Improved Doubly Robust Estimation when Data are Monotonely Coarsened, with Application to Longitudinal Studies with Dropout

Anastasios A. Tsiatis[1,*], **Marie Davidian**[1], and **Weihua Cao**[2]

[1] Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A

[2] Division of Biostatistics, Office of Surveillance and Biometrics, U.S. Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, Maryland, 20993, U.S.A

Summary

A routine challenge is that of making inference on parameters in a statistical model of interest from longitudinal data subject to drop out, which are a special case of the more general setting of monotonely coarsened data. Considerable recent attention has focused on doubly robust estimators, which in this context involve positing models for both the missingness (more generally, coarsening) mechanism and aspects of the distribution of the full data, that have the appealing property of yielding consistent inferences if only one of these models is correctly specified. Doubly robust estimators have been criticized for potentially disastrous performance when both of these models are even only mildly misspecified. We propose a doubly robust estimator applicable in general monotone coarsening problems that achieves comparable or improved performance relative to existing doubly robust methods, which we demonstrate via simulation studies and by application to data from an AIDS clinical trial.

Keywords

Coarsening at random; Discrete hazard; Dropout; Longitudinal data; Missing at random

1. Introduction

Studies in which data are to be collected longitudinally according to a pre-determined schedule are often complicated by dropout, where some subjects leave the study prematurely and do not return, so that the intended data from the point of dropout onward are missing. Ordinarily, interest focuses on questions that can be formalized within a statistical model describing aspects of the distribution of the full data, the data that would have been collected on a subject had dropout not occurred. Failure to take dropout into account in analyses based on the observed data, which are curtailed due to dropout for some participants, can lead to biased inferences on full data model parameters, and a vast literature exists on methods for making valid inferences based on the observed data under different assumptions regarding

the dropout mechanism; e.g., Hogan, Roy, and Korkontzelou (2004), Philipson, Ho, and Henderson (2008), and Molenberghs and Fitzmaurice (2009) and the references therein.

As a running example, we consider data from AIDS Clinical Trials Group (ACTG) Protocol 175 (Hammer et al., 1996), where subjects infected with human immunodeficiency virus (HIV) were randomized to four antiretroviral regimens: zidovudine (ZDV), ZDV +didanosine (ZDV+ddI), ZDV+zalcitabine (ZDV+ddC), and didanosine (ddI). On each, CD4 T-cell count (cells/mm^3 blood), a measure of immunologic status, was measured at baseline and, ideally, at 20±5, 40±5, 60±5, and 96±5 weeks post-baseline, along with several baseline covariates. As the latter three regimens showed no differences, we focus on estimating mean CD4 count at 96±5 weeks for the population of subjects assigned to any of the three. Of 1838 such participants, 12%, 30%, 38%, and 49% had dropped out by each of the visit times, respectively. Clearly, the substantial dropout complicates inference on the population mean.

Missingness due to dropout in a longitudinal study is a special case of monotone coarsening. Under coarsening, for each subject, one of a set of $M + 1$ many-to-one functions of the full data, indexed by $r = 1, \ldots, M, \infty$, is observed (Heitjan and Rubin, 1991; Gill, van der Laan, and Robins, 1997; Tsiatis, 2006). With monotone coarsening, the many-to-one function for any $r = 1, \ldots, M$ is itself a many-to-one function of the $(r + 1)$th function, so that $r = 1$ corresponds to the "most coarsened" data and $r = M$ to the least, and ∞ denotes no coarsening (the full data are observed). Monotone dropout in a longitudinal study fits into this framework, with r indexing $M+1$ planned data collection times, where $r = 1$ corresponds to baseline. Here, the coarsened data at level r are the data that would be observed on a subject who is present for the rth visit and then drops out prior to the $(r + 1)$th visit.

Analogous to the notion of missing at random (MAR), the mechanism leading to coarsening is coarsening at random (CAR, Heitjan and Rubin, 1991) if, for each r, the probability that, given the full data, the data are coarsened at level r depends only on the coarsened data (so not on data not observed at level r). Whether or not the CAR assumption is reasonable must of course be critically evaluated by the analyst; when it is plausible, a number of approaches have been proposed for making inference on full data model parameters based on the observed, coarsened data. These include likelihood methods, where a parametric model for the entire full data distribution may be posited, from which the likelihood based on the coarsened data can be deduced without the need to specify the coarsening mechanism (e.g., Birmingham, Rotnitzky, and Fitzmaurice, 2003; Little, 2009). These methods will yield valid inferences as long as the posited full data model is correct, but can lead to bias otherwise. In contrast, inverse probability weighted methods (IPW) (Robins, Rotnitzky, and Zhao, 1994, 1995; Rotnitzky, Robins, and Scharfstein, 1998; Rotnitzky, 2009) require specification of models for the coarsening probabilities, and the resulting estimators are consistent only if these models are correct and can be unstable in practice if some probabilities of observing the full data are close to zero, leading to large inverse weights. Robins et al. (1994) identified a class of "augmented" IPW (AIPW) estimators that, in the present context, involve (parametric) modeling of both the coarsening probabilities and the conditional expectations of certain functions of the full data given the coarsened data for each level of coarsening. The efficient member of this class, with smallest asymptotic variance, is obtained when both sets of models are correctly specified. Scharfstein, Rotnitzky, and Robins (1999) noted that estimators in this class are consistent even if one of the sets of models (but not both) is misspecified. Such estimators are referred to as "doubly robust" (DR) and have been advocated owing to the protection this feature affords (Bang and Robins, 2005). Bang and Robins (2005) described a DR estimator in the case of a longitudinal study with dropout and provided simulation evidence demonstrating the DR property; see also Seaman and Copas (2009).

Despite their obvious appeal, DR estimators have been vigorously criticized. Kang and Schafer (2007) presented simulations in the simple situation of estimation of a population mean from an iid sample with MAR response showing that the usual DR estimator can exhibit severe bias when both sets of models are only "slightly" misspecified and/or when some probabilities of observing full data are close to zero and argued against use of DR estimators. In this setting, however, Tan (2006, 2007, 2008) and Cao, Tsiatis, and Davidian (2009) showed how to construct DR estimators that do not have these shortcomings; see also Goetgeluk, Vansteelandt, and Goetghebeur (2009). Cao et al. (2009) set out expressly to identify the "best" DR estimator, that with smallest asymptotic variance if the coarsening probabilities are correctly specified regardless of whether or not the conditional expectation models are, and demonstrated that these estimators are relatively more efficient and exhibit superior robustness to slight modeling mishaps relative to other DR estimators.

In this paper, we extend these ideas to the general setting of monotonely coarsened data. In Section 2, we introduce notation and formalize the CAR assumption. We state the inferential objectives and describe the general form of DR estimators in Section 3, and in Section 4 propose an improved DR estimator, which we specialize to the case of a longitudinal study with dropout through application to the ACTG 175 data in Section 5. Simulations presented in Section 6 exhibit the improved performance of the proposed methods.

2. General Coarsened Data Framework and Coarsening at Random

We follow Tsiatis (2006, Section 7.1). Denote the full data by Z; ideally, then, the data intended to be collected are realizations of independent and identically distributed (iid) Z_1, …, Z_n. Let C be a discrete coarsening variable with possible values 1, …, M, ∞ corresponding to $M + 1$ levels of coarsening. When $C = r$, $r = 1, …, M$, we observe $G_r(Z)$, a many-to-one function of Z. When $C = \infty$, we observe $G_\infty(Z) = Z$; i.e., there is no coarsening, and the full data are observed. Under monotone coarsening, $G_r(Z)$ is a many-to-one function of $G_{r+1}(Z)$; i.e., $G_r(Z) = f_r\{G_{r+1}(Z)\}$, $r = 1, …, M$, where $G_{M+1}(Z) = G_\infty(Z)$. Thus, $G_1(Z)$ are the most coarsened data, $G_2(Z)$ are less so, and so forth, up to $G_\infty(Z) = Z$, where there is no coarsening. The observed data are realizations of iid $\{C_i, G_{C_i}(Z_i)\}$, $i = 1$, …, n.

As is customary in general missing data problems, we assume that there is a positive probability of observing the full data; i.e., we make the positivity assumption $P(C = \infty|Z) \geq \varepsilon > 0$ almost everywhere. The CAR assumption may be expressed as

$$P(C=r|Z)=\pi\{r, G_r(Z)\}, \quad r=1, \dots, m, \infty; \tag{1}$$

i.e., the probability of coarsening at level r depends on the full data Z only as a function $\pi\{r, G_r(Z)\}$ of the observed data $G_r(Z)$. As $G_\infty(Z) = Z$, write $\pi\{\infty, G_\infty(Z)\} = \pi(\infty, Z)$.

We now demonstrate how data from a longitudinal study with dropout fit into this framework, where we use notation popularized by Robins and colleagues (e.g., Bang and Robins, 2005). Let L_j be the vector of information collected at visit time t_j, $j = 1, …, M +1$. Let R be a dropout indicator such that, if $R = j$, the subject is last seen at the jth visit, and the observed data are $\bar{L}_j = (L_1, …, L_j)$, $j = 1, …, M + 1$; write $\bar{L} = \bar{L}_{M+1}$. In the coarsened data framework, the full data Z are thus $Z = G_\infty(Z) = \bar{L}$; the coarsening indicator C corresponds to R, $C = 1, …, M, \infty$, where $C = \infty$ is the same as $R = M + 1$; and the coarsened data $G_r(Z) = \bar{L}_r$, $r = 1, …, M$. The observed data from a sample of size n are then iid $(R_i, \bar{L}_{R_i})$, $i = 1, …, n$. Thus, in ACTG 175, $M = 4$, $t_1 = 0$ (baseline); $(t_2, t_3, t_4, t_5) = (20, 40, 60, 96) \pm 5$ weeks; $L_1 = (X, Y_1)$, say, where X are baseline covariates and Y_1 is baseline CD4 count; and $L_j = Y_j$,

$j = 2, \ldots, M + 1 = 5$, where Y_j is CD4 count at t_j. The positivity and CAR assumptions (1) become $P(R = M + 1|\bar{L}) \geq \varepsilon > 0$ almost everywhere and $P(R = j|\bar{L}) = \pi(j, \bar{L}_j), j = 1, \ldots, M, M + 1$, respectively.

3. Inferential Objective and Doubly Robust Estimators

We suppose that the analyst has specified a semiparametric model for the full data Z corresponding to density $p_Z(z; \beta, \eta)$, say, where β ($p \times 1$) is the finite dimensional parameter of interest; here, then, $p_Z(z; \beta, \eta)$ embodies the features of the full data that the analyst is willing to assume. The goal is to estimate β based on the sample of observed, monotonely coarsened data. Ordinarily, η is an infinite dimensional nuisance parameter representing aspects of the full data distribution about which nothing is assumed. If η were finite dimensional, $p_Z(z; \beta, \eta)$ is a fully parametric model, in which case inference on β based on the observed data under the CAR assumption could be carried out via maximum likelihood (ML) techniques. We consider estimators for β calculable under a more general semiparametric model.

For ACTG 175, with $Y = Y_5 =$ CD4 count at 96 ± 5 weeks, $\beta = E(Y)$. With no further assumptions, $p_Z(z; \beta, \eta)$ is completely nonparametric except for the restriction of finite β; if full data were available, the obvious estimator is the sample mean at 96 ± 5 weeks. If instead one assumed $E(Y_j|X) = \beta_0 + \beta_1 t_j + \beta_x^T X$ and interest focused on the "slope" β_1, $\beta = (\beta_0, \beta_1, \beta_x^T)^T$, then $p_Z(z; \beta, \eta)$ would be the semiparametric model imposing this conditional (on X) mean structure, with all other aspects of the full data distribution unspecified. With full data, β could be estimated by solving a set of generalized estimating equations (GEEs).

In general, we assume that estimators for β exist based on the full data, defined by ($p \times 1$) unbiased estimating functions $m(Z, \beta)$; i.e., such that $E\{m(Z, \beta)\} = 0$ for all β (or at least for β in a neighborhood of β_0, the true value). An estimator would solve $\sum_{i=1}^{n} m(Z_i, \beta) = 0$ and, under regularity conditions, would be consistent and asymptotically normal by standard M-estimator theory (Stefanski and Boos, 2002). For the sample mean at 96 ± 5 weeks in ACTG 175, $m(Z, \beta) = Y - \beta$; for the slope parameter, $m(Z, \beta)$ would be a GEE estimating function, perhaps involving nuisance parameters in a "working" correlation structure.

We start with the premise that the analyst has fully specified the coarsening probabilities (1), so that they involve no unknown parameters, a requirement we relax in Section 4. For general monotonely coarsened data, the theory of Robins et al. (1994) implies that, under CAR, if the coarsening probabilities $\pi\{r, G_r(Z)\}$ are correctly specified, members of the class of all regular, asymptotically linear estimators (Tsiatis, 2006, Chapter 3) for β using the observed data solve estimating equations based on an AIPW estimating function of form

$$\frac{I(C = \infty)m(Z, \beta)}{\pi(\infty, Z)} + \sum_{r=1}^{M} \frac{dM_r\{G_r(Z)\}}{K_r\{G_r(Z)\}} \mathcal{L}_r\{G_r(Z)\}$$

(2)

(Tsiatis, 2006, Chapter 10) where $\pi\{r, G_r(Z)\}$ is as in (1); the coarsening discrete hazard $\lambda_r \{G_r(Z)\} = P(C = r|C \geq r, Z)$ and survival function $K_r\{G_r(Z)\} = P(C > r|Z)$ and $\mathcal{L}_r \{G_r(Z)\}$ are arbitrary functions of $G_r(Z)$; and $dM_r\{G_r(Z)\} = I(C = r) - \lambda_r\{G_r(Z)\} I(C \geq r)$, all for $r = 1, \ldots, M$ (Tsiatis, 2006, Theorem 9.2). Under the CAR assumption,

$K_r\{G_r(Z)\} = 1 - \sum_{j=1}^{r} \pi\{j, G_j(Z)\}$, $r = 1, \ldots, M$; $\lambda_1 \{G_1(Z)\} = \pi \{1, G_1(Z)\}$; and, for $r = 2, \ldots,$

M, $\lambda_r\{G_r(Z)\}=\pi\{r, G_r(Z)\}/\left[1-\sum_{j=1}^{r-1}\pi\{j, G_j(Z)\}\right]$. If $\mathcal{L}_r\{G_r(Z)\}=0$, $r=1, \ldots, M$, (2) reduces to the simple IPW estimating function depending only on data from the "complete cases," subjects for whom full data are observed; thus, the second "augmentation" term in (2) seeks to improve efficiency of estimation of β by exploiting information from all subjects.

When the $\pi\{r, G_r(Z)\}$ are correct, estimators for β solving estimating equations based on (2) will be consistent and asymptotically normal regardless of the choice of $\mathcal{L}_r\{G_r(Z)\}$, $r=1, \ldots, M$, and the optimal choice yielding the estimator for β with smallest asymptotic variance is $E\{m(Z, \beta_0)|G_r(Z)\}$ (Tsiatis, 2006, Sections 9.1, 9.2, 10.3). As these conditional expectations may be unknown, it is natural to model them via functions $h_r\{G_r(Z), \xi\}$, $r=1, \ldots, M$, where ξ is a finite dimensional parameter, and estimate β by solving

$$\sum_{i=1}^{n}\left[\frac{I(C_i=\infty)m(Z_i, \beta)}{\pi(\infty, Z_i)}+\sum_{r=1}^{M}\frac{dM_r\{G_r(Z_i)\}}{K_r\{G_r(Z_i)\}}h_r\{G_r(Z_i), \widehat{\xi}\}\right]=0,$$

(3)

where $\hat{\xi}$ is some estimator for ξ. The method of estimating ξ is key; see Section 4.

Writing $m(Z)=m(Z, \beta_0)$, if the coarsening probabilities are correctly specified and $\hat{\xi}$ converges in probability to some ξ^*, say, an estimator for β solving (3) will be consistent and asymptotically normal regardless of whether or not $h_r\{G_r(Z), \xi^*\}=E\{m(Z)|G_r(Z)\}$, $r=1, \ldots, M$. Moreover, the form of the asymptotic variance of the estimator is identical whether $\hat{\xi}$ or the value ξ^* is substituted in (3) (Tsiatis, 2006, Theorem 10.3), so that the asymptotic variance does not depend on the sampling variation of $\hat{\xi}$ but depends only on its limit in probability ξ^*. If $h_r\{G_r(Z), \xi^*\}=E\{m(Z)|G_r(Z)\}$, $r=1, \ldots, M$, does hold, then the estimator for β will be optimal (have smallest asymptotic variance) among all such estimators. If $h_r\{G_r(Z), \xi^*\}=E\{m(Z)|G_r(Z)\}$, $r=1, \ldots, M$, but the coarsening probabilities are misspecified, the estimator for β will still be consistent. Accordingly, such estimators are doubly robust, as only one set of models need be correct to ensure consistency.

In the context of a longitudinal study with dropout, Bang and Robins (2005) described estimators for β that are solutions to (3); we present details in the next section.

4. Existing and Proposed Doubly Robust Estimators

We continue to assume that the coarsening probabilities $\pi\{r, G_r(Z)\}$ are fully specified, which we relax shortly. Different methods for estimating ξ in the models $h_r\{G_r(Z), \xi\}$ will lead to different estimators for β solving (3). Bang and Robins (2005) advocate one such method, described later in this section. We seek to define an estimator $\hat{\xi}_{opt}$ for ξ in the spirit of Cao et al. (2009); i.e., that (i) is "optimal" when the $\pi\{r, G_r(Z)\}$, $r=1, \ldots, M, \infty$, are correctly specified, even if the $h_r\{G_z(Z)\}$, $r=1, \ldots, M$, are not, in the sense of yielding an estimator $\hat{\beta}_{opt}$ solving (3) with smallest asymptotic variance; and (ii) $\hat{\beta}_{opt}$ is doubly robust. Moreover, $\hat{\xi}_{opt}$ requires no further assumptions beyond specification of the models $h_r\{G_r(Z), \xi\}$.

Denote the true coarsening probabilities as $\pi_0\{r, G_r(Z)\}$, and define the true discrete hazards and survival functions as $\lambda_{r0}\{G_r(Z)\}$ and $K_{r0}\{G_r(Z)\}$, where $K_M\{G_M(Z)\}=\pi(\infty, Z)$ and $K_{M0}\{G_M(Z)\}=\pi_0(\infty, Z)$; and write $dM_{r0}\{G_r(Z)\}$ when $\lambda_{r0}\{G_r(Z)\}$ is substituted for $\lambda_r\{G_r(Z)\}$ in $dM_r\{G_r(Z)\}$. With the coarsening probabilities correct, whether or not the $h_r\{G_r(Z), \xi\}$ are correct, it is straightforward to deduce that minimizing the variance of an estimator for β solving (3) involves minimizing in ξ^*

$$E\left[\frac{I(C=\infty)m(Z)}{\pi_0(\infty,Z)}+\sum_{r=1}^{M}\frac{dM_{r0}\{G_r(Z)\}}{K_{r0}\{G_r(Z)\}}h_r\{G_r(Z),\xi^*\}\right]^2,$$

(4)

where ξ^* is the value to which the estimator $\hat{\xi}$ used converges in probability. Denote this minimizing value by ξ^{opt}. If the models $h_r\{G_r(Z),\xi\}$ are correctly specified, so that there is some ξ_0 such that $h_r\{G_r(Z),\xi_0\}=E\{m(Z)|G_r(Z)\}$, $r=1,\ldots,M$, then in fact $\xi^{opt}=\xi_0$; if not, such a ξ^{opt} still exists. Accordingly, to satisfy (i), we require that the desired $\hat{\xi}_{opt}$ converge in probability to ξ^{opt}. To ensure (ii), when the $h_r\{G_r(Z),\xi\}$ are correctly specified but the coarsening probabilities may not be, $\hat{\xi}_{opt}$ must converge in probability to ξ_0.

From (4), ξ^{opt} must satisfy

$$E\left(\left[\sum_{r=1}^{M}\frac{dM_{r0}\{G_r(Z)\}}{K_{r0}\{G_r(Z)\}}h_{r\xi}\{G_r(Z),\xi\}\right]\left[\frac{I(C=\infty)m(Z)}{\pi(\infty,Z)}+\sum_{r=1}^{M}\frac{dM_{r0}\{G_r(Z)\}}{K_{r0}\{G_r(Z)\}}h_r\{G_r(Z),\xi\}\right]\right)=0,$$

where $h_{r\xi}\{G_r(Z),\xi\}$ is the column vector of partial derivatives of $h_r\{G_r(Z)\}$ with respect to ξ. Using Lemmas 10.1–10.3 of Tsiatis (2006), this expression can be written as

$$E\left[-m(Z)\sum_{r=1}^{M}\frac{\lambda_{r0}\{G_r(Z)\}}{K_{r0}\{G_r(Z)\}}h_{r\xi}\{G_r(Z),\xi\}+\sum_{r=1}^{M}\frac{\lambda_{r0}\{G_r(Z)\}}{K_{r0}\{G_r(Z)\}}h_{r\xi}\{G_r(Z),\xi\}h_r\{G_r(Z),\xi\}\right]=0$$

(5)

We now derive an estimator $\hat{\xi}_{opt}$ for ξ that converges to ξ^{opt} satisfying (5) when the coarsening probabilities are correctly specified but the models $h_r\{G_z(Z),\xi\}$ may not be and that converges to ξ_0 in the converse situation. We propose estimating ξ by solving estimating equations corresponding to the estimating function

$$\sum_{r=1}^{M}I(C>r)q_r\{G_r(Z),\xi\}[h_{r+1}\{G_{r+1}(Z),\xi\}-h_r\{G_r(Z),\xi\}],$$

(6)

where $q_r\{G_r(Z),\xi\}$ is a vector of functions with dimension equal to that of ξ, $I(C>M)=I(C=\infty)$, and $h_{M+1}\{G_{M+1}(Z),\xi\}=m(Z)$; note that (6) is a function of the observed data $\{C, G_C(Z)\}$. We show in Web Appendix A that (6) is an unbiased estimating function for ξ when the coarsening probabilities may not be correctly specified but the models $h_r\{G_z(Z),\xi\}$ are and that estimators for ξ based on (6) converge in probability to ξ_0 for arbitrary choice of the $q_r\{G_r(Z),\xi\}$. Thus, the proposed estimator $\hat{\xi}_{opt}$, which involves a particular choice of these functions, converges in probability to ξ_0 under these conditions, as required for (ii).

We propose the estimator $\hat{\xi}_{opt}$ found by taking

$$q_r\{G_r(Z),\xi\}=-[K_r\{G_r(Z)\}]^{-1}\sum_{j=1}^{r}\frac{\lambda_j\{G_j(Z)\}}{K_j\{G_j(Z)\}}h_{j\xi}\{G_j(Z),\xi\},\quad r=1,\ldots,M.$$

(7)

With (7) substituted, it may be shown (see Web Appendix B) that (6) has expectation zero at $\xi = \xi^{opt}$, where ξ^{opt} solves (5), when the coarsening probabilities are correctly specified but the functions $h_r\{G_r(Z), \xi\}$ may or may not be and hence is an unbiased estimating function under these conditions, so that $\hat{\xi}_{opt}$ converges in probability to ξ^{opt}, ensuring (i).

Summarizing, the proposed estimator $\hat{\beta}_{opt}$, found by using (7) in the estimating function (6) for ξ to obtain $\hat{\xi}_{opt}$, will be doubly robust and achieve smallest asymptotic variance among estimators in class (3) when the coarsening probabilities are correctly specified but the $h_r\{G_r(Z), \xi\}$ may not be. Bang and Robins (2005) proposed an alternative approach, which is effectively equivalent to modeling $E\{m(Z)|G_r(Z)\}$, $r = 1, \ldots, M$, by functions $h_r^*\{G_r(Z), \xi_r\}$ corresponding to a generalized linear model with canonical link, where the parameter ξ_r is specific to the rth level of coarsening; and $q_r\{G_r(Z), \xi_r\}$ analogous to those in (6) for each r are dictated by the gradient and variance function of the generalized linear model. The resulting estimator for β is doubly robust, but, as it does not exploit the "optimal" choice (7) in estimation of the ξ_r, it will not in general achieve the minimum asymptotic variance when the coarsening probabilities are correct unless the $h_r^*\{G_r(Z), \xi_r\}$ are also correct.

For either estimator, in order to be feasible models for $E\{m(Z)|G_r(Z)\}$, $r = 1, \ldots, M$, the $h_r\{G_r(Z), \xi\}$ must satisfy $E[h_{r+1}\{G_{r+1}(Z), \xi\}|G_r(Z)] = h_r\{G_r(Z), \xi\}$, and similarly for the $h_r^*\{G_r(Z), \xi_r\}$. Accordingly, as demonstrated in Sections 5 and 6, the analyst may specify a model for the part of the joint distribution of Z that allows one to identify the conditional expectations $E\{m(Z)|G_r(Z)\}$ so that they meet this requirement.

The coarsening probabilities are unlikely to be known except in a study where coarsening is by design. Thus, it is natural to postulate and fit parametric models for the coarsening mechanism (Tsiatis, 2006, Section 8.2); e.g., model the discrete hazards $\lambda_r\{G_r(Z)\}$, $r = 1, \ldots, M$, in terms of a finite dimensional parameter ψ via logistic regression and write $\lambda_r\{G_r(Z), \psi\}$, and estimate ψ by ML. This implies corresponding models $\pi\{r, G_r(Z), \psi\}$ and $K_r\{G_r(Z), \psi\}$. Letting $\lambda_{r\psi}\{G_r(Z), \psi\}$ be the column vector of partial derivatives of $\lambda_r\{G_r(Z_i), \psi\}$ with respect to ψ, we show in Web Appendix C that the score vector for ψ is

$$S_\psi\{C, G_C(Z), \psi\} = \sum_{r=1}^{M} dM_r\{G_r(Z), \psi\} K_{r-1}\{G_r(Z), \psi\} \lambda_{r\psi}\{G_r(Z), \psi\} / [K_r\{G_r(Z), \psi\} \lambda_r\{G_r(Z), \psi\}]$$
.

As detailed in Tsiatis (2006, Chapters 8–10), there is an effect on the asymptotic distribution of an estimator for β solving (3) when the coarsening probabilities are modeled and ψ is estimated by the maximum likelihood estimator (MLE) $\hat{\psi}$. In particular, it follows from Theorem 9.1 of Tsiatis (2006) that, when the models for the coarsening probabilities are correctly specified, so that there exists ψ_0 such that $\lambda_r\{G_r(Z), \psi_0\} = \lambda_{r0}\{G_r(Z)\}$, the estimator for β that solves the estimating equation

$$\sum_{i=1}^{n} \left[\frac{I(C_i = \infty) m(Z_i, \beta)}{\pi_0(\infty, Z_i, \hat{\psi})} + \sum_{r=1}^{M} \frac{dM_r\{G_r(Z_i), \hat{\psi}\}}{K_r\{G_r(Z_i, \hat{\psi})\}} h_r\{G_r(Z_i), \hat{\xi}\} \right] = 0$$

(8)

for some $\hat{\xi}$ is asymptotically equivalent to that solving

$$\sum_{i=1}^{n}\left(\frac{I(C_i=\infty)m(Z_i,\beta)}{\pi(\infty,Z_i,\psi_0)}+\sum_{r=1}^{M}\left[\frac{dM_r\{G_r(Z_i),\psi_0\}}{K_r\{G_r(Z_i),\psi_0\}}h_r\{G_r(Z_i),\xi^*\}-\theta_{proj}^{T}S_{\psi}\{C_i,G_c(Z_i),\psi_0\}\right]\right)$$

$$=\sum_{i=1}^{n}\left(\frac{I(C_i=\infty)m(Z_i,\beta)}{\pi_0(\infty,Z_i,\psi_0)}+\sum_{r=1}^{M}\frac{dM_r\{G_r(Z_i),\psi_0\}}{K_r\{G_r(Z_i),\psi_0\}}\times\left[h_r\{G_r(Z_i),\xi^*\}-\theta_{proj}^{T}\frac{K_{r-1}\{G_r(Z_i),\psi_0\}\lambda_{r\psi}\{G_r(Z_i),\psi_0\}}{\lambda_r\{G_r(Z_i),\psi_0\}}\right]\right)=0. \tag{9}$$

Here, ξ^* is the limit in probability of $\hat{\xi}$, and θ_{proj} is the value of θ that minimizes

$$E\left[\frac{I(C=\infty)m(Z)}{\pi(\infty,Z_i,\psi_0)}+\sum_{r=1}^{M}\frac{dM_r\{G_r(Z_i),\psi_0\}}{K_r\{G_r(Z,\psi_0\}}\tilde{h}_r\{G_r(Z_i),\tilde{\xi}\}\right]^2,\quad \tilde{\xi}=(\xi^T,\theta^T)^T, \tag{10}$$

when ξ^* is substituted for ξ, and

$$\tilde{h}_r\{G_r(Z),\tilde{\xi}\}=h_r\{G_r(Z),\xi\}-\theta^T\frac{K_{r-1}\{G_r(Z),\psi_0\}\lambda_{r\psi}\{G_r(Z_i),\psi_0\}}{\lambda_r\{G_r(Z),\psi_0\}}. \tag{11}$$

Referring to (4), which defines ξ^{opt} assuming ψ_0 is known, and examining (9) suggests that the "optimal" estimator for ξ for estimators for β in the class given by (8) should converge to the value ξ^{opt*} that minimizes (10) simultaneously in ξ^* and θ. Identifying $\tilde{h}_r\{G_r(Z),\tilde{\xi}\}$ in (10) with $h_r\{G_r(Z),\xi\}$ in (4) shows that finding ξ^{opt} minimizing (4) is analogous to finding the optimal $\tilde{\xi}$, and hence ξ^{opt*}, minimizing (10). We can use this correspondence to propose an approach to estimating $\tilde{\xi}$ that will lead to $\hat{\tilde{\xi}}_{opt*}$, say, such that (i) using $\hat{\tilde{\xi}}_{opt*}$ in (8) yields the estimator $\hat{\beta}_{opt*}$ with smallest asymptotic variance among estimators solving (8) when the coarsening probabilities are correctly modeled, and (ii) $\hat{\beta}_{opt*}$ is doubly robust.

As, in practice, ψ_0 is unknown, write $\tilde{h}_r\{G_r(Z),\tilde{\xi},\psi\}$ to denote (11) treating ψ in the coarsening model as a free parameter. Analogous to (16) of Cao et al. (2009), we propose estimating $\tilde{\xi}$ by solving estimating equations corresponding to the estimating function

$$\sum_{r=1}^{M}I(C>r)\tilde{q}_r\{G_r(Z),\tilde{\xi},\hat{\psi}\}[\tilde{h}_{r+1}\{G_{r+1}(Z),\tilde{\xi},\hat{\psi}\}-\tilde{h}_r\{G_r(Z),\tilde{\xi},\hat{\psi}\}], \tag{12}$$

where $\tilde{q}_r\{G_r(Z),\tilde{\xi},\psi\}$ is the extension of (7), namely,

$$\tilde{q}_r\{G_r(Z),\tilde{\xi},\psi\}=-[K_r\{G_r(Z),\psi\}]^{-1}\sum_{j=1}^{r}\frac{\lambda_j\{G_j(Z),\psi\}}{K_j\{G_j(Z),\psi\}}\left[\begin{array}{c}\tilde{h}_{j\xi}\{G_j(Z),\tilde{\xi},\psi\}\\ \tilde{h}_{j\theta}\{G_j(Z),\tilde{\xi},\psi\}\end{array}\right]; \tag{13}$$

and $\tilde{h}_{j\theta}\{G_j(Z),\tilde{\xi},\psi\}=-K_{j-1}\{G_j(Z),\psi\}\lambda_{j\psi}\{G_j(Z),\psi\}/\lambda_j\{G_j(Z),\psi\}$ and $\tilde{h}_{j\xi}\{G_j(Z),\tilde{\xi},\psi\}=h_{j\xi}\{G_j(Z),\xi,\psi\}$ are column vectors of partial derivatives of (11) with respect to θ and ξ.

Noting that $\hat{\psi}$ converges in probability to ψ_0 when the coarsening probabilities are modeled correctly, if they are correct but the $h_r\{G_r(Z),\xi\}$ may not be, by an argument analogous to

that in Web Appendix B, $\hat{\xi}_{opt*}$ solving the estimating equations defined by (12), jointly in θ, will converge to ξ^{opt*}. Likewise, if $\hat{\psi}$ converges to some ψ^* when the coarsening probabilities may not be correct, if this is the case but the models $h_r\{G_r(Z), \xi\}$ are correct, analogous to Web Appendix A, the expectation of (12) evaluated at ψ^* may be shown to be equal to zero when $\tilde{\xi} = (\xi, \theta) = (\xi_0, 0)$. Thus, (i) and (ii) are satisfied; i.e., the estimator $\hat{\beta}_{opt*}$ obtained by solving (8) with the MLE $\hat{\psi}$ and the estimator $\hat{\xi}_{opt*}$ solving the estimating equations implied by (12) substituted is doubly robust and has smallest asymptotic variance among all estimators solving (8) when the coarsening probabilities are correct but the models $h_r\{G_r(Z), \xi\}$ may not be. Accordingly, $\hat{\beta}_{opt*}$ should be more efficient under the latter conditions than the doubly robust estimator for β of Bang and Robins (2005) obtained when the coarsening probabilities are modeled and ψ is estimated by ML, $\hat{\beta}_{br*}$, say.

Because $(\widehat{\beta}_{opt*}^T, \widehat{\xi}_{opt*}^T, \widehat{\psi}^T)^T$ is an M-estimator, and similarly for $\hat{\beta}_{br*}$, the asymptotic covariance matrix for each may be approximated by the empirical sandwich method (Stefanski and Boos, 2002) and will be consistent for the true sampling covariance matrices regardless of whether or not one or both sets of models is misspecified; see Web Appendix D.

An alternative approach to estimation of ξ would be to extend the methods of Tan (2006, 2007, 2008) to the setting of monotonely coarsened data. Given that the proposed approach is optimal when the discrete hazard models are correct, theoretically, such an extension would be no more efficient in this case. In simulations in Cao et al. (2009), the proposed approach outperformed that of Tan for estimation of a single population mean under both correct and incorrect models, and we would expect to see similar relative performance here.

5. Application to ACTG 175

We now demonstrate how the foregoing development is specialized to a longitudinal study with dropout by application to ACTG 175. Recall that interest focuses on $\beta = E(Y)$, where $Y = Y_5 = $ CD4 count at 96±5 weeks, the mean CD4 count for the HIV-infected population if assigned to regimens ZDV+ddI, ZDV+ddC, or ddI; $M = 4$; and $m(Z, \beta) = Y - \beta$. The baseline covariate vector X includes age (years); weight (kg); Karnofsky score (karnof), an index reflecting ability to perform activities of daily living (0 to 100); days of prior antiretroviral therapy (antidays); and binary indicator variables for hemophilia (hemo), homosexual activity (homo), history of intravenous drug use (drug), ZDV within 30 days of the trial, race (0 = white), gender (0 = female), antiretroviral history (hist; 0 = naive, 1 = experienced), and symptomatic status (symp; 0 = asymptomatic).

We consider estimation of β by the simple IPW estimator, which corresponds to solving (8) with all of the $h_r\{G_r(Z), \xi\}$ set equal to zero, $\hat{\beta}_{ipw}$; two versions of the estimator $\hat{\beta}_{br*}$ of Bang and Robins (2005); and two versions of the proposed estimator $\hat{\beta}_{opt*}$. The CAR assumption (1) is not unreasonable; it is widely acknowledged in longitudinal HIV studies that subjects with baseline characteristics such as intravenous drug use and/or lower evolving CD4 counts prior to dropout, reflecting compromised immunologic status, may be more likely to drop out. Under CAR/MAR, the naive estimator, the sample mean of CD4 counts for the complete cases at 96±5 weeks, equal to 348.7 cells/mm^3 with standard error (SE) 5.76, thus may be an overestimate if subjects with poorer immunologic status are more likely to drop out.

Using the notation at the end of Section 2, we represent the models we now present by replacing C by R and $G_r(Z)$ by $\bar{L}_j$ and indexing visits by j in obvious fashion. For use with all estimators, logistic regression models for the discrete hazards at each j were developed with main effects in elements of $\bar{L}_j$ identified via separate ML fits at each j to the data on all

subjects with $R \geq j$ using forward selection with entry level of significance 0.15; we also considered other levels, with no qualitative differences. This yielded models

$\lambda_j(\bar{L}_j, \psi) = \text{expit}(\psi_i^T \tilde{\bar{L}}_j), j = 1, \ldots, 4$, where $\text{expit}(u) = e^u/(1 + e^u)$, $\psi = (\psi_1^T, \ldots, \psi_4^T)^T$, and $\tilde{\bar{L}}_j$ is the subset of $\bar{L}_j$ selected;

$\tilde{\bar{L}}_1 = (Y_1, \text{age, drug, karnof, antidays, race, hist, symp}), \tilde{\bar{L}}_2 = (Y_2, \text{age, home, drug, antidays, karnof}), \tilde{\bar{L}}_3 = Y_3$, and $\tilde{\bar{L}}_4 = (Y_1 Y_3, \text{hemo, drug, karnof, race})$. Finding the MLE $\hat{\psi}$ then reduced to carrying out individual ML fits of these models for each j.

Noting that $E\{m(Z)|\bar{L}_j\} = E(Y|\bar{L}_j) - \beta$ for each j, developing models $h_j(\bar{L}_j, \xi)$ and $h_j^*(\bar{L}_j, \xi_j), j = 1, \ldots, 4$, for $\hat{\beta}_{opt*}$ and $\hat{\beta}_{br*}$, respectively, corresponds to developing models for the regression of 96±5 week CD4 count on $\bar{L}_j$; i.e., for $E(Y|Y_1, \ldots, Y_j, X)$. To develop models $h_j(\bar{L}_j, \xi)$, we assumed that the longitudinal data follow the linear mixed model

$$Y_{ij} = \alpha_{0i} + \alpha_{1i} t_{ij} + \gamma^T \tilde{X}_i + e_{ij}, \tag{14}$$

where $\alpha_i = (\alpha_{0i}, \alpha_{1i})^T \sim N\{(\mu_{\alpha 0}, \mu_{\alpha 1})^T, \Sigma_\alpha\}$; $e_{ij} \sim N(0, \sigma_e^2)$ are iid for all i, j; the $\alpha_i, i = 1, \ldots, n$, are independent of each other and all e_{ij}; and $\tilde{X} = (\text{weight, karnof, hist, symp})$ was identified by fitting (14) by ML with all of X included and retaining only those elements for which the usual t-test of whether or not the associated coefficient is equal to zero had p-value less than 0.05. Under (14), standard results for the multivariate normal distribution yield the required conditional expectations $E(Y|Y_1, \ldots, Y_j, X) = E(Y|Y_1, \ldots, Y_j, \tilde{X})$, all of which depend on the

common $\xi = \{\mu_{\alpha 0}, \mu_{\alpha 1}, \text{vech}(\sum_\alpha)^T, \sigma_e^2, \gamma^T\}^T$; see Web Appendix E. To obtain the first version of $\hat{\beta}_{opt*}, \widehat{\beta}_{opt*}^{(1)}$, say, we estimated ξ in the implied models $h_j(\bar{L}_j, \xi)$ using (12). For direct comparison of the Bang-Robins approach to the proposed method using the same covariate information, we let $h_j^*(\bar{L}_j, \xi_j)$ for each j be linear regression models including main effects in all CD4 counts up through j and $\tilde{X}$ and estimated the ξ_j by separate ordinary least squares (OLS) regressions for each j based on the observed data at j; denote the resulting estimator by $\widehat{\beta}_{br*}^{(1)}$. For a second version of $\hat{\beta}_{br*}$, denoted $\widehat{\beta}_{br*}^{(2)}$, we instead considered for each j all of $Y_1, \ldots, Y_j, X$ as potential main effects in linear models, and developed and fit these separately by OLS with forward selection on the elements of X. The resulting $h_j^*(\bar{L}_j, \xi_j)$ contained (age,karnof,race,gender,hist), (age,hemo,drug,karnof,antidays,gender,symp), (age,hemo,karnof,gender), and (age,hemo,karnof) for $j = 1, 2, 3, 4$, respectively, along with $(Y_1, \ldots, Y_j)$. We implemented both $\widehat{\beta}_{br*}^{(1)}$ and $\widehat{\beta}_{br*}^{(2)}$ as described by Bang and Robins (2005, Section 3). A second version of the proposed estimator, $\widehat{\beta}_{opt*}^{(2)}$, was derived by, rather than taking ξ common across j, letting the models implied by (14) for each j have j-specific parameters ξ_j. We then let $\xi = (\xi_1^T, \ldots, \xi_4^T)^T$, and estimated ξ using (12). For all estimators, we obtained SEs via the sandwich technique.

Estimation of ξ by solution of the estimating equations based on (12) may be carried out via standard techniques, such as a Newton-Raphson updating scheme. Thus, in principle, implementation is no more complex than for the Bang-Robins approach, where the ξ_r are estimated by separate solutions to M sets of estimating equations. Computation of $\hat{\xi}_{opt*}$ is likely a higher-dimensional problem than the separate ones; however, here and in Section 6, we encountered no numerical difficulties with either method.

For comparison, we also fit the mixed model (14) directly by normal ML using SAS proc mixed (SAS Institute, 2009) and estimated β by the marginal predicted value $\hat{\beta}_{mixed}$ at 96 ± 5 weeks obtained by setting X equal to its sample mean with SE from the associated estimate statement, which treats the sample mean of X as fixed.

The resulting $\hat{\beta}_{ipw} = 332.96$, (SE 5.10), $\widehat{\beta}_{br*}^{(1)} = 333.34$ (4.96), $\widehat{\beta}_{opt*}^{(1)} = 333.15$ (4.90), $\widehat{\beta}_{br*}^{(2)} = 333.44$ (4.96), $\widehat{\beta}_{opt*}^{(2)} = 333.35$ (4.76). Recognizing that this is a single data set, it is encouraging to note that the estimates are virtually identical, and, consistent with the theory, the IPW estimator is inefficient relative to the AIPW competitors on the basis of estimated SE. Moreover, both versions of the proposed estimator achieve or surpass the performance of the Bang and Robins estimators, although not dramatically, and all estimates are indeed smaller than the naive estimate, as expected. We also obtained $\hat{\beta}_{mixed} = 346.20$ (4.92); in contrast to the AIPW estimates, this estimate is not appreciably different from the naive.

We deliberately chose the ACTG 175 study to demonstrate the methods because of a unique feature that highlights the advantage of consideration of the general setting of monotone coarsening. Although subjects in the study ceased to attend clinic visits and provide CD4 counts after some time point, so effectively did "drop out" of the study with respect to the response of interest, follow-up of all subjects continued. Thus, additional information on each subject throughout the entire 96-week period, regardless of whether or not s/he ceased to attend clinic visits, is available, which we summarize in four time-dependent covariates $dis_{ij} = I\{$subject i discontinued study treatment during $(t_j, t_{j+1}]\}$, $j = 1, \ldots, 4$; we did not include dis_j in the definitions of L_j in the foregoing analysis for illustrative simplicity, although we could have done so. Acknowledging these data takes this situation out of the realm of the standard longitudinal dropout setting and notation at the end of Section 2, which assumes that no data are available beyond visit j if the subject was last seen at j. However, the present setting may still be cast as a case of monotone coarsening and these additional data incorporated in the analysis, as we now demonstrate.

Reverting to the general notation, $Z = (X, Y_1, Y_2, Y_3, Y_4, Y, dis_1, dis_2, dis_3, dis_4)$; and, with $C = r$ indicating that the subject last provided a CD4 count at visit r, we observe $G_r(Z) = (X, Y_1, \ldots, Y_r, dis_1, dis_2, dis_3, dis_4)$, $r = 1, \ldots, 4$, and $G_\infty(Z) = Z$ for $r = \infty$. Clearly, the coarsened data satisfy the monotonicity requirement. This demonstrates that one need not think strictly temporally in characterizing monotone coarsening in longitudinal data.

Recall that the goal is to estimate β = mean CD4 count at 96 ± 5 weeks for the population *assigned* to ZDV+ddI, ZDV+ddC, or ddI, so regardless of whether or not subjects stayed on these regimens for the entire 96 weeks. We illustrate by calculating $\hat{\beta}_{opt*}$ and $\hat{\beta}_{br*}$ as follows. For both estimators, we derived the discrete hazard models by the same strategy as in the previous analysis, considering all elements of $G_r(Z)$ as possible main effects in the linear predictor of a logistic regression model for each r and retaining a subset of these terms by forward selection. This yielded logistic regression models $\lambda_r\{G_r(Z), \psi_r\}$ that included main effects for $(Y_1, age, drug, karnof, antidays, race, hist, symp)$, $(Y_2, age, homo, drug, antidays, karnof, dis_1, dis_2)$, (Y_3, dis_1, dis_2), and $(Y_1, Y_3, hemo, karnof, race, dis_2, dis_4)$ for $r = 1, 2, 3, 4$, respectively. To derive models $h_r\{G_r(Z), \xi\}$ for $\hat{\beta}_{opt*}$, we used the form of $E(Y|X, Y_1, \ldots, Y_r, dis_1, dis_2, dis_3, dis_4)$ implied by the linear mixed model $Y_{ir} = \alpha_{0i} + \alpha_{1i}t_{ir} + \gamma^T\tilde{X}_i + \varphi_1 I(r \geq 3)dis_{i2} + \varphi_2 I(r = 5)dis_{i4} + e_{ir}$, where the random effects and within-subject deviations are normal as above, and now $\tilde{X} = (weight, karnof, symp)$; see Web Appendix E. The common

$$\xi = \{\mu_{\alpha 0}, \mu_{\alpha 1}, \mathrm{vech}(\textstyle\sum_\alpha)^T, \sigma_e^2, \gamma^T, \varphi_1, \varphi_2\}^T$$ was then estimated via (12). For $\hat{\beta}_{br*}$, we took

$h_r^*\{G_r(Z), \xi_r\} = r_r^T \widetilde{X} + \varphi_{1,r} \text{dis}_2 + \varphi_{2,r} \text{dis}_4 + \zeta_r^T(Y_1, \ldots, Y_r)$, so $\xi_r = (\gamma_r^T, \varphi_{1,r}, \varphi_{2,r}, \zeta_r^T)^T$, which was estimated by OLS for each r. Using these estimated discrete hazards to also calculate $\hat{\beta}_{ipw}$, $\hat{\beta}_{ipw} = 325.32\ (5.80)$, $\hat{\beta}_{opt*} = 328.10\ (5.05)$, and $\hat{\beta}_{br*} = 327.46\ (5.49)$. As before, performance of the estimators based on estimated SEs is consistent with the theory.

6. Simulation Studies

We carried out several simulations to assess the performance of the proposed methods in the case of a longitudinal study with dropout, which we describe using the notation at the end of Section 2. To obtain data for subject i, $i = 1, \ldots, n$, we generated baseline covariates ($t_1 = 0$) $X_i = (X_{i1}, X_{i2})^T$, where $X_{i1} \sim N(5, 1)$, and $X_{i2} \sim$ Bernoulli(0.5). For visit times $(t_1, t_2, t_3) = (0, 1, 2)$, we generated longitudinal responses via the mixed model $Y_{ij} = \alpha_{0i} + \alpha_{1i}t_{ij} + \gamma^T X_i + e_{ij}$, where $(\alpha_{0i}, \alpha_{1i})^T \sim N\{(1.0, 2.5)^T, \Sigma\}$, vech$(\Sigma) = (0.3, 0.1, 0.2)^T$, $\gamma = (1, -1)^T$, and $e_{ij} \sim N(0, 1)$. Thus, $L_1 = (X, Y_1)$, $L_2 = Y_2$, and $L_3 = Y_3 = Y$. As in ACTG 175, we focus on estimation of $\beta = E(Y) = 10.5$. This setup implies that, in truth, $E(Y|\bar{L}_1) = \gamma^T X + \mu_3(X, Y_1) + t_3\mu_4(X, Y_1)$ and $E(Y|\bar{L}_2) = \gamma^T X + \mu_1(X, Y_1, Y_2) + t_3\mu_2(X, Y_1, Y_2)$, where the forms of $\mu_1, \ldots, \mu_4$ are given in Web Appendix F. We considered two dropout scenarios; in both, letting $U_1 = I(Y_1 > 5.8)$ and $U_2 = I(Y_2 > 6.2)$, dropout was induced according to the discrete hazards $\lambda_1(\bar{L}_1, \psi) = \text{expit}(\psi_{0,1} + \psi_{1,1}U_1)$ and $\lambda_2(\bar{L}_2, \psi) = \text{expit}(\psi_{0,2} + \psi_{1,2}U_1 + \psi_{2,2}U_2)$, so $\psi = (\psi_{0,1}, \psi_{1,1}, \psi_{0,2}, \psi_{1,2}, \psi_{2,2})^T$. A concern with methods involving inverse weighting is the influence of extreme estimated inverse weights. In the "moderate" scenario, $\psi = (-2.0, 2.5, -2.0, 2.0, 2.5)^T$, representing the potential for "moderately large" estimated inverse weights and resulting in 36% and 70% missing Y_2 and Y on average, respectively. In the "extreme" scenario, $\psi = (-3.5, 5.0, -2.1, 2.0, 2.89)^T$, with 40% and 71% missing Y_2 and Y, yielding more "extreme" inverse weights; see below. For each situation below, we estimated β by $\hat{\beta}_{ipw}$, $\hat{\beta}_{opt*}$, and $\hat{\beta}_{br*}$, with SEs obtained by the sandwich method; and by $\hat{\beta}_{mixed}$ obtained by fitting the mixed model above, analogous to the approach in Section 5, with SE calculated as in that section.

For each inverse weight scenario, we considered the four situations of all combinations of correct or incorrect regression models $h_j(\bar{L}_j, \zeta)$ and $h_j^*(\bar{L}_j, \xi_j)$ for $E(Y|\bar{L}_j)$, $j = 1, 2$, and correct or incorrect discrete hazard models $\lambda_j(\bar{L}_j, \psi)$. Incorrect discrete hazard models were specified by replacing (U_1, U_2) in the logistic regressions above by (Y_1, Y_2). Incorrect models for $E(Y|\bar{L}_j)$ were obtained by eliminating all terms involving X and replacing (Y_1, Y_2) by $[\exp\{(Y_1/9)^2\}, (Y_1 + 3)/\{1 + \exp(Y_2)\} + 1]$ in $\mu_1, \ldots, \mu_4$ above. Correct and incorrect discrete hazard models were fit by ML. For $\hat{\beta}_{opt*}$, the implied ζ in the correct or incorrect models was estimated based on (12); for $\hat{\beta}_{br*}$, the implied ξ_j were estimated by separate OLS regressions at each j. We considered $n = 500, 1000$ and $n = 500$ for the "moderate" and "extreme" scenarios, with 1000 Monte Carlo data sets for each n-situation combination.

Table 1 summarizes the distributions of estimated inverse weights $\{\pi(\infty, Z, \hat{\psi})\}^{-1}$ and $[K_r\{G_r(Z), \hat{\psi}\}]^{-1}$ in (8) for the "moderate" and "extreme" scenarios, showing that, in both, rather large inverse weights are possible. Results for estimation of β are presented in Tables 2 and 3. When both sets of models are correct, $\hat{\beta}_{opt*}$ and $\hat{\beta}_{br*}$ exhibit virtually identical performance and considerable efficiency gains over $\hat{\beta}_{ipw}$ in the "moderate" scenario, as expected; interestingly, $\hat{\beta}_{br*}$ compares poorly to both $\hat{\beta}_{ipw}$ and $\hat{\beta}_{opt*}$ in the "extreme" scenario, suggesting a possible finite sample effect of more extreme weights. Bias of $\hat{\beta}_{ipw}$, $\hat{\beta}_{br*}$, and $\hat{\beta}_{opt*}$ is inconsequential in all situations for the "moderate" scenario, although showing a trend consistent with theory when one or both models are incorrect. When the discrete hazards are correct but the regression models are not, $\hat{\beta}_{opt*}$ shows efficiency gains over $\hat{\beta}_{br*}$ in both scenarios, as expected from its construction; $\hat{\beta}_{br*}$ performs poorly in the "extreme" scenario. In the converse situation, performance of the two estimators is

comparable in the "moderate" scenario, but $\hat{\beta}_{br*}$ is highly variable in the "extreme" scenario, and $\hat{\beta}_{ipw}$ is substantially biased, as expected. When both sets of models are incorrect, $\hat{\beta}_{opt*}$ shows a large gain in efficiency over $\hat{\beta}_{br*}$ in the "moderate" scenario. Under the "extreme" scenario, all estimators are biased, but $\hat{\beta}_{opt*}$ is considerably more efficient than $\hat{\beta}_{ipw}$ or $\hat{\beta}_{br*}$. In all situations in both scenarios, except when both sets of models are misspecified, leading to estimated SEs that do not reflect the true sampling variability, confidence intervals based on all estimators for the most part achieve nominal coverage. Not surprisingly, in both scenarios, when the mixed model fitted is correctly specified, $\hat{\beta}_{mixed}$ exhibits considerably better precision; however, SEs calculated as conventional in practice lead to coverage that falls short of the nominal level, reflecting failure to account for the variation in $\tilde{X}$.

These results suggest that the proposed approach may be more stable than competing methods in situations with extreme estimated inverse weights. The method seeks to obtain as efficient an estimator of β as possible under correct discrete hazards models, where the estimator for ξ serves only to increase efficiency and is not of inherent interest. Our estimator for ξ minimizes the expected squared residual in (4) or (10), where one may think of the residual as subtracting the augmentation term (involving ξ) from the IPW complete case term $I(C = \infty)m(Z)/\pi(\infty, Z)$. We conjecture that the resulting choice of ξ acts automatically to counteract, to the extent possible, the destabilizing effects that large inverse weights $[\pi(\infty, Z)]^{-1}$ in particular have on the variance of the estimator for β.

7. Discussion

We have proposed doubly robust estimators for general semiparametric full data model parameters based on data subject to monotone coarsening at random. A special case is that of longitudinal data subject to MAR dropout. As for a population mean under MAR response as in Cao et al. (2009), the methods are designed to equal or exceed the asymptotic efficiency relative to other doubly robust estimators when models for the coarsening mechanism are correctly specified, even when regression models incorporated to increase efficiency are not. In contrast to simulations by Kang and Schafer (2007), our empirical studies show that doubly robust estimators need not exhibit disastrous performance, even when both sets of models are incorrectly specified, and that the proposed estimator may outperform competing methods and be more stable in the presence of very large inverse weights.

As noted in Section 4, the regression models must be consistent with one another for each level of coarsening, which may be ensured through specification of a part of the joint distribution of the full data. The impact of such specification and more generally the role of model selection for both the regressions and coarsening mechanism merits formal study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005; 61:962–972. [PubMed: 16401269]

Birmingham J, Rotnitzky A, Fitzmaurice GM. Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns. Journal of the Royal Statistical Society, Series B. 2003; 65:275–297.

Cao W, Tsiatis AA, Davidian M. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. Biometrika. 2009; 96:723–734. [PubMed: 20161511]

Gill, RD.; van der Laan, MJ.; Robins, JM. Coarsening at random: characterizations, conjectures and counter examples. Proceedings of The First Seattle Symposium in Biostatistics: Survival Analysis; New York: Springer; 1997. p. 255-294.

Goetgeluk S, Vansteelandt S, Goetghebeur E. Estimation of controlled direct effects. Journal of the Royal Statistical Society, Series B. 2009; 70:1049–1066.

Hammer SM, Katzenstein DA, Hughes MD, Gundaker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. for the AIDS Clinical Trials Group Study 175 Study Team. A trial comparing nucleoside monotherapy with combination therapy in HIV infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine. 1996; 335:1081–1089. [PubMed: 8813038]

Heitjan DF, Rubin DB. Ignorability and coarse data. The Annals of Statistics. 1991; 19:2244–2253.

Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Statistics in Medicine. 2004; 23:1455–1497. [PubMed: 15116353]

Kang DY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion and rejoinder). Statistical Science. 2007; 22:523–380.

Little, RJA. Selection and pattern-mixture models. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. Longitudinal Data Analysis. Boca Raton, FL: Chapman and Hall/CRC; 2009. p. 409-431.

Molenberghs, G.; Fitzmaurice, G. Incomplete data: Introduction and overview. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. Longitudinal Data Analysis. Boca Raton, FL: Chapman and Hall/CRC; 2009. p. 395-408.

Philipson PM, Ho WK, Henderson R. Comparative review of methods for handling drop-out in longitudinal studies. Statistics in Medicine. 2008; 27:6276–6298. [PubMed: 18937274]

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association. 1994; 89:846–866.

Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995; 90:106–121.

Rotnitzky, A. Inverse probability weighted methods. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. Longitudinal Data Analysis. Boca Raton, FL: Chapman and Hall/CRC; 2009. p. 453-476.

Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. Journal of the American Statistical Association. 1998; 93:1321–1339.

SAS Institute. SAS/STAT 9.2 User's Guide. Cary NC: SAS Institute Inc; 2009.

Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable dropout using semiparametric nonresponse models. (with discussion and rejoinder). Journal of the American Statistical Association. 1999; 94:1096–1146.

Seaman S, Copas A. Doubly robust generalized estimating equations for longitudinal data. Statistics in Medicine. 2009; 28:937–955. [PubMed: 19153970]

Stefanski LA, Boos DD. The calculus of M-estimation. The American Statistician. 2002; 56:29–38.

Tan Z. A distributional approach for causal inference using propensity scores. Journal of the American Statistical Association. 2006; 101:1619–1637.

Tan Z. Understanding OR, PS and DR. Statistical Science. 2007; 22:560–568.

Tan Z. Comment: Improved Local Efficiency and Double Robustness. The International Journal of Biostatistics. 2008; 4(1) Article 10. 10.2202/1557–4679.1109

Tsiatis, AA. Semiparametric Theory and Missing Data. New York: Springer; 2006.

Table 1

Summaries of distributions of estimates of inverse weights for simulated subjects across 1000 Monte Carlo data sets. Min and Max are the minimum and maximum values across all subjects for whom the indicated inverse weights were calculated across all 1000 data sets, and SD is the standard deviation. The Moderate and Extreme scenarios are described in the text. Results for the Moderate scenario for $n = 1000$ were similar to those for $n = 500$.

	Min	Percentiles									Max	Mean	SD
		1%	5%	10%	25%	50%	75%	90%	95%	99%			
Moderate scenario, $n = 500$													
Correct models													
$[K_1\{G_1(Z),\psi\}]^{-1}$	1.07	1.08	1.10	1.11	1.13	1.18	2.64	2.85	2.97	3.21	3.48	1.87	0.78
$[K_2\{G_2(Z),\psi\}]^{-1}$	1.14	1.20	1.24	1.27	1.37	3.02	4.92	43.27	44.19	66.07	146.59	9.44	15.65
$[\pi\{\infty,Z,\psi\}]^{-1}$	1.14	1.19	1.23	1.24	1.28	1.40	3.12	4.79	6.01	36.97	146.59	3.31	6.12
Inorrect models													
$[K_1\{G_1(Z),\psi\}]^{-1}$	1.00	1.04	1.08	1.12	1.24	1.49	2.01	2.95	3.92	7.46	50.0	1.89	1.46
$[K_2\{G_2(Z),\psi\}]^{-1}$	1.01	1.17	1.35	1.51	1.95	2.98	5.56	12.19	21.46	71.28	1960.13	7.04	22.37
$[\pi\{\infty,Z,\psi\}]^{-1}$	1.01	1.13	1.26	1.37	1.64	2.18	3.33	5.75	8.52	22.74	815.62	3.50	7.57
Extreme scenario, $n = 500$													
Correct models													
$[K_1\{G_1(Z),\psi\}]^{-1}$	1.00	1.01	1.02	1.02	1.03	1.05	5.48	6.13	6.50	7.50	11.40	3.19	2.34
$[K_2\{G_2(Z),\psi\}]^{-1}$	1.06	1.09	1.12	1.13	1.19	3.14	3.66	15.42	91.54	199.21	1333.93	13.79	41.12
$[\pi\{\infty,Z,\psi\}]^{-1}$	1.06	1.08	1.10	1.12	1.14	1.20	3.22	3.79	8.72	50.91	359.18	3.40	11.70
Inorrect models													
$[K_1\{G_1(Z),\psi\}]^{-1}$	1.00	1.01	1.01	1.02	1.09	1.48	3.50	12.04	28.20	159.82	156770.00	14.85	442.93
$[K_2\{G_2(Z),\psi\}]^{-1}$	1.01	1.09	1.22	1.34	1.70	2.56	4.86	12.39	29.51	274.38	258602.00	36.10	1360.00
$[\pi\{\infty,Z,\psi\}]^{-1}$	1.01	1.07	1.15	1.23	1.43	1.86	2.88	5.61	9.49	43.22	240730.00	8.36	658.50

Table 2

Simulation results for the "moderately large" inverse weight scenario; 1000 Monte Carlo replications. Bias is Monte Carlo bias, RMSE is root mean square error, MCSD is Monte Carlo standard deviation, AveSE is average of sandwich standard errors, Cov is Monte Carlo coverage of 95% Wald confidence intervals, R denotes regression models, and DH denotes discrete hazard models. True value of $\beta = 10.5$. Smallest, median, second largest, and largest standard errors for table entries: Bias, (0.019, 0.033, 0.067, 0.077); AveSE, (0.004, 0.022, 0.086, 0.340); Cov, (0.007, 0.008, 0.010, 0.011)

	Bias	RMSE	MCSD	AveSE	Cov	Bias	RMSE	MCSD	AveSE	Cov
n = 1000										
	R correct, DH correct					**R correct, DH incorrect**				
β_{ipw}	−0.03	0.85	0.85	0.87	0.94	0.79	2.07	1.91	1.16	0.94
β_{br*}	−0.02	0.60	0.60	0.58	0.95	−0.01	0.61	0.61	0.60	0.95
β_{opt*}	−0.01	0.60	0.60	0.61	0.95	−0.03	0.64	0.64	0.66	0.95
β_{mixed}	−0.02	0.28	0.28	0.25	0.91	−0.02	0.28	0.28	0.25	0.91
	R incorrect, DH correct					**R incorrect, DH incorrect**				
β_{ipw}	−0.03	0.85	0.85	0.87	0.94	0.79	2.07	1.91	1.16	0.94
β_{br*}	0.01	0.81	0.81	0.83	0.94	0.03	2.11	2.11	1.71	0.90
β_{opt*}	−0.04	0.72	0.72	0.73	0.94	−0.39	1.51	1.46	1.39	0.85
β_{mixed}	−3.79	3.80	0.30	0.19	0.00	−3.79	3.80	0.30	0.19	0.00
n = 500										
	R correct, DH correct					**R correct, DH incorrect**				
β_{ipw}	0.00	1.37	1.37	1.31	0.94	0.70	2.53	2.43	1.41	0.92
β_{br*}	0.00	0.86	0.86	0.90	0.95	0.02	1.03	1.03	1.02	0.95
β_{opt*}	−0.01	0.86	0.86	0.89	0.95	0.01	1.04	1.04	1.14	0.94
β_{mixed}	−0.01	0.41	0.41	0.35	0.91	−0.01	0.41	0.41	0.35	0.91
	R incorrect, DH correct					**R incorrect, DH incorrect**				
β_{ipw}	0.00	1.37	1.37	1.31	0.94	0.70	2.53	2.43	1.41	0.92
β_{br*}	0.04	1.11	1.11	1.01	0.94	0.17	2.52	2.51	1.17	0.94
β_{opt*}	−0.02	1.04	1.04	1.10	0.94	−0.31	1.51	1.48	1.89	0.88
β_{mixed}	−3.80	3.82	0.42	0.27	0.00	−3.80	3.82	0.42	0.27	0.00

Table 3

Simulation results for the "extreme" inverse weight scenario, 1000 Monte Carlo data sets; entries and true value of β are as in Table 2. Smallest, median, second largest, and largest standard errors for table entries: Bias, (0.033, 0.182, 1.008, 3.920); AveSE, (0.020, 0.022, 0.180, 1.26); Cov, (0.005, 0.006, 0.007, 0.009)

	Bias	RMSE	MCSD	AveSE	Cov	Bias	RMSE	MCSD	AveSE	Cov
	n = 500									
	R correct, DH correct					*R correct, DH incorrect*				
β_{ipw}	−0.09	2.65	2.65	2.61	0.95	13.84	124.74	123.97	15.68	0.91
β_{br*}	−0.06	5.27	5.27	5.52	0.94	0.41	31.89	31.89	30.99	0.94
β_{opt*}	0.06	1.15	1.15	1.10	0.93	−0.04	1.05	1.05	1.12	0.96
β_{mixed}	−0.00	0.44	0.44	0.36	0.89	−0.00	0.44	0.44	0.36	0.89
	R incorrect, DH correct					*R incorrect, DH incorrect*				
β_{ipw}	−0.09	2.65	2.65	2.61	0.95	13.84	124.74	123.97	15.68	0.91
β_{br*}	1.06	6.31	6.22	5.97	0.93	8.07	49.53	8.87	17.24	0.97
β_{opt*}	−0.35	2.51	2.49	2.53	0.94	3.14	11.84	11.42	10.10	0.92
β_{mixed}	−3.86	3.88	0.45	0.27	0.00	−3.86	3.88	0.45	0.27	0.00