Addressing an idiosyncrasy in estimating survival curves using double-sampling in the presence of self-selected right censoring.

Constantine E. Frangakis
Department of Biostatistics, Johns Hopkins University
615 N. Wolfe St., Baltimore, MD 21205, U.S.A.
cfrangak@jhsph.edu

and

Donald B. Rubin
Department of Statistics, Harvard University
1 Oxford St., Cambridge, MA 02138, U.S.A.
rubin@stat.harvard.edu

In Biometrics (2001, 57: 333-353), with discussion.

SUMMARY. We investigate the use of follow-up samples of individuals to estimate survival curves from studies that are subject to right-censoring from two sources: (i) early termination of the study, namely, administrative censoring, or (ii) censoring due to lost data prior to administrative censoring, so-called dropout. We assume that, for the full cohort of individuals, administrative censoring times are independent of the subjects' inherent characteristics, including survival time. To address the loss to censoring due to dropout, which we allow to be possibly selective, we consider an intensive second phase of the study where a representative sample of the originally lost subjects is subsequently followed and their data recorded. As with double-sampling designs in survey methodology, the objective is to provide data on a representative subset of the dropouts. Despite assumed full response from the follow-up sample, we show that, in general in our setting, administrative censoring times are not independent of survival times within the two subgroups, nondropouts and sampled dropouts. As a result, the stratified Kaplan-Meier estimator is not appropriate for the cohort survival curve. Moreover, using the concept of potential outcomes, as opposed to observed outcomes, and thereby explicitly formulating the problem as a missing data one, reveals and addresses these complications. We present an estimation method based on the likelihood of an easily observed subset of the data, and study its properties analytically for large samples. We evaluate our method in a realistic situation by simulating data that match published margins on survival and dropout from an actual hip-replacement study. Limitations and extensions of our design and analytic method are discussed.

KEY WORDS: Double-sampling; Dropouts; Loss to follow-up; Potential outcomes; Rubin Causal Model.

1. Introduction

1.1 Motivation.

When studies follow subjects over time and investigate a time-to-event outcome T, such as survival, the outcome may not be available for all subjects at the end of the study. We focus on the situation with the simultaneous occurrence of: (i) administrative censoring, that is, censoring due to early termination of the study, and (ii) dropout, or loss to follow-up, which occurs when the subject interrupts contact with the investigator before the end of the study and before the survival time is observed. An important example is patients who have surgery in which a human joint, e.g, the hip, is replaced with an artificial one, a prosthesis. Often, the prosthesis wears out and needs replacement, and surgeons are interested in the time, T, between the first surgery and the next replacement, that is, the survival time of the prosthesis. Our work is motivated by the large number of dropout patients often occurring after surgery of joints (e.g., Gartland, 1988; Dorey and Amstutz, 1989; Murray, Carr, and Bulstrode, 1993). Our goal is to estimate the survival function of the cohort of all subjects entering the study.

In order to estimate this survival function in the presence of censoring, it is common to assume that entry times into the study do not relate to survival and thus, essentially, that administrative censoring is ignorable in the sense of Rubin (1976, 1978). Moreover, it is sometimes assumed that dropout is also ignorable, possibly after conditioning on observed covariates (e.g., Dobbs, 1980). Ignorability of dropout then essentially requires that a subject whose survival is censored by dropout after time t, say, from entry, has comparable survival to a subject whose survival is censored after time t by administrative censoring. Although ignorability of administrative censoring can be plausible, ignorability of dropout is suspect (Sims, 1973; Austin *et al.*, 1979), and one plan to lessen the reliance on this assumption is to use a random (or representative) sample of the original dropouts. Assume that, for this subset of subjects, the investigator successfully obtains the data intended to have been recorded in the absence of dropout, namely either the actual failure time T or knowledge that it would have been administratively censored. Similar two-phase designs are known in the survey literature as double-sampling designs (Cochran, 1963; Glynn, Laird, and Rubin, 1993; Rotnitzky and Robins, 1995; Zanutto, 1998) and the analyses of their data are typically simple when the outcome is either fully observed or fully missing. Problems involving double-sampling with competing risks have been discussed by Hogan and Laird (1996), and Flehinger, Reiser and Yashchin (1998), although under different frameworks, assumptions and goals from ours, whereas Baker, Wax, and Patterson (1993) and Wax, Baker, and Patterson (1993) discussed double-sampling of dropouts assuming discrete survival data and in the absence of competing varying administrative censoring.

1.2 Purpose and outline.

We have two aims. First, by formulating an appropriate framework for our problem that links the concept of potential outcomes (Neyman, 1923; Rubin, 1974, 1978) to competing risks, we show how to use the data from the double-sampled subjects to draw inference for the cohort survival function. Obtaining valid inference using double-sampling with survival data is more subtle than with survey data because of an idiosyncrasy that can render usual analyses incorrect when estimating the cohort survival function. Demonstrating this fact is our second aim, because, to our knowledge, this has not been discussed in the literature.

The procedure we focus on needs only a stratification of the intended data, and, for example, does not rely on the times of study entry and of study-dropout for each dropout subject (such times might not be available to the analyst due to confidentiality constraints). When more than the intended data are available for analysis, our procedure will not be fully efficient, and one could construct locally semiparametric efficient estimators (e.g., Robins, Rotnitzky, and Zhao, 1994), or could use fully parametric methods, and it would be interesting in future work to compare these procedures to the proposed method in realistic finite samples.

In the next section, we develop our framework using potential outcomes that are characteristics inherent to subjects. We posit an assumption that treats administrative censoring as a randomization variable, and thus independent of the potential outcomes. This reveals a specific structure on the two subgroups of subjects for whom the only reason for possibly missing data is administrative censoring: the nondropout group, and the recovered subset of the dropouts. In particular, for each of these two groups, it is tempting to apply standard survival methods that assume independence between survival and administrative censoring times, and then combine these two separate estimates as with two-phase designs in survey literature. Under our assumptions, however, we show that within each of the two subgroups, the administrative censoring times generally correlate with survival times. Consequently, Kaplan-Meier estimators (Kaplan and Meier, 1958) that stratify by these groups, or, alternatively, likelihood-based methods that ignore this correlation, generally do not appropriately estimate the survival curves within either of the subgroups, nor hence, the survival curve for the cohort. In Section 3, we discuss addressing the problem from the perspective of missing data and we derive an estimation procedure that, under our assumptions, appropriately estimates the cohort survival curve. In Section 4, we evaluate our procedure in situations that are realistic in practice. Although double-sampling when faced with both dropout and administrative censoring is natural, we do not have a completely real data example available for demonstration. Consequently, we use published margins for survival curve, dropout rate, sample size accrual rate, and available length of follow-up from an actual hip-replacement study, to project simulated double-sampling data, which we use to compare our procedure to the re-weighted stratified Kaplan-Meier estimators. The final section gives some further remarks for the design and estimation methods. The appendix provides technical details.

2. The Problem

2.1 Potential outcomes and goal.

Consider a population P of subjects, for example, having surgery to have the human hip joint replaced by a prosthesis. A common complication is that subjects drop out, that is, are lost to follow-up, in the sense that they interrupt contact with the investigator before their survival times becomes known. For clarity, we define potential outcomes (Neyman, 1923; Rubin, 1974, 1978) for subjects of the population P before making probability statements, as in the Rubin Causal Model (Holland, 1986). Suppose that subjects $i = 1, ..., n$ from the population participate in a hypothetical study depicted in Fig. 1. Let E_i be the calendar time of entry into the study, e.g., the time of surgery, and T_i be the survival time of the prosthesis, counting from surgery for each subject (Fig. 1(a)). Also in part (a) of the figure, let L_i be the length of time, counting from entry, for which subject i would remain in the study, i.e., without dropping out, if the study remained open indefinitely, where $L_i \leq T_i$. The potential outcome L_i is treated as a characteristic inherent to subject i at this time in this study, like birth-date and gender. For subjects i who would drop out in that hypothetical scenario, $L_i < T_i$, whereas $L_i = T_i$ for subjects who would not drop out before failure of their prosthesis no matter when the study would end, so we say that the indicator $R_i = \delta(L_i = T_i)$ is the true dropout status, where $\delta(\cdot)$ is the indicator function. [Figure 1 about here.]

Now that the potential outcomes have been defined, we consider the actual study. Let E_{max} be the calendar time when the study actually ends, e.g., June 1 1999, as in Fig. 1(b). Using standard notation, we let $C_i = E_{max} - E_i$ be the administrative censoring time, and $X_i = \min(T_i, C_i)$ and $\Delta_i = \delta(T_i < C_i)$, be, respectively, the length of survival time that lies within the administrative censoring time, and the indicator for whether that length is the full survival time. The data (X_i, Δ_i) are the intended data, that is, the information the investigator would record in the absence of any dropout. But when $L_i < X_i$, neither X_i nor Δ_i is observed, but rather (L_i, R_i^{obs}) where $R_i^{obs} := \delta(L_i \geq X_i)$; $R_i^{obs} = 1$ for subjects i who are not dropouts in the study, whereas $R_i^{obs} = 0$ for the dropouts in the study (Fig. 1(c)).

Note that the observed, study-dropout status, R_i^{obs}, is not always equal to the true dropout status R_i. For example, by comparing Fig. 1 (a) to (c), subjects 4 and 5 who are administratively censored in the study (and thus have $R_i^{obs} = 1$) are a mixture of true nondropouts $(R_i = 1)$ and true dropouts $(R_i = 0)$.

We let $S(t)$ be the fraction of all subjects in P whose survival time T exceeds t, for $t > 0$. Our goal is to learn about the survival function $S(t)$. Our methods can be extended to allow for observed covariates. However, to communicate our main arguments without complicating notation, we assume that no covariates are recorded or, alternatively, that we are already within cells of observed covariates.

2.2 Initial aspects of study design.

Assume that subjects $i = 1, ..., n$ are a simple random sample from the population P, and that P is large enough so that observations on different sampled subjects can be treated as independent. We focus discussion on cohorts P of subjects that are homogeneous (e.g., in terms of a common surgical method over time) in the sense that entry times E_i are not related with either survival times, T_i, or true dropout times L_i. We can express this in the following assumption, which we will make throughout the rest of the paper.

ASSUMPTION 1. *Randomness of entry times:*

$$\Pr(T_i, L_i, E_i) = \Pr(T_i, L_i) \Pr(E_i),$$

where, here and in the sequel, the measure Pr() is the one induced by the random sampling from the population P.

One implication of Assumption 1 is that, in the full cohort of subjects, survival and administrative censoring times are independent: $\Pr(T_i, C_i) = \Pr(T_i)\Pr(C_i)$, which is a common assumption (Fleming and Harrington, 1991, and references therein; Hougaard, 1999). Another immediate implication is that true dropout status and administrative censoring times are independent: $\Pr(R_i, C_i) = \Pr(R_i)\Pr(C_i)$.

2.3 Double-sampling and observed data.

In joint-replacement studies, dropout occurs for a variety of reasons (e.g., the subject moves, has pain and visits a different physician, feels very well and interrupts the hospital visits). Because dropout is possibly related to survival (as in similar settings, Sims, 1973; Austin *et al.*, 1979), we consider a second phase of data collection, where the investigator selects a subset of the study-dropouts, (i.e., with $R_i^{obs} = 0$), and pursues them intensively enough to record their intended data (X_i, Δ_i), which would have been observed at calendar time E_{max}. We let $S_i = 1$ if subject i has $R_i^{obs} = 0$ and is pursued in this second phase, such as subject 2 in Fig. 1(d), and we let $S_i = 0$ otherwise. Pursuing dropouts is costly (e.g., Dorey and Amstutz, 1989), so the size of the sample would depend on available resources and rarely include all those with $R_i^{obs} = 0$. For situations where (X_i, Δ_i) is still not recorded among some subjects with $S_i = 1$, see Section 5. Also relevant to the design of the second phase of data collection is the following issue.

The data (X_i, Δ_i) for dropouts with early entry times generally carry more information than those with later entry times in the sense that the former subjects have been in the study longer. Moreover, early entry subjects are more likely to be study-dropouts because they have been followed for a longer period. Therefore, a simple random sample from $\{i : R_i^{obs} = 0\}$ will automatically oversample study-dropouts with earlier entry times. For this reason, and because we generally will not know exactly the optimal design without prior knowledge of the survival times of the dropouts, we assume for simplicity that, from those with $R_i^{obs} = 0$, the investigator chooses the set $\{i : S_i = 1\}$ by simple random sampling that, for convenience, we treat as independent Bernoulli trials with common and known selection probability Pr$(S_i = 1|R_i^{obs} = 0) = p^{(S)}$. We comment on extensions in Section 5.

2.4 Idiosyncrasy.

After the double-sampling, we have two groups of subjects with (X_i, Δ_i): the nondropouts in phase 1, $\{i : R_i^{obs} = 1\}$, and the sampled subset $\{i : S_i = 1\}$ of the dropouts; for both groups, the only reason for missing survival time is administrative censoring. The two groups can have different survival curves, and, because only some subjects of those with $R_i^{obs} = 0$ are recovered, an analysis that ignores the group classification, $\{R_i^{obs} = 1\}$ versus $\{S_i = 1\}$, after double-sampling (e.g., as in Dorey and Amstutz, 1989), is not appropriate generally.

Therefore, in this section we consider the consequence of using the following procedure: (i) constructing Kaplan-Meier estimators within these two observed subgroups, and then (ii) combining the two estimators, weighted by the appropriate proportions of $\{i : R_i^{obs} = 1\}$ and $\{i : R_i^{obs} = 0\}$, to estimate the cohort survival function $S(t)$. This stratified Kaplan-Meier (SKM) statistic is the standard estimator when censoring is solely administrative. The statistic SKM would be consistent for $S(t)$ if the administrative censoring times C_i were independent of the survival times T_i for study-nondropouts, i.e., with $R_i^{obs} = 1$, and for study-dropouts, i.e., with $R_i^{obs} = 0$. However, this independence does not generally hold under Assumption 1 in either group; the following result is derived as an application of Bayes' theorem and after some algebra (proof omitted).

RESULT 1. *Under Assumption 1, and with positive c and t,*

$$\begin{aligned}\Pr(T > t|R^{obs} = 1, C = c) =& \Pr(T > t|R = 1)\\ & \times \left\{\frac{\Pr(R = 1) + u(c,t)\Pr(L > c, R = 0)}{\Pr(R = 1) + \Pr(L > c, R = 0)}\right\},\end{aligned}$$

$$\textit{where} \quad u(c,t) := \frac{\Pr(T > t|L > c, R = 0)}{\Pr(T > t|R = 1)}, \quad \textit{and}$$
$$\Pr(T > t|R^{obs} = 0, C = c) = \Pr(T > t|R = 0, L < c).$$

In Result 1, $\Pr(T > t|R^{obs}, C = c)$ is generally a function of c and, therefore, survival and administrative censoring times are correlated within both strata defined by observed dropout status R^{obs}. This complication is not due to the simple random sampling design from the study-dropouts in the second phase because the correlation between survival and censoring times is also present in the study-nondropouts. Focusing on the latter group, Result 1 shows that C_i and T_i would be independent if $u(c,t) = 1$ for all c, t. A necessary and sufficient set of conditions for this is that: (i) survival times T_i be independent of true dropout statuses R_i, and (ii) among true dropouts, $\{i : R_i = 0\}$, survival times T_i be independent of dropout times L_i. However, by definition for true dropout, $L_i < T_i$ for all i. If, in addition, there is a survival time in the cohort that is smaller than the largest observed dropout time, which is expected to be true in most realistic cases, then it can be easily shown that conditions (i) and (ii) above cannot both hold. It follows that the straightforward within-stratum Kaplan-Meier estimators are generally inconsistent and, hence, that generally the combined statistic SKM is not an appropriate estimator for the cohort survival curve $S(t)$. Result 1 also implies that standard "ignorable" (Rubin, 1978) likelihood or Bayesian survival methods that stratify on R^{obs} and ignore the correlation between survival and administrative censoring times would also suffer from bias, even in large samples with the correct model for $S(t)$.

The spurious correlation within observed groups in Result 1 occurs because the observed data are non-trivial functions of the potential outcomes that reflect scientifically relevant characteristics of subjects. The distinction between potential outcomes and observed data occurs, more generally, in studies that suffer from deviations from protocol, including treatment-noncompliance and incomplete outcomes, and can be critical in defining and estimating quantities of interest (e.g., Rubin, 1978; Robins and Greenland, 1994; Angrist, Imbens and Rubin, 1996; Frangakis and Rubin, 1999; Robins, Greenland and Hu, 1999; Rubin and Frangakis, 1999).

3. Addressing the Problem: Missing Data Perspective

3.1 General.

Because the selection mechanism of study-dropouts at the second phase is known conditionally on the first phase, any data (X_i, Δ_i) missing in stratum $(R_i^{obs} = 0)$ after double-sampling are missing at random (Rubin, 1976), a fact that we can use for robust estimation using likelihood principles. Under Assumption 1, the likelihood function of data

$$D = \{R_i^{obs}, C_i, S_i\} \cup \{X_i, \Delta_i : R_i^{obs} = 1\} \cup \{L_i : R_i^{obs} = 0, S_i = 0\} \cup \{X_i, \Delta_i, L_i : R_i^{obs} = 0, S_i = 1\}$$

is proportional to

$$\begin{aligned} L(\theta \mid D) = \prod_i & \text{Pr}(X_i, \Delta_i, C_i, R_i^{obs} = 1; \theta)^{R_i^{obs}} \\ & \times \text{Pr}(C_i, L_i, R_i^{obs} = 0; \theta)^{(1-R_i^{obs})(1-S_i)}\, \text{Pr}(X_i, \Delta_i, C_i, L_i, R_i^{obs} = 0; \theta)^{S_i} \end{aligned}$$

where θ represents parameters governing the distribution of all observables. Consider also the likelihood function of the reduced data $D^* = \{R_i^{obs}, S_i\} \cup \{(X_i, \Delta_i) : R_i^{obs} = 1 \text{ or } S_i = 1\}$, proportional to

$$L(\theta \mid D^*) = \prod_i \left\{\Pr(X_i, \Delta_i | R_i^{obs} = 1; \theta)\right\}^{R_i^{obs}} \left\{\Pr(X_i, \Delta_i | R_i^{obs} = 0; \theta)\right\}^{S_i}$$
$$\times \Pr(R_i^{obs} = 1; \theta)^{R_i^{obs}} \{\Pr(R_i^{obs} = 0; \theta)\}^{1-R_i^{obs}}.$$

Making inference on $S(t)$ by maximizing $L(\theta \mid D)$ when θ is unrestricted is generally not possible. Alternative ways to address the problem, under Assumption 1, include

(i) maximizing the reduced data likelihood $L(\theta \mid D^*)$ with unrestricted θ;

(ii) positing semiparametric submodels, θ_s, for θ;

(iii) positing fully parametric submodels for θ.

For approach (i), the components of θ that are identifiable from $L(\theta \mid D^*)$ identify $S(t)$, so this method will produce a consistent estimator of the survival curve for general distributions θ. This method provides the basic intuition for how the problem can be addressed. Because this method uses only the R^{obs}–stratification of the intended data (X_i, Δ_i) of nondropouts and of dropouts that have been double-sampled, it can be applied even in constrained situations, for example, when the entry times $E_i (= E_{max} - C_i)$ and dropout times L_i are hidden from the analyst because of confidentiality concerns.

For approach (ii), work of Robins, Rotnitzky, and Zhao (1994) can be used to construct estimators for $S(t)$ that take the form of inverse probability of censoring weighted (IPCW) statistics and include estimator (i) as a member. The θ_s–optimal estimators within that class asymptotically: are efficient when the working model θ_s is correct; remain consistent when θ_s is incorrect; and are more efficient than that of (i) when the additional data $D - D^*$ are available. In approach (iii) maximum likelihood or Bayesian inference can be used.

We focus on approach (i), thereby providing insight into the situation in the simple case with minimal data available. Nevertheless, it would be interesting to compare the three methods to see how much efficiency is lost using (i) relative to (ii) and (iii) in realistic settings where the extra information is available.

3.2 Maximum likelihood from $L(\theta \mid D^*)$.

Assuming that the random variables T and C are absolutely continuous, first we re-parameterize the problem in terms of hazard functions. Define the net hazard function of interest for the full cohort of subjects, $\lambda^{\text{net}}(t) := \lim_{d\to 0}\{\text{Pr}\ (t \le T < t+d|T \ge t)/d\}$. By Assumption 1, survival and administrative censoring times are independent in the full cohort. Therefore, the net hazard function in the full cohort is equal to the crude hazard function in the full cohort, defined as $\lambda^{\text{crd}}(t) := \lim_{d\to 0}\{\text{Pr}(t \le T < t+d|X \ge t)/d\}$ (e.g., Fleming and Harrington, 1991). Furthermore, we can generally decompose the crude hazard function of the full cohort to the crude hazard functions for the two subgroups defined by the observed dropout status R^{obs}, $\lambda_g^{\text{crd}}(t) := \lim_{d\to 0}\{\text{Pr}\ (t \le T < t+d|X \ge t, R^{obs} = g)/d\}$, using

$$\begin{aligned} \lambda^{\text{crd}}(t) &= \sum_{g=0,1} w_g(t)\lambda_g^{\text{crd}}(t), \qquad\qquad \text{where} \\ w_g(t) :=& \text{Pr}(R^{obs} = g|X \ge t) = \frac{\pi_g(t)p_g}{\sum_{g'=0,1}\pi_{g'}(t)p_{g'}}, \end{aligned} \tag{3.2}$$

and where $\pi_g(t) = \text{Pr}(X \ge t|R^{obs} = g)$ are the probabilities of being in the "risk set" at time t, $p_g = \text{Pr}(R^{obs} = g)$, for $g = 0, 1$, and the expressions in (3.2) follow from Bayes' theorem. The time dependent weights $w_g(t)$ reflect that (3.2) is a stratification at every time point. At this stage, the crude hazard functions $\lambda_g^{\text{crd}}(t)$, and the probabilities $\pi_g(t)$ and p_g can be estimated by maximizing the likelihood function $L(\theta \mid D^*)$ with no further modeling assumptions.

For convenience in presentation, abbreviate the two groups $\{i : R_i^{obs} = 1\}$ and $\{i : S_i = 1\}$ by H_1 and H_0, respectively. Using standard notation, define the "risk set" and failure processes:

for each individual, by $Y_i^*(t) := \delta(X_i \geq t)$ and $N_i^*(t) := \delta(X_i \leq t)\Delta_i$; and for the groups $H_g, g = 0, 1$, by $Y_g(t) := \sum_{i \in H_g} Y_i^*(t)$ and $N_g(t) := \sum_{i \in H_g} N_i^*(t)$. Using the notation $dQ(t)$ for the differential of right-continuous processes $Q(t)$, we define the processes $\hat{\Lambda}_g^{\text{crd}}(t) := \int_{(0,t)} dN_g(u)/Y_g(u)$, and

$$\begin{aligned} \hat{\Lambda}(t) &:= \sum_{g=0,1} \int_0^t \hat{w}_g(u) d\hat{\Lambda}_g^{\text{crd}}(u), && \text{where} \\ \hat{w}_g(t) &:= \frac{\hat{\pi}_g(t)\hat{p}_g}{\sum_{g'=0,1} \hat{\pi}_{g'}(t)\hat{p}_{g'}}, && \hat{\pi}_g(t) := Y_g(t)/n_g, \end{aligned} \tag{3.3}$$

and where n_g is the number of people in group H_g, and $\hat{p_g} = \sum_i \delta(R_i^{obs} = g)/n$, for $g = 0, 1$. The processes $\hat{\Lambda}_g^{\text{crd}}(t)$ maximize the likelihood $L(\theta \mid D^*)$ with respect to, and are centered approximately at, the cumulative crude hazard functions $\Lambda_g^{\text{crd}}(t) := \int_{(0,t)} \lambda_g^{\text{crd}}(u)du$, regardless of dependence between survival and administrative censoring times. Similarly, the empirical distributions $\hat{\pi}_g(t)$ and the ratios $\hat{p_g}$ maximize $L(\theta \mid D^*)$ with respect to, and are centered approximately at, $\pi_g(t)$ and p_g respectively. Then, by (3.2) and (3.3), the process $\hat{\Lambda}(t)$ will be centered approximately at $\int_{(0,t)} \lambda^{\text{crd}}(u)du$, and thus, by Assumption 1, at the net cumulative hazard function $\Lambda^{\text{net}}(t) := \int_{(0,t)} \lambda^{\text{net}}(u)du$.

The estimator $\hat{\Lambda}(t)$ in (3.3) can also be viewed as having a generalized form of a "weighted logrank" statistic (Fleming and Harrington, 1991), where, here, the weights vary both across time and between the two estimated crude hazard functions. However, the approximate variability of $\hat{\Lambda}(t)$ around the estimand, $\sum_{g=0,1} \int_0^t w_g(u)\lambda_g^{\text{crd}}(u)du = \Lambda^{\text{net}}(t)$, (given in (A.2)), has a different form from that of usual weighted logrank statistics. For a time $t_m < E_{max}$ such that Pr$(T > t_m | R^{obs} = g)$ and Pr$(C > t_m | R^{obs} = g)$ are positive for $g = 0, 1$, the following result, whose proof is outlined in the appendix, summarizes the large-sample distribution of $\hat{\Lambda}(t)$.

RESULT 2. *Under Assumption* 1, *and for* $0 < t < t_m$,

$$n^{\frac{1}{2}} \left\{ \hat{\Lambda}(t) - \Lambda^{\text{net}}(t) \right\} \to W(t),$$

weakly in distribution, as $n \to \infty$, *where* $W(t)$ *is a Gaussian process with* $E\{W(t)\} = 0$. *Also, for* s *with* $0 < s, t < t_m$, *the covariance* $\operatorname{cov}\{W(s), W(t)\}$ *has the form* (A.2) *given in the appendix.*

The expression for $\operatorname{cov}\{W(s), W(t)\}$ in (A.2) involves quantities, defined in the appendix, that are functions of the unknowns $\Lambda_g^{\text{crd}}(t)$, $\pi_g(t)$, and p_g, $g = 0, 1$. By replacing the latter with their sample analogues $\hat{\Lambda}_g^{\text{crd}}(t)$, $\hat{\pi}_g(t)$, and $\hat{p}_g$ in all the defining expressions in the appendix, we let the statistic $\hat{V}(s, t)$ be the resulting sample analogue of expression (A.2). Then, using Result 2 and Slutsky's theorem, it can be easily shown that, for times t with $0 < t < t_m$,

$$n^{\frac{1}{2}} \left\{ \hat{\Lambda}(t) - \Lambda^{\text{net}}(t) \right\} \left\{ \hat{V}(t, t) \right\}^{-\frac{1}{2}} \to N(0, 1), \tag{3.5}$$

in distribution, as $n \to \infty$. We can then use (3.5) to obtain confidence intervals (pointwise) for $\Lambda^{\text{net}}(t)$. Using the identity $S(t) = \exp\{-\Lambda^{\text{net}}(t)\}$, we obtain an estimator, $\hat{S}(t) := \exp\{-\hat{\Lambda}(t)\}$, that is consistent for the survival curve. We can use the same relation to obtain confidence intervals for $S(t)$ as the reciprocals of the exponentiated confidence intervals for $\Lambda^{\text{net}}(t)$.

Note that, because this estimation method is based on likelihood $L(\theta \mid D^*)$ instead of $L(\theta \mid D)$, it requires only the implication of Assumption 1 that survival T_i and administrative censoring times C_i are independent in the cohort, and partially unobserved, group of subjects. The full structure of Assumption 1 was important in Result 1 to demonstrate that stratified Kaplan-Meier estimators are generally not appropriate in this setting.

4. Projected Simulations from Actual Study

4.1 Setting.

Results for the "Tharies" type of hip-replacement prostheses were reported by Dorey and Amstutz (1989), who discussed 1985 data for the first 100 such prostheses implanted in their hospital starting in 1975. These prostheses were implanted in the first 2-year window (1975-1977) and, although they all had a potential follow-up time C_i of at least 8 years, 45 (45%) of the patients were lost to follow-up within 8 years following their surgery and before prosthesis failure (i.e., $L_i < 8$, although the dropout times are not known to us). Among those original dropouts, the intended data (X_i, Δ_i) were recovered for 35 patients (78%=35/45) after an intensive search reported by the investigators. Subsequently, the survival curve for the cohort was computed using a single Kaplan-Meier estimator, where the non-sampled dropouts were treated as administratively censored. Because the general inconsistency of this unstratified Kaplan-Meier estimator in our framework follows from techniques used in Sec. 2.4, we do not simulate detailed results for it, but we use their reported curve as a reference "true curve" below. In addition, the study had, essentially, no administrative censoring because patients who got Tharies in the 8-year window 1977-1985 were ignored. In this Section, by projecting information from the Tharies study to allow potential inclusion of patients from the whole 10-year window available in 1985, we will compare our method with the stratified Kaplan-Meier when both dropout and administrative censoring are present, in conditions summarized in Table 1.

To reflect looking at accrual within the 10-year window, in all 12 conditions of Table 1, we set $E_{max} = 10$ years and we simulated entry times E_i uniformly in (0,10) [so, $C_i = (10 - E_i) \sim U(0, 10)$]. We matched points on the Tharies empirical survival curve of Dorey and Amstutz (1989; Fig. 1, "complete analysis") to a model for T_i where $T_i^* := \log(T_i)$ is normally distributed, giving mean $\mu_{T^*} = \log(9.3 \text{ years})$ and standard deviation $\sigma_{T^*} = 0.728$. The resulting survival probabilities $S(t)$ at times $t = 6, 7,$ and 8 years are 72.65%, 65.19%, and 58.20%, and, in the following, will be treated as true and will be our estimation goal. The simulations follow Assumption 1.

The Tharies' study-dropout rate of 45% in the first 8 years of follow-up means that the fraction of true dropouts $(R_i = 0)$ would be larger than 45%. To reflect this, we fixed $\Pr(R_i = 0) = 50\%$, and used the models described in the next paragraphs to get study-dropout rates $p^{mis}_{C>8} := \Pr(R^{obs}_i = 0|C_i > 8)$ in the range 40% to 46% when, as with the Tharies' 45% study-dropout, we condition on $C_i > 8$. We also report our rates $p^{mis} := \Pr(R^{obs}_i = 0)$ over the full 10-year window. To investigate plausible conditions for the remaining associations between true dropout status, survival, and dropout time among true dropouts, for which we do not have further information from the Tharies study, we posit the following models.

We draw R_i conditionally on the survival times following the probit model $\Pr(R_i = 1|\log(T_i) = t) = \Phi\{\alpha_R+\delta_R t\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Because we have fixed the marginal probability of being a true dropout to 50%, we vary $\delta_R = -1, 0, 1$, which determines $\alpha_R\,(= 2.23, 0, -2.23)$. Under this setting, $\log(T)$ has the same standard deviation within both sets $\{i : R_i = 0\}$ and $\{i : R_i = 1\}$, and $|\delta_R| = 1$ generates a difference of 1.05 standard deviations between the means of $\log(T)$ in the two groups defined by R. So, the meaning of δ_R is how long the true nondropout patients' Tharies would last compared to true dropouts.

Among true dropouts, $\{i : R_i = 0\}$, we allow the dropout time $L_i \le T_i$ to be related to the survival time T_i. To do this, we simulate L_i from a log-normal regression on $T^*(= \log(T_i))$ and right-truncated at T^*_i. That is,

$$\log(L)|(T^* = t, R = 0) \sim \Pr\{L^*|T^* = t, R = 0, L^* < T^*\}, \text{ where}$$
$$\Pr(L^*|T^* = t, R = 0) = N\left\{\mu_{L^*} + r^*\frac{\sigma_{L^*}}{\sigma_{T^*}}(t - \mu_{T^*}), \sigma^2_{L^*}(1 - r^{*2})\right\},$$

where we fix the mean and standard deviation of L^*, $\mu_{L^*} = \log(E_{max})$ and $\sigma_{L^*} = \sigma_{T^*}$, respectively, for simplicity. By varying the parameter r^* we induced different values of $r := \text{corr}(L_i, T_i | R_i = 0)$, the correlation between patient dropout times and prosthesis survival among the true dropouts.

To match the observed sample size accrual rate of 100 patients in the actual Tharies 2-year window, for each individual study and in each of the 12 conditions defined by the above settings we simulated $n = 500$ patients in the 10-year window. In each individual study, we subsequently double-sampled study-dropouts with a 50% probability each.

We set our goal to compare performance, in terms of coverage rates of nominal 95% confidence intervals and mean squared errors for the survival probability $S(t)$ of Tharies at $t = 6, 7$ and 8 years, and evaluated over 2500 replicated studies for each condition in Table 1, using (i) the procedure described in Section 3, labeled IML because it is the maximum likelihood of the R^{obs}–stratification of the intended data; and (ii) the procedure labeled SKM, based on the stratified Kaplan-Meier estimator described in Section 2.4 and a normal approximation to its distribution. The standard error for SKM is obtained by the delta method; these standard errors were very close to the sampling standard deviations of SKM as computed over the simulation replications for each condition (not shown). An S-plus5 program with FORTRAN subroutines for general implementation of procedure IML, and an S-plus5 program for generating the simulations described here are available from the authors.

4.2 Results.

Based on the conditions in Table 1, when the true dropout patients' Tharies have longer survival times than those of the true nondropouts (conditions with $\delta_R = -1$), SKM and IML perform comparably except for the cases where the correlation r between patients' dropout time and Tharies survival is 0.34 or 0.16, where IML is superior to SKM. These cases, among all those with $\delta_R = -1$, have the highest fraction of study-dropouts $p^{mis}_{C>8}$ and closest to the observed 45% based on the 100 patients reported by Dorey and Amstutz (1989). This indicates that, even with relatively small association between survival time and dropout time among true dropouts, the fraction of study-dropouts influences the performance of the SKM procedure. [Table 1 here]

This influence can also be seen in the conditions where true dropouts would have the same Tharies survival distribution as the true nondropouts (conditions with $\delta_R = 0$). Because the survival distribution for the whole cohort is the same across all sensitivity conditions, the prostheses for true dropout patients must have shorter survival times under conditions $\delta_R = 0$ than under conditions $\delta_R = -1$. Consequently, by comparing the results assuming $\delta_R = -1$ to those assuming $\delta_R = 0$ with comparable values of the correlation r, we observe that (i) there are larger study-dropout rates, $p^{mis}_{C>8}$ and p^{mis}, and (ii) in these cases, SKM notably undercovers the true probability values. Nevertheless, IML has sufficient coverage, and is also more accurate than SKM as indicated by the mean squared errors.

Finally, the conditions of Table 1 in which the prostheses for true dropout patients have smaller survival times than the prostheses for true nondropouts (conditions with $\delta_R = 1$) are associated with high correlation, r, between patients' dropout time and Tharies survival. Also, these conditions, among all conditions in the Table, give the highest study-dropout rates $p^{mis}_{C>8}$, and the closest to the observed 45% from the actual study. In these conditions, SKM performs very poorly, whereas IML performs quite well.

5. Further Remarks and Extensions

Among the dropouts pursued in the second phase, there can be a small number of cases for which the data (X_i, Δ_i) still do not get recorded, perhaps if a person has died before E_{max}. If, in these cases, one still wished to project a hypothetical status of the prosthesis at E_{max}, which would have to be assessed based on other observed information, we assume these special cases would be treated as administratively censored, and, therefore, practically we can still assume that if $S_i = 1$, then the data (X_i, Δ_i) are recovered.

As noted in Section 2.3, the data (X_i, Δ_i) for dropouts with early entry times generally carry more information than those with later times. However, a nonprobability sampling in the second phase that would include only study-dropouts with early entry times cannot generally give appropriate inference for the cohort without using parametric model extrapolation because, by Result 1, the study-dropouts with early entry times have different survival distributions from the study-dropouts with later entry times.

We restricted attention to the simple random sampling at the second phase, for convenience in demonstrating estimation and to clarify the complication of using the stratified Kaplan-Meier estimator even in this simple design. Alternative designs in the second phase can be implemented that use different probabilities of selection for different subjects. When these probabilities depend on continuous, as opposed to discrete, covariates, such as the entry times, the method proposed here based solely on stratification on R^{obs} status would need adjustment to reflect the design probabilities of selection when estimating the cohort survival curve.

As mentioned in Section 3.1, when, in addition to the data D^*, the data D are available for analysis, the proposed method can be further improved by semiparametric or parametric methods. Moreover, the implementation of estimation can be simulation-based, as opposed to analytic, and can be motivated even within our simple setting where closed-form approximate inference exists. For example, one could use the permutation distribution of the data (X_i, Δ_i) of the subsample $\{i : S_i = 1\}$ of study-dropouts to “impute” the missing data (X_i, Δ_i) for the non-subsampled study-dropouts. For each such configuration, a Kaplan-Meier statistic could be calculated for the “imputed” full cohort, where censoring and survival times are independent, and these statistics could then be averaged to give an estimate for the cohort survival curve.

Because the total number of such permutations can be very large, this approach practically would rely on principled simulation methods coupled with principled methods of analysis of multiply imputed datasets (Rubin, 1987; Glynn, Laird and Rubin, 1993; Tu, Meng, and Pagano, 1993; Efron, 1994; Lin, Fleming and Wei, 1994; Rubin, 1996; Wang and Robins, 1998). Simulation-based implementation of estimation can be particularly relevant in extensions of our setting when analytic derivation becomes less tractable, for example, in situations when Assumption 1 is relaxed to allow for non-homogeneous cohorts with calendar time trends in survival and true dropout behavior.

A final comment is that we expect many studies of survival employ at least some type of informal double-sampling of dropouts, and we hope that the availability of our framework and methods stimulate the study and use of double-sampling to address dropout with survival data.

Acknowledgement

The authors thank the orthopedic clinic of IKA Hospital in Melissia, Athens, Greece, for hosting in part this work; Stuart G. Baker and Emanuel Frangakis for useful discussions; the editor and reviewers for helpful comments; and the support from the U.S. National Institute of Child Health and Human Development (R01 HD38209), National Institute of Mental Health (R01 MH56639), National Institute for Drug Abuse (R01 DA10184), the Johns Hopkins H.-C. Yang Memorial Faculty Fund, and the National Science Foundation.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with Discussion). *Journal of the American Statistical Association* **91**, 444–472.

Austin, M. A., Berreyesa, S., Elliott, J. L., Wallace, R. B., Barret–Connor, E., and Criqui, M. H. (1979). Methods for determining long-term survival in a population based study. *American Journal of Epidemiology* **110**, 747–752.

Baker, S. G., Wax, Y., and Patterson, B. H. (1993). Regression analysis of grouped survival data: informative censoring and double sampling. *Biometrics*, **49**, 379–389.

Billingsley, P. (1979). *Probability and Measure*. New York: Wiley.

Cochran, W. (1963). *Sampling Techniques*. New York: Wiley.

Dobbs, H. S. (1980). Survivorship of total hip replacements. *The Journal of Bone and Joint Surgery [Br]* **62**, 168–172.

Dorey, F. and Amstutz, H. C. (1989). The validity of survivorship analysis in total joint arthroplasty. *The Journal of Bone and Joint Surgery [Am]* **71**, 544–548.

Efron, B. (1994). Missing data, imputation, and the Bootstrap (with Discussion). *Journal of the American Statistical Association* **89**, 463–478.

Flehinger, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika* **85**, 151–164.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Frangakis, C. E. (1999). Coexistent complications with noncompliance with study-protocols, and implications for statistical analysis. Unpublished PhD thesis (appendix of Part 3), Department of Statistics, Harvard University.

Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86**, 365–379.

Gartland, J. J. (1988). Orthopaedic clinical research. Deficiencies in experimental design and determination of outcome. *The Journal of Bone and Joint Surgery [Am]* **70**, 1357–1364.

Glynn, R. J., Laird, N. M., and Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. *Journal of the American Statistical Association* **88**, 984–993.

Hogan, J. W. and Laird, N. M. (1996). Intention-to-treat analyses for incomplete repeated measures data. *Biometrics* **52**, 1002–1017.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945-970.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics* **55**, 13–22.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

Murray, D. W., Carr, A. J., and Bulstrode, C. (1993). Survival analysis of joint replacements. *The Journal of Bone and Joint Surgery [Br]* **75**, 697–704.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: essay on principles, Section 9. Translated in *Statistical Science*, **5**, 465–480, 1990.

Robins, J. M. and Greenland, S. (1994). Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* **89**, 737–479.

Robins, J. M., Greenland, S., and Hu, F.-C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome (with Discussion). *Journal of the American Statistical Association* **94**, 687–712..

Robins, J. M., Rotnitzky, A., and Zhao, L.-P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

Rotnitzky, A. and Robins, J. M. (1995). Semiparametric regression with follow-up of nonrespondents. Technical report, Department of Biostatistics, Harvard School of Public Health.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Rubin, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6**, 34–58.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, D. B. (1996). Multiple imputation after 18+ years (with Discussion). *Journal of the American Statistical Association* **91**, 473–489.

Rubin, D. B. and Frangakis, C. E. (1999). Comment on “Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome” by J. M. Robins, S. Greenland, and F.-C. Hu. *Journal of the American Statistical Association* **94**, 702–704.

Sims, A. C. P. (1973). Importance of a high tracing rate in long-term medical follow-up studies. *Lancet* No. 2, 433–435.

Tu, X. M., Meng, X. L., and Pagano, M. (1993). The AIDS epidemic: estimating survival after AIDS diagnosis from surveillance data. *Journal of the American Statistical Association* **88**, 26–36.

Wang, N. and Robins, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85**, 935–948.

Wax, Y., Baker, S. G., and Patterson, B. H. (1993). A score test for non-informative censoring using doubly sampled grouped survival data. *Applied Statistics* **42**, 159–172.

Zanutto, E. L. (1998). Imputation for unit nonresponse: modeling sampled nonresponse follow-up, administrative records, and matched substitutes. PhD thesis, Department of Statistics, Harvard University.

APPENDIX: OUTLINE OF PROOF OF RESULT 2

To state the approximate variability of $\hat{\Lambda}(t)$, we will use some additional definitions. For $t > 0$, let $D(t) = \{\sum_{g'=0,1} \pi_{g'}(t) p_{g'}\}^{-2}$, and, for $g = 0, 1$, let: $w_{g,\{\pi_0\}}(t) := (-1)^g p_0 p_1 \pi_1(t)\ D(t)$; $w_{g,\{\pi_1\}}(t) := (-1)^{g+1} p_0 p_1 \pi_0(t) D(t)$; and $w_{g,\{p_1\}}(t) := (-1)^{g+1} \pi_1(t) \pi_0(t) D(t)$. In the above, $w_{g,\{\pi_0\}}(t)$ is the quantity obtained when the the weight $w_g(t)$ in (3.2) is regarded as a function of $\pi_0(t)$, $\pi_1(t)$, and p_1, and is differentiated partially with respect to $\pi_0(t)$. The quantities $w_{g,\{\pi_1\}}(t)$ and $w_{g,\{p_1\}}(t)$ have similar interpretation. In an analogous way, we define the functions $\Lambda^{\text{crd}}_{\{\pi_0\}}(t) := \sum_{g=0,1} \int_0^t w_{g,\{\pi_0\}}(u) d\Lambda^{\text{crd}}_g(u)$, $\Lambda^{\text{crd}}_{\{\pi_1\}}(t) := \sum_{g=0,1} \int_0^t w_{g,\{\pi_1\}}(u) d\Lambda^{\text{crd}}_g(u)$, and $\Lambda^{\text{crd}}_{\{p_1\}}(t) := \sum_{g=0,1} \int_0^t w_{g,\{p_1\}}(u) d\Lambda^{\text{crd}}_g(u)$. Using these definitions, we express $W^{(n)}(t) := n^{\frac{1}{2}} \left\{ \hat{\Lambda}(t) - \Lambda^{\text{net}}(t) \right\}$ in a linearized form, in Lemma A.

LEMMA A. *Under Assumption* 1, *and for* $0 < t < t_m$,

$$\begin{aligned} W^{(n)}(t) =& W^{(n)}_{\{p_1\}}(t) + \sum_{g=0,1} W^{(n)}_{\{\pi_g\}}(t) + W^{(n)}_{\{\Lambda_g\}}(t) + O_p(n^{-\frac{1}{2}}), \quad \textit{where} \\ W^{(n)}_{\{p_1\}}(t) &= n^{\frac{1}{2}}(\hat{p}_1 - p_1)\int_0^t d\Lambda^{\text{crd}}_{\{p_1\}}(u), \\ W^{(n)}_{\{\pi_g\}}(t) &= n_g^{-\frac{1}{2}} \sum_{i \in H_g} J_{g,i}(t), \quad J_{g,i}(t) = k_g^{\frac{1}{2}} \int_0^t \{Y_i^*(u) - \pi_g(u)\} d\Lambda^{\text{crd}}_{\{\pi_g\}}(u), \\ W^{(n)}_{\{\Lambda_g\}}(t) &= n_g^{-\frac{1}{2}} \sum_{i \in H_g} K_{g,i}(t), \quad K_{g,i}(t) = k_g^{\frac{1}{2}} \int_0^t \frac{w_g(u)}{\pi_g(u)} dM_{g,i}(u), \\ \textit{and} \qquad & M_{g,i}(t) = N_i^*(t) - \int_0^t Y_i^*(u) d\Lambda^{\text{crd}}_g(u). \end{aligned}$$

Proof. We have the identity

$$W^{(n)}(t) = \sum_{g=0,1} n^{\frac{1}{2}} \left\{ \int_0^t \frac{\hat{w}_g - w_g}{Y_g} dN_g + \int_0^t \frac{w_g}{Y_g} \sum_{i \in H_g} dM_{g,i} \right\}, \tag{A.1}$$

where we have suppressed the variable of integration. The first summands in the right hand side of (A.1) are asymptotically equivalent to $n^{\frac{1}{2}} \int_0^t (\hat{w}_g - w_g) d\Lambda^{\text{crd}}_g$. Moreover, by a Taylor expansion we have

$$n^{\frac{1}{2}}(\hat{w}_g - w_g) = n^{\frac{1}{2}}(\hat{p}_1 - p_1) w_{g,\{p_1\}} + n^{\frac{1}{2}} \sum_{g^*=0,1} (\hat{\pi}_{g^*} - \pi_{g^*}) w_{g,\{\pi_{g^*}\}} + O_p(n^{-\frac{1}{2}}).$$

Substituting this expansion in (A.1), and noting that Y_g/n_g converges uniformly in probability to π_g in $[0, t_m]$ as $n_g \to \infty$, by the Glivenko-Cantelli Theorem (Billingsley, 1979, Theorem 20.6), Lemma A follows after some algebra.

Using Corollary B.1.1 of Fleming and Harrington (1991), in order to show Result 2 it suffices to show two conditions: (i) for any finite set of times $t_1, ..., t_k$, the vector $(W^{(n)}(t_1), ..., W^{(n)}(t_k))$ converges in distribution to $(W(t_1), ..., W(t_k))$, where W is a Gaussian process with moments as in Result 2, and (ii) the process $W^{(n)}$ is “tight” in the sense of Fleming and Harrington (1991, p. 340).

The large-sample normality in condition (i) follows from applying the central limit theorem and Theorem 2.4.4 of Fleming and Harrington (1991) on Lemma A. This application also shows that, for fixed times $0 < s, t < E_{max}$, the covariance cov$\{W(s), W(t)\}$ is given by,

$$V_{\{p_1\}}(s,t) + \sum_{g=0,1} V_{\{\pi_g\}}(s,t) + V_{\{\Lambda_g\}}(t) + V_{\{\pi_g,\Lambda_g\}}(s,t) + V_{\{\pi_g,\Lambda_g\}}(t,s), \qquad \text{(A.2)}$$

where

$$V_{\{p_1\}}(s,t) := p_1 p_0 \int_0^t d\Lambda^{\text{crd}}_{\{p_1\}}(u) \int_0^s d\Lambda^{\text{crd}}_{\{p_1\}}(u^*),$$

and, for $k_0 = 1/\{p_0 p^{(S)}\}$, $k_1 = 1/p_1$, and $g = 0, 1$,

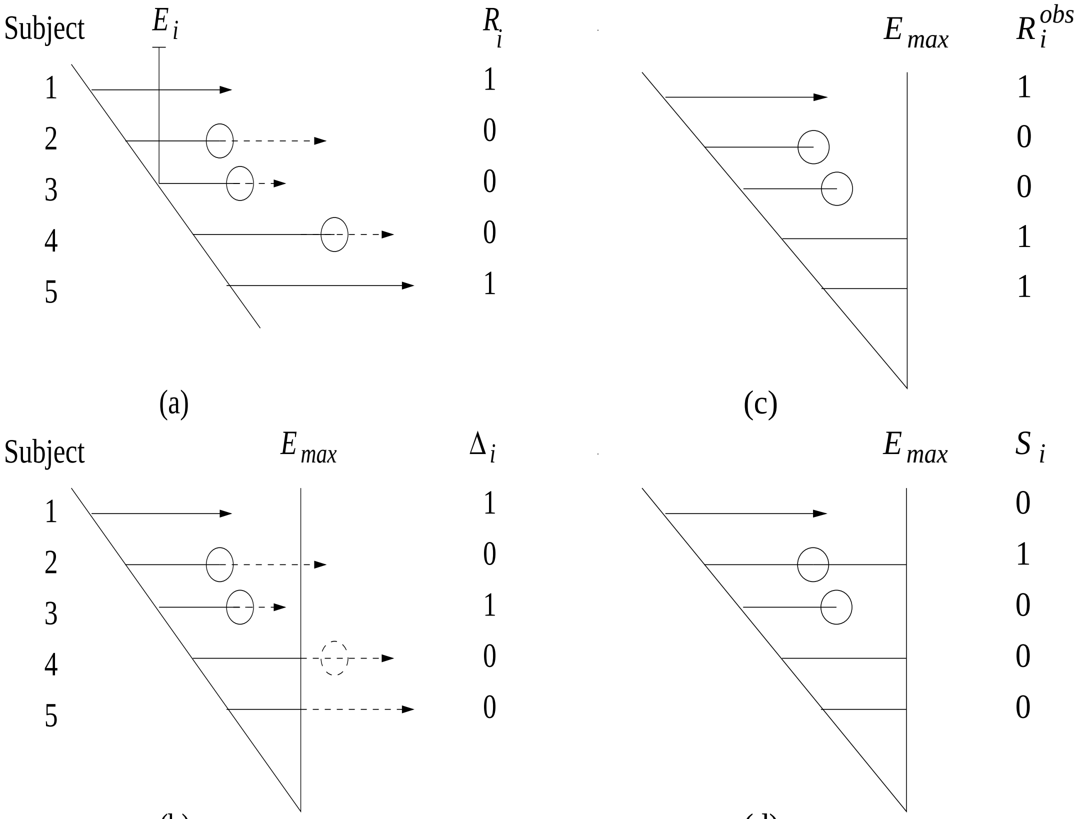
$$V_{\{\pi_g\}}(s,t) := k_g \int_0^t \int_0^s \left[\pi_g\{\max(u,u^*)\} - \pi_g(u)\pi_g(u^*)\right] d\Lambda^{\text{crd}}_{\{\pi_g\}}(u^*) d\Lambda^{\text{crd}}_{\{\pi_g\}}(u),$$

$$V_{\{\Lambda_g\}}(s,t) := k_g \int_0^{\min(s,t)} \frac{\{w_g(u)\}^2}{\pi_g(u)} d\Lambda^{\text{crd}}_g(u), \qquad \text{and}$$

$$V_{\{\pi_g,\Lambda_g\}}(s,t) := -k_g \int_0^{\min(s,t)} \int_u^s \frac{\pi_g(u^*)}{\pi_g(u)} w_g(u) d\Lambda^{\text{crd}}_{\{\pi_g\}}(u^*) d\Lambda^{\text{crd}}_g(u).$$

The proof of condition (ii) is not difficult but is more laborious and is omitted here. Further details can be found in Frangakis (1999, appendix of Part 3).

Figure 1. Potential outcomes and observed data with double-sampling. For each part, solid lines represent observed information, dashed lines represent unobserved information: (a) survival times, (arrows), and true dropouts (circles); (b) administrative censoring with true dropout; (c) study-dropouts; (d) subject 2 is double-sampled.


Subject
E_i
R_i
1
2
3
4
5
1
0
0
0
1
E_{max}
R_i^{obs}
1
0
0
1
1
Subject
E_{max}
Δ_i
1
2
3
4
5
1
0
1
0
0
E_{max}
S_i
0
1
0
0
0

(a) (c)

(b) (d)

Table 1

Projected Tharies prosthesis study of Sec. 4: sensitivity inference for the prostheses survival probabilities $S(t)$ at $t = 6, 7$ and 8 years from primary surgery; coverage of nominal 95% confidence intervals and mean squared error.

δ_R	-1				0				1			
r^*	-0.4	-0.2	0.0	0.2	-0.4	-0.2	0.0	0.2	-0.4	-0.2	0.0	0.2
r	0.16	0.34	0.46	0.60	0.34	0.45	0.57	0.67	0.55	0.63	0.68	0.72
$p^{mis}_{C>8}$ (%)	40.3	38.8	36.7	34.6	43.0	42.2	41.4	40.5	45.8	44.9	44.2	44.0
p^{mis} (%)	19.7	19.4	18.3	16.4	23.1	23.2	22.5	21.5	26.5	26.8	27.3	27.0
$S(6)$												
Coverage (%)												
SKM	93.1	94.4	93.9	95.4	86.4	85.7	84.3	82.8	36.2	36.8	40.5	38.0
IML	94.2	95.5	94.4	94.6	94.2	95.3	94.3	93.8	94.0	94.1	94.1	94.3
MSE $(\times 10^3)$												
SKM	1.10	0.97	0.96	0.87	1.35	1.32	1.37	1.46	4.35	4.35	4.17	4.13
IML	0.91	0.84	0.92	0.88	1.06	0.99	1.03	1.03	1.10	1.05	1.11	1.05
$S(7)$												
Coverage (%)												
SKM	92.0	93.3	93.9	94.9	86.2	85.2	84.2	81.5	33.8	34.7	35.4	35.7
IML	94.5	95.6	93.5	95.0	95.3	94.8	95.6	94.1	94.4	94.0	93.9	94.5
MSE $(\times 10^3)$												
SKM	1.65	1.43	1.33	1.25	1.84	1.91	1.98	2.13	6.34	6.20	6.20	6.07
IML	1.28	1.19	1.28	1.25	1.34	1.37	1.36	1.40	1.38	1.44	1.43	1.38
$S(8)$												
Coverage (%)												
SKM	90.3	92.4	93.5	94.1	87.6	85.8	85.2	83.4	44.6	45.5	47.5	46.8
IML	94.1	94.9	94.2	94.8	94.8	94.6	94.4	94.5	94.2	94.2	94.2	94.7
MSE $(\times 10^3)$												
SKM	2.78	2.36	2.08	1.94	2.43	2.48	2.58	2.73	7.46	7.35	7.17	7.07
IML	1.83	1.69	1.77	1.70	1.90	1.90	1.89	1.85	1.94	1.94	1.85	1.83

$p^{mis}_{C>8} = \Pr(R^{obs}_i = 0|C_i > 8)$; $p^{mis} = \Pr(R^{obs}_i = 0)$; only positive values of $r = \text{corr}(L_i, T_i|R_i = 0)$ are considered because $L_i < T_i$ for true dropouts.

SKM, stratified Kaplan-Meier method; IML, method of Section 3.2.