalen type estimators of the net hazard and cause-specific hazard functions for the true and surrogate endpoints on the time-transformed scales. Denote the solution to equations (10) and (11) by $\eta^* = (\beta^*, \alpha^*)'$. Note that the equations (10) and (11) are constructed conditional on the observed data and that the only stochastic component to them is $(G_1, \ldots, G_n)$.

Define $\mathbf{U}(\eta) = \{U_1(\beta), U_2(\eta)\}$. Obviously, $\mathbf{U}(\eta^*) = \sum_{i=1}^{n} \widehat{\psi}_i G_i$, where $\hat{\psi}_i = (\psi\,\widehat{}_{1i}, \psi\,\widehat{}_{2i})'$. It then follows from (16) that

$$n^{1/2}(\eta^* - \eta) = \mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^{n} \widehat{\psi}_i G_i + o_P(1).$$

(12)

The conditional distribution of $n^{-1/2} \sum_{i=1}^{n} \widehat{\psi}_i G_i$ $(i = 1, \ldots, n)$, given the data, is zero-mean normal with covariance matrix $n^{-1} \sum_{i=1}^{n} \widehat{\psi}_i \widehat{\psi}_i'$, which converges in probability to $\mathbf{V}$. Therefore, the conditional distribution of $n^{1/2}(\eta^* - \eta)$ is asymptotically equivalent to the unconditional distribution of $n^{1/2}(\hat{\eta} - \eta)$. To approximate the distribution of η, we obtain a large number of realizations of η^* by repeatedly generating the random samples $(G_1, \ldots, G_n)$ while fixing the data and solving equations (10) and (11). The empirical distribution of η^* can then be used to

perform hypothesis testing or interval estimation on α and β. For the purposes of surrogacy, the empirical distribution of $\boldsymbol{\eta}^*$ yields an empirical distribution of the relative effect measure, which can then be used to construct a 95% CI for RE.

4.2 Adjusted Association

We now consider estimation of the adjusted association measure for surrogacy. We consider use of the Clayton-Oakes copula model in this paper. Because the covariate Z is typically discrete, this leads to a very simple two-step algorithm for estimation of the AA measure. First, within each level of Z, estimate the dependence parameter using the approach of Fine et al. (2001). This leads to an estimator $\hat{\theta}_Z$ defined by

$$\widehat{\theta}_Z = \frac{\sum_{i<j} D_{ijZ} \Delta_{ijZ}}{\sum_{i<j} D_{ijZ}(1 - \Delta_{ijZ})},$$

(13)

where $D_{ijZ} = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij}, Z_i = Z_j = Z)$ and $\Delta_{ijZ} = I\{(X_i - X_j)(Y_i - Y_j) > 0, Z_i = Z_j = Z\}$, $\tilde{C}_{ij} = C_i \wedge C_j, \tilde{X}_{ij} = X_i \wedge X_j, \tilde{Y}_{ij} = Y_i \wedge Y_j, i, j = 1, \ldots, n$. The advantage of this type of estimator is that it does not require estimation of high-dimensional nuisance parameters or the marginal distribution of S. Applying arguments similar to those in Fine et al. (2001), we can show that $n^{1/2}(\hat{\theta}_Z - \theta_Z)$ converges in distribution to a normal random variable with mean zero and variance $\sigma_Z^2 \equiv I(\theta_Z)/J(\theta_Z)^2$, where

$$I(\theta_Z) = \lim_{n \to \infty} 2n^{-3} \sum_{k<l<m} Q_{kl}(\theta_Z)Q_{km}(\theta_Z) + Q_{kl}(\theta_Z)Q_{lm}(\theta_Z) + Q_{km}(\theta_Z)Q_{lm}(\theta_Z),$$

$$Q_{ij}(\theta_Z) = D_{ijZ}\{\Delta_{ijZ} - \theta_Z/(1 + \theta_Z)\},$$

and

$$J(\theta_Z) = \lim_{n \to \infty} n^{-2} \sum_{i<j} D_{ijZ}(1 + \theta_Z)^{-2}.$$

Simple and consistent estimators of σ_Z^2 are given by $\widehat{\sigma}_Z^2 \equiv I(\widehat{\theta}_Z)/J(\widehat{\theta}_Z)^2$ for $Z = 0$ and $Z = 1$.

At the second step, given the estimates $\hat{\theta}_Z$, we construct a weighted average of the estimates $\widehat{\theta} \equiv \sum_{Z=0}^{1} w_z \widehat{\theta}_z$ this serves as our measure of the AA parameter. It is straightforward to see that $n^{1/2}(\hat{\theta} - \theta)$ converges in distribution to a mean-zero normal random variable with variance $w_0^2 \sigma_0^2 + w_1^2 \sigma_1^2$. Using this variance, we are able to derive the optimal weights for combining estimators as $w_z^{\text{opt}} = \sigma_{1-z}^2 (\sigma_0^2 + \sigma_1^2)^{-1}$. Again, this is easy to estimate using the observed data. The main appeal of the estimation procedure proposed here is ease of calculation. One could also adapt the estimators proposed by Wang (2003) to this situation, but the computation of the estimator becomes more cumbersome.

In theory, it is possible to consider estimators of θ_Z using weighted versions of (13). However, both Fine et al. (2001) and Ghosh (2006) have found that unweighted versions of these dependence estimators tend to perform quite competitively relative to weighted estimators, so we focus on unweighted estimates of θ_Z in this manuscript. Note that a crucial assumption is that there is no interaction between the dependence parameter in the Clayton-Oakes model and

treatment group. If there is such an interaction, then use of the adjusted association measure would be statistically invalid and lead to erroneous conclusions. We instead would wish to report the estimates $\hat{\theta}_0$ and $\hat{\theta}_1$, along with their associated standard errors.

5. Numerical Examples

5.1 Acute Myelogeneous Leukemia Data

The motivating data example comes from a Phase III clinical trial (Berman et al., 1991) comparing two treatments for acute myelogenous leukemia. The first treatment is the standard, daunorubicin (DNR) and cytosine arabinoside (Ara-C), while the experimental therapy was idarubicin (IDR) and Ara-C, all three of which are chemotherapeutic agents. The surrogate endpoint of interest here is time to complete remission, defined as the leukemia cells being killed by the chemotherapies. There are 130 patients in this randomized study; 65 are in each treatment arm. In this example, the standard therapy was the DNR/Ara-C, so it was coded as $Z = 0$, while $Z = 1$ corresponds to the IDR/Ara-C combination.

While IDR and DNR are have similar functions, there are slight differences in the molecular structure of the two compounds that lead clinicians to believe that the IDR might be less toxic than DNR. The goal of the study is to determine if compare the efficacy of the DNR/Ara-C combination versus that of IDR/Ara-C with respect to several study endpoints, among them complete remission and survival, as discussed by Berman et al. (1991). Complete remission refers to leukemia cells being killed by the therapy. Here and in the sequel, we will use the terms complete remission and remission interchangeably. Detailed descriptives of the surrogate and true endpoints can be found in Ghosh (2008).

In a prior analysis of these data using AFT models, Ghosh (2008) obtained estimated effects of −0.69 for treatment on remission (standard error = 0.34) and 0.51 for treatment on survival (standard error = 0.24). Thus, we find that patients on IDR/Ara-C tend to have longer survival and shorter times to remission than those on DNR/Ara-C. Combining the regression parameter estimates yields an estimated relative effect of 0.74 in magnitude; a 95% nonparametric bootstrap gave a confidence interval of (0.40, 1.15) for the RE.

The analysis done by Ghosh (2008) treated death as a censoring event for the surrogate endpoint. Doing so models the cause-specific hazard of remission (Kalbeisch and Prentice 2002, §8.2.2), which is a fundamentally different quantity than the net hazard of remission, which is what gets modelled here. Applying the proposed methodology, we get an estimated treatment effect on survival of 0.59 (standard error = 0:28) and −2.00 for treatment on remission (standard error = 0.41). This leads to a relative effect estimate of 0.30 in magnitude; the proposed resampling method yields a 95% CI of (0.11, 0.49). In terms of the parameter estimates, the reason for the difference in the treatment effect on survival is because in Ghosh (2008), the method of Jin et al. (2003) was used, while here, the method of Louis (1981) was used. This leads to the slight difference in parameter estimates and standard errors. For the effect of treatment on complete remission, we are modelling a much different quantity for remission than in Ghosh (2008). The standard error increases for the proposed methodology because of the artificial censoring. 18 of the 89 complete remission events have been artificially censored here; this is needed so as to obtain an approximately unbiased estimate of the effect of treatment on time to complete remission. A key finding is that we have less evidence for surrogacy based on using the relative effect measure.

Next, we describe the estimation and inference for adjusted association. In Ghosh (2008), we obtained an adjusted association parameter of 0.61 along with an associated 95% CI of (0.47, 0.83). Thus, after adjusting for treatment, there is a significant negative association between time to complete remission and time to death. This makes biological sense; if the cancer cells

die sooner, then this would correlate with longer survival. Applying the proposed methodology in Section 4.2, we obtain an estimate of the adjusted association parameter of 0.42 and associated 95% confidence interval of (0.34, 0.50). This is even more evidence of negative association between the two endpoints than in the original analysis of Ghosh (2008), as the AA estimate deviates more from one. This suggests greater evidence of surrogacy based on this criteria. One other point to note is that the confidence intervals for the adjusted association and relative effect measures become narrower.

Applying the proposed methodology gives conflicting evidence of surrogacy in this example. A potential way to resolve this would be to have data from multiple trials testing these two treatments. For that scenario, the AA measure would assess individual-level surrogacy, while the RE would be a measure of trial-level surrogacy.

5.2 Simulation Study

We also performed a simulation study to assess the finite-sample properties of the RE and AA estimators from Section 4. For the RE simulation, data are generated using marginally standard normal distributions for ($logT, logS$) conditional on Z with means μ_{TZ} and μ_{SZ}. Since we are mimicking a clinical trial situation, we assume that $P(Z = 1) = 0.5$ for the simulation. Dependence between $\log S$ and $\log T$ was generated by a Clayton-Oakes model with dependence parameter θ common to the treatment and control groups.

In this model, the population relative effect is given by $RE = (\mu_{T1} - \mu_{T0})/(\mu_{S1} - \mu_{S0})$. We considered ($\mu_{S0}, \mu_{S1}, \mu_{T0}, \mu_{T1}$) = (0, 1, 0, 0.8) and (0, 1, 0, 0.4). These configurations yield RE values equal to 0.8 and 0.4. The adjusted association is given by θ, which we took to equal 1.1 and 1.6. For each simulation setting, 1000 simulation samples were generated and 1000 resamplings were generated for each simulation sample. As noted by Fine et al. (2001, §4), there are issues in attempting to generate data from a proper probabilistic model on the region $S \leq T$. We generate (S, T) conditional on Z using the above model and create the observed data structure $(X_i, \delta_i^X, Y_i, \delta_i^Y, Z_i)$, $i = 1,\ldots,n$. We seek to estimate RE and AA using the data observed on the region $S \leq T$. In particular, this model implies that $P(S < T/Z = 0) = 0.5$ for both scenarios and $P(S < T/Z = 1) \approx 0.33$ and 0.44 for the two scenarios (i.e., RE = 0.4 and = 0.8, respectively). Censoring (C), distributed independently of (S,T), was generated using the logarithm of a Uniform[0,3] random variable. This leads to approximately 60% and 40% censoring for (S,T) for the RE = 0.4 and RE = 0.8 situations. We also considered the naive approach to estimation of relative effect, which was described in Section 4.

The results are presented in Table 1. We find that the proposed estimators perform satisfactorily in finite samples. There is a small bias in the relative effect estimator that diminishes with larger sample size. By contrast, the naive approach is biased for all scenarios considered and yields poor coverage for the 95% confidence intervals. The resampling-based method tends to provide slightly conservative coverage for the 95% confidence intervals.

We also assessed the finite-sample properties of the proposed methodology for AA. Conditional on Z, we generated data from a bivariate Clayton-Oakes model under two scenarios. In the first, $\theta_1 = \theta_2 = 1.6$, while in the second, $\theta_1 = 2$ and $\theta_2 = 4$. The second situation is to consider the performance of the proposed methodology under model misspecification. The marginal distributions of T and S were generated as for the data considered in Table 1. The results are presented in Table 2. We note that despite in the case where the simulation scenario is correctly specified, the proposed methodology has good small-sample properties. In the situation where the model is misspecified because of the interaction between the dependence parameter with treatment, there is a bias that appears to persist with larger sample sizes.

6. Discussion

In this paper, we have described a semi-competing risks approach to the analysis of surrogate endpoints. Using this formulation clarifies inherent asymmetry of the surrogate and true endpoints. This is in keeping with the spirit of the surrogate endpoint serving to potentially replace the true endpoint in clinical practice. By incorporating the constraint that $S \leq T$, there is also a narrowing of the confidence intervals in the measures of surrogacy, leading to a gain in confidence about this assessment. While we have focused on existing measures of surrogacy, many limitations and complications were discovered, primarily in the case of quantities such as the PTE. New measures that take into account the data structure may be necessary.

One important issue is appropriate selection of an appropriate copula model. While such methods have been recently proposed (e.g., Wang and Wells, 2000), we have focused on the Clayton-Oakes model primarily because of its widespread use in the semi-competing risks literature and because it leads to very simple estimation procedures that do not involve estimation of nuisance parameters. Obviously, model assessment would be important to perform in practice for the data analyst.

As was alluded to in the paper, it would be important to also consider the meta-analysis setting for the purposes of assessing surrogacy. A natural approach would be to consider a hierarchical model that would incorporate both within- and between-trial information. While Daniels and Hughes (1997), Gail et al. (2000) and Buyse et al. (2000) have described methods for assessing surrogacy in the meta-analysis setting, they did not extend their methodologies to accommodate semi-competing risks data. A natural extension to models (1) and (2) in the meta-analysis problem is

$$
\begin{aligned}
\log T &= (\beta + b_i)Z + \varepsilon_1^i \\
\log S &= (\alpha + a_i)Z + \varepsilon_2^i,
\end{aligned}
$$

where i indexes study and (a_i, b_i) are random effects for the ith study, assumed to be drawn from some distribution. Research on this topic is currently under investigation.

Acknowledgments

Appendix

Appendix

Asymptotic results for $(\hat{\alpha}, \hat{\beta})$ and $\widehat{RE}$: Define $N_{1i}(t;\beta) = \delta_i^Y I\{\tilde{Y}_i(\beta) \leq t\}$ and

$$
M_{1i}(t;\beta) = N_{1i}(t;\beta) - \int_0^t I\{\tilde{Y}_i(\beta) \geq s\}\lambda_0^T(s)\mathrm{d}s,
$$

where $\lambda_0^T(t)$ is the hazard function of T when $Z = 0$. Then

$$U_1(\beta) = \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i - \overline{Z}^{(1)}(t;\beta) \right\} dN_{1i}(t;\beta)$$
$$= \sum_{i=1}^{n} \int_0^\infty \{ Z_i - \overline{Z}^{(1)}(t;\beta) \} dM_{1i}(t;\beta),$$

Where $\overline{Z}^{(1)}(t;\beta) = \sum_{j=1}^{n} I\{\tilde{Y}_j(\beta) \geq t\} Z_j / \sum_{j=1}^{n} I\{\tilde{Y}_j(\beta) \geq t\}$. Because $M_{1i}(t) \equiv M_{1i}(t, \beta_0)$ ($i = 1, \ldots, n$) are martingales with respect to the marginal filtration $\mathcal{F}_t = \sigma\{N_{1i}(s, \beta_0), I\{\tilde{Y}_i(\beta_0) \geq s\}, Z_i; 0 \leq s \leq t, i = 1, \ldots, n\}$, the martingale central limit theorem (Fleming and Harrington, 1991, Theorem 5.3.5) implies that $n^{-1/2} U_1(\beta_0)$ is asymptotically zero-mean normal and that

$$n^{-1/2} U_1(\beta_0) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \{ Z_i - \bar{z}^{(1)}(t) \} dM_{1i}(t) + o_P(1),$$

(14)

where $z^{-(1)}(t)$ is the probability limit of $Z^{-(1)}(t; \beta_0)$. In (14), $o_P(1)$ denotes a random variable that converges in probability to zero.

Similarly, define

$$M_{2i}(t;\eta) \equiv M_{2i}(t;\eta) = \tilde{\delta}_i^X(\eta) I\{\tilde{X}_i(\eta) \leq t\} - \int_0^t I\{\tilde{X}_i(\eta) \geq u\} \lambda_0^S(u),$$

where λ_0^S it the cause-specific hazard function for S when $Z = 0$. By applying similar martingale arguments as above, we can establish that

$$n^{-1/2} U_2(\eta_0) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty \{ Z_i - \bar{z}^{(2)}(t) \} dM_{2i}(t) + o_P(1),$$

(15)

where $z^{-(2)}(t)$ is the probability limit of $Z^{-(2)}(t; \eta_0)$.

Write $\mathbf{U}(\eta) = \{U_1(\beta), U_2(\alpha; \beta)\}'$. Because both the right-hand sides of (14) and (15) consist of sums of n independent random variables, the multivariate central limit theorem implies that $n^{-1/2} \mathbf{U}(\eta_0)$ is asymptotically bivariate normal with mean zero and covariance matrix

$$\mathbf{V} \equiv \lim_{n \to \infty} n^{-1} \sum_{i-1}^{n} E \begin{bmatrix} \psi_{1i}^2 & \psi_{1i}\psi_{2i} \\ \psi_{1i}\psi_{2i} & \psi_{2i}^2 \end{bmatrix},$$

where $\psi_{1i} = \int_0^\infty \{ Z_i - \bar{z}^{(1)}(t) \} dM_{1i}(t)$ and $\psi_{2i} = \int_0^\infty \{ Z_i - \bar{z}^{(2)}(t) \} dM_{2i}(t)$. By extending the asymptotic linearity arguments of Ying (1993), we can show that, for η in a small neighborhood of η_0,

$$n^{-1/2} \mathbf{U}(\eta_0) = n^{-1/2} \mathbf{U}(\eta_0) + \mathbf{A} n^{1/2}(\eta - \eta_0) + o_P(1),$$

(16)

where $\mathbf{A}$ is the asymptotic slope matrix of $n^{-1}\mathbf{U}(\eta_0)$. It follows that $n^{1/2}(\eta - \eta_0)$ is asymptotically zero-mean normal with covariance matrix $\mathbf{A}^{-1}\mathbf{V}\mathbf{A}^{-1}$. The asymptotical normality of $n^{1/2}(\widehat{RE} - RE)$ then immediately follows by the multivariate delta method.

References

Begg CB, Leung DHY. On the use of surrogate endpoints in randomized trials (with discussion). Journal of the Royal Statistical Society, Series A 2000;163:15–28.

Berger VW. Does the Prentice criterion validate surrogate endpoints? Statistics in Medicine 2004;23:1571–1578. [PubMed: 15122737]

Berman E, Heller G, Santorsa J, McKenzie S, Gee T, Kempin S, Gulati S, Andreeff M, Kolitz J, Gabrilove J. Results of a randomized trial comparing idarubicin and cytosine arabinoside with daunorubicin and cytosine arabinoside in adult patients with newly diagnosed acute myelogenous leukemia. Blood 1991;77:1666–1674. [PubMed: 2015395]

Biomarkers Working Group. Biomarkers and surrogate endponts: preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics 2001;69:89–95. [PubMed: 11240971]

Burzykowski, T.; Molenberghs, G.; Buyse, M. The Evaluation of Surrogate Endpoints. New York: Springer-Verlag; 2005.

Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. Applied Statistics 2001;50:405–422.

Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 1998;54:1014–1029. [PubMed: 9840970]

Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analysis of randomized experiments. Biostatistics 2000;1:49–67. [PubMed: 12933525]

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.

Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Statistics in Medicine 1997;16:1515–1527. [PubMed: 9249922]

Day R, Bryant J, Lefkopolou M. Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. Biometrika 1997;84:45–56.

Fine JP, Jiang H, Chappell R. On semi-competing risks data. Biometrika 2001;88:907–919.

Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. New York: John Wiley and Sons; 1991.

Frangakis C, Rubin D. Principal stratification in causal inference. Biometrics 2002;58:21–29. [PubMed: 11890317]

Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic disease. Statistics in Medicine 1992;11:167–178. [PubMed: 1579756]

Gail MH, Pfeiffer R, van Houwelingen HC, Carroll RJ. On meta-analytic assessment of surrogate outcomes. Biostatistics 2000;1:231–246. [PubMed: 12933506]

Ghosh D. Semiparametric inferences for association with semi-competing risks data. Statistics in Medicine 2006;25:2059–2070. [PubMed: 16196081]

Ghosh D. Semiparametric inference for surrogate endpoints with bivariate censored data. Biometrics 2008;64:149–156. [PubMed: 17651457]

Jin Z, Ying Z, Wei LJ. A simple resampling method by perturbing the minimand. Biometrika 2001;88:381–390.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. New York: Wiley; 2002.

Lin DY, Fleming TR, DeGruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Statistics in Medicine 1997;16:1515–1527. [PubMed: 9249922]

Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. Biometrika 1996;83:381–393.

Louis TA. Nonparametric analysis of an accelerated failure time model. Biometrika 1981;68:381–390.

Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Controlled Clinical Trials 2002;23:607–625. [PubMed: 12505240]

Nelsen, R. An Introduction to Copulas. New York: Springer; 1999.

Oakes D. A model for association in bivariate survival data. J. R. Statist. Soc. B 1982;44:414–422.

Oakes D. Semiparametric inference in a model for association in bivariate survival data. Biometrika 1986;73:353–361.

Oakes D. Bivariate survival models induced by frailties. J. Am. Statist. Assoc 1989;84:487–493.

Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. Biometrika 1994;81:341–350.

Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. Journal of the American Statistical Association 2006;101:1085–1093.

Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in Medicine 1989;8:431–440. [PubMed: 2727467]

Reid N. A conversation with Sir David Cox. Statistical Science 1994;9:439–455.

Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, Francini G, Grothey A, O'Connell M, Catalano PJ, Blanke CD, Kerr D, Green E, Wolmark N, Andre T, Goldberg RM, De Gramont A. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. Journal of Clinical Oncology 2005;23:8664–8670. [PubMed: 16260700]

Robins JM. An analytic method for randomized trials with informative censoring: part I. Lifetime Data Analysis 1995;1:241–254. [PubMed: 9385104]

Taylor JM, Wang Y, Thiébaut R. Counterfactual links to the proportion of treatment effect explained by a surrogate marker. Biometrics 2005;61:1102–1111. [PubMed: 16401284]

Tsai WY. Testing the assumption of independence of truncation time and failure time. Biometrika 1990;77:169–177.

Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. Proc Natl Acad Sci USA 1975;72:20–22. [PubMed: 1054494]

Wang W. Estimating the association parameter for copula models under dependent censoring. Journal of the Royal Statistical Society Series B 2003;65:257–273.

Wang W, Wells MT. Model selection and semiparametric inference for bivariate failure-time data (with discussion). Journal of the American Statistical Association 2000;95:62–76.

Wang Y, Taylor JM. A measure of the proportion of treatment effect explained by a surrogate marker. Biometrics 2002;58:803–812. [PubMed: 12495134]

Ying Z. A large sample study of rank estimation for censored regression data. Annals of Statistics 1993;21:76–99.

Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". Journal of Educational and Behavioral Statistics 2003;28:353–368.

Table 1

Simulation results for relative effect

RE	θ	n	Naive					Proposed				
			Bias	SE	SEE	CI1	CI2	Bias	SE	SEE	CI1	CI2
0.4	1.1	50	0.19	0.38	0.37	0.88	0.87	0.01	0.43	0.41	0.97	0.96
		100	0.25	0.27	0.25	0.81	0.79	−0.02	0.32	0.32	0.95	0.96
		150	0.32	0.20	0.21	0.68	0.71	−0.03	0.24	0.24	0.95	0.95
	1.6	50	0.23	0.33	0.31	0.85	0.84	0.02	0.37	0.35	0.96	0.96
		100	0.35	0.23	0.22	0.72	0.75	−0.02	0.26	0.24	0.96	0.96
		150	0.41	0.19	0.18	0.66	0.64	0.01	0.39	0.35	0.97	0.97
0.8	1.1	50	0.25	0.36	0.35	0.9	0.91	0.01	0.39	0.35	0.97	0.97
		100	0.36	0.25	0.25	0.78	0.76	−0.03	0.28	0.26	0.96	0.96
		150	0.43	0.20	0.19	0.69	0.68	0.02	0.22	0.20	0.96	0.95
	1.6	50	0.33	0.33	0.32	0.88	0.87	0.01	0.38	0.37	0.96	0.96
		100	0.39	0.23	0.22	0.79	0.81	−0.01	0.25	0.24	0.95	0.95
		150	0.47	0.18	0.19	0.66	0.65	0.02	0.21	0.20	0.96	0.95

Note: θ denotes correlation between S and T in bivariate Clayton-Oakes model used to simulate data. SEE denotes average of standard errors based on empirical distribution of η^*; CI1 denotes coverage probability of 95% CI using standard error calculated from empirical distribution of η^*; CI2 denotes coverage probability of 95% CI using 2.5th and 97.5th percentiles calculated from empirical distribution of η^*.

Table 2

Simulation results for adjusted association

| | $\theta_0 = \theta_1 = 1.6$ | | | | $\theta_0 = 2; \theta_1 = 4$ | | | |
n	Bias	SE	SEE	95% CI	Bias	SE	SEE	95% CI
50	−0.02	0.39	0.37	0.93	−0.54	0.42	0.40	0.80
100	−0.01	0.28	0.26	0.94	−0.67	0.30	0.29	0.65
150	0.01	0.23	0.22	0.95	−0.89	0.24	0.24	0.51

noindent **Note:** θ_0 and θ_1 denote dependence parameters in Clayton-Oakes model for control and treatment groups, respectively. In situation where $\theta_1 \neq \theta_2$, bias is calculated using the true theta value is 2. SE is the standard error of θ across simulation samples; SEE denotes average of standard errors based on empirical distribution of θ.