BIOMETRICS 57, 103–113
March 2001

Methods for Conducting Sensitivity Analysis of Trials with Potentially Nonignorable Competing Causes of Censoring

Andrea Rotnitzky,[1] Daniel Scharfstein,[3,*] Ting-Li Su,[4] and James Robins[1,2]

Departments of [1]Biostatistics and [2]Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.

[3]Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland 21205, U.S.A.

[4]Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, England, U.K.

*email: dscharf@jhsph.edu

SUMMARY. We consider inference for the treatment-arm mean difference of an outcome that would have been measured at the end of a randomized follow-up study if, during the course of the study, patients had not initiated a nonrandomized therapy or dropped out. We argue that the treatment-arm mean difference is not identified unless unverifiable assumptions are made. We describe identifying assumptions that are tantamount to postulating relationships between the components of a pattern-mixture model but that can also be interpreted as imposing restrictions on the cause-specific censoring probabilities of a selection model. We then argue that, although sufficient for identification, these assumptions are insufficient for inference due to the curse of dimensionality. We propose reducing dimensionality by specifying semiparametric cause-specific selection models. These models are useful for conducting a sensitivity analysis to examine how inference for the treatment-arm mean difference changes as one varies the magnitude of the cause-specific selection bias over a plausible range. We provide methodology for conducting such sensitivity analysis and illustrate our methods with an analysis of data from the AIDS Clinical Trial Group (ACTG) study 002.

KEY WORDS: Attrition; Augmented inverse probability of censoring weighted estimation; Curse of dimensionality; Longitudinal data; Noncompliance; Pattern-mixture models; Selection bias; Selection models.

1. Introduction

The AIDS Clinical Trial Group (ACTG) study 002 was a double-blind, randomized clinical trial designed to compare the efficacy and safety of high-dose AZT (1500 mg/day) versus low-dose AZT (1200 mg/day for 4 weeks and 600 mg/day thereafter) in AIDS patients (Fischl et al., 1990). Between December 2, 1986, and November 12, 1987, 520 subjects were enrolled and randomized to receive one of the two treatments, with 261 subjects assigned to high-dose AZT and 259 assigned to low-dose AZT. The design of the study called for clinic visits to be made every 8 weeks, at which time data on CD4 lymphocyte count, total white blood cell count (WBC), number of bouts of *Pneumocystis carinii* pneumonia (PCP), and whether or not the subject was still taking the assigned AZT treatment were to be obtained. In this article, we will be concerned with the analysis of data from the first four postbaseline clinic visits.

Some patients stopped coming to their scheduled clinic visits and others had intermittent clinic visits, i.e., missed a clinic visit but returned for a later one. In the high-dose AZT arm, 137 subjects stopped their clinic visits and 22 had intermittent clinic visits. For the low-dose AZT arm, these figures were 114 and 25, respectively.

During the course of the trial, evidence from other studies pointed to the potential benefits of prophylaxis therapy for PCP. In August 1987, the study was revised to allow the use of prophylaxis therapy for subjects who experienced a second bout of PCP. In April 1988, the study was further changed to allow all subjects to use prophylaxis therapy. Except for general guidelines, the decision whether or when to administer prophylaxis therapy was left up to the patients and their physicians. As a result, prophylaxis therapy was a nonrandomized treatment embedded in the randomized trial. Of the 261 (259) patients in the high-dose (low-dose) AZT arm, 18 (30) had initiated prophylaxis at or prior to week 32.

The initiation of a nonrandomized therapy like prophylaxis complicates the analysis of even an ideal and unrealistic trial in which no patient misses a clinic visit. If the nonrandomized therapy influences the outcome of interest and the rates of therapy differ in the two treatment arms, then differences in the observed outcome distributions of the two treatment arms do not necessarily reflect treatment effects because these differences might only be due to the fact that the rate of therapy is higher in one arm than in the other. In the ACTG study, the rate of prophylaxis therapy was higher in the low-dose arm, and it is biologically plausible that prophylaxis therapy

might have a beneficial effect on CD4 count by preventing PCP since PCP is an opportunistic infection that can stimulate HIV replication.

Therefore, one important public health question raised by ACTG 002 is whether or not high-dose AZT would have had a different effect than low-dose AZT had all subjects received identical prophylaxis therapy. In this article, we are concerned with addressing the following specific question: Would there have been a difference between the means of CD4 count at week 32 in the two treatment arms in the hypothetical scenario in which no subject missed the clinic visit at week 32 or started prophylaxis therapy during the course of the trial?

In Section 6, we note that, under our analytic model, unless all patients return to the clinic visit at week 32, data obtained on patients subsequent to missing a clinic visit or initiating prophylaxis therapy is, asymptotically, uninformative for inference about the treatment mean difference of interest. Thus, in our analyses, we disregard the data obtained on subjects subsequent to their first missed clinic visit or their time of initiation of prophylaxis. With a slight abuse of terminology, we refer to the first time a subject misses a clinic visit as his/her drop-out time. In our approach, we regard drop-out and initiation of prophylaxis as competing causes of censoring.

In Section 3, we argue that the treatment-arm-specific means of interest and hence their difference, denoted as Δ, are not identifiable from the data recorded in ACTG 002. To identify them, we must make nonverifiable assumptions. We then describe a class of nonverifiable, identifying assumptions. In doing so, our purpose is twofold. First, we want to isolate a set of untestable assumptions, i.e., assumptions that can't be rejected by any statistical test, that are sufficient for identification of Δ. Our second objective is to provide various interpretations of these identifying assumptions with the goal of facilitating communication with the field investigators. We argue that these assumptions essentially encode the *a priori,* unverifiable belief about the influences of unmeasured prognostic factors for CD4 count at week 32 on the decision of subjects to drop out of the study and of doctors to prescribe prophylaxis therapy. Our proposal is to model these influences via (nonidentified) cause-specific selection-bias functions. We recommend conducting a sensitivity analysis to examine how inference about Δ changes as one varies the selection-bias functions over plausible ranges.

2. ACTG 002 Data Configuration

Throughout, we refer to a subject that has not dropped out or started prophylaxis at or prior to visit t as uncensored at visit t. We say that the subject is censored due to drop-out (prophylaxis use) by visit t if he/she missed a clinic visit (started prophylaxis) at or prior to visit t and prior to initiation of prophylaxis (missing a clinic visit). Let $X = 0$ if the subject is assigned to the low-dose AZT arm and $X = 1$ otherwise. Define $\boldsymbol{V}_t$, $t = 0, \ldots, 4$, as the vector $(CD4_t, WBC_t, PCP_t, AZT_t)'$, denoting the CD4 count and white blood cell count at visit t, i.e., at week $8t$, the number of PCP bouts experienced prior to visit t, and the indicator that the subject is taking some AZT at visit t that would be recorded (possibly contrary to fact) if the subject is not censored at or prior to visit t. Let $Y = CD4_4$. Our goal is to conduct inferences about the treatment-arm-specific means $\mathrm{E}(Y \mid X = x)$ and their difference $\Delta = \mathrm{E}(Y \mid X = 1) - \mathrm{E}(Y \mid X = 0)$.

For $t = 1, \ldots, 4$, define the cause-specific censoring indicator,

$$R_t = \begin{cases} 0 & \text{if the subject is uncensored at visit } t, \\ 1 & \text{if the subject was censored by visit } t \\ & \text{due to drop-out,} \\ 2 & \text{if the subject was censored by visit } t, \\ & \text{due to prophylaxis use,} \end{cases}$$

and set $R_{(-1)} = R_0 = 0$. Let $C = 5$ if $R_4 = 0$ and $C = \min\{t : R_t \neq 0\}$ otherwise. We regard the observed data in trial 002 as 520 realizations of the random vector $\boldsymbol{O} = (\boldsymbol{R}, X, \bar{\boldsymbol{V}}_C)$, where $\boldsymbol{R} = (R_1, \ldots, R_4)$ and, for any $t = 1, \ldots, 5$, $\bar{\boldsymbol{V}}_t = (\boldsymbol{V}_0, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{t-1})$.

In Table 1, we present the cumulative percentages of subjects who were censored due to drop-out, censored due to prophylaxis use, and remained uncensored at each clinic visit, stratified by treatment arm. Note that 60% (7%) of subjects are censored due to drop-out (prophylaxis use) in the high-dose arm compared with 51% (12%) in the low-dose arm. Table 2 presents summary statistics of the time-varying covariates $\boldsymbol{V}_t$ for uncensored subjects at each visit time, stratified by treatment group.

3. Identifiability and a Sensitivity Analysis Philosophy

In the ACTG 002 study, the observed data $\mathbf{O}$ does not identify either $\mathrm{E}(Y \mid X = x)$ or Δ unless $R_4 = 0$ with probability one. By this we mean that, even if the law of $\mathbf{O}$ was entirely known, this would not suffice to determine what the true values of $\mathrm{E}(Y \mid X = x)$ or Δ are. To see this, write $f(Y \mid X)$ as the

Table 1

Cumulative percentages of subjects who were censored due to drop-out, censored due prophylaxis use, and remained uncensored at each clinic visit, stratified by treatment arm

Group	Status	Week 8	Week 16	Week 24	Week 32
High dose	Censored due to drop-out	21.07%	35.25%	48.24%	60.15%
	Censored due to prophylaxis use	1.15%	1.92%	4.21%	6.90%
	Uncensored	77.78%	62.48%	47.51%	32.95%
Low dose	Censored due to drop-out	13.90%	27.41%	37.84%	50.97%
	Censored due to prophylaxis use	1.54%	3.47%	6.18%	11.58%
	Uncensored	84.56%	69.11%	55.98%	37.45%

Table 2

Mean CD4, WBC, PCP, and percentage of subjects who took their assigned AZT treatment among uncensored subjects at each clinic visit, stratified by treatment group

Variable	Group	Baseline	Week 8	Week 16	Week 24	Week 32
CD4	High dose	85.43	140.70	108.89	83.45	72.67
	Low dose	86.50	146.99	104.32	77.94	76.22
WBC/100	High dose	38.95	37.74	32.76	31.38	32.29
	Low dose	38.00	41.93	37.56	34.52	33.58
PCP	High dose	0.00	0.05	0.11	0.19	0.45
	Low dose	0.00	0.04	0.10	0.31	0.47
AZT(%)	High dose	100.00	86.21	91.46	93.55	90.70
	Low dose	100.00	91.32	94.41	93.10	95.88

mixture

$$f(Y \mid X) = f(Y \mid R_4 = 0, X)P(R_4 = 0 \mid X) + \sum_{j=1}^{2}\sum_{t=1}^{4} P\left(R_t - R_{(t-1)} = j \mid X\right) f_{tj}(Y \mid X), \tag{1}$$

where

$$f_{tj}(Y \mid X) = \int f(Y \mid R_t - R_{(t-1)} = j, \bar{V}_t, X) \times f(\bar{V}_t \mid R_t - R_{(t-1)} = j, X)d\bar{V}_t.$$

In this mixture, the law $f(Y \mid R_t - R_{(t-1)} = j, \bar{V}_t, X)$, being the distribution of Y in a censored population, is not identified by the observed data $\boldsymbol{O}$. This implies that $f(Y \mid X)$ and $\mathrm{E}(Y \mid X)$ are also not identified from $\boldsymbol{O}$.

Our strategy to resolve the identification problem is to postulate a known exponential tilt relationship between $f(Y \mid R_t = 0, \bar{V}_t, X)$ (the distribution of Y for uncensored subjects at each occasion t with a given recorded history $\bar{V}_t$) and $f(Y \mid R_t - R_{(t-1)} = j, \bar{V}_t, X)$ (the distribution of Y for patients with the same recorded past that are censored at time t due to a specific cause); i.e., for $t = 1, \ldots, 4$, we specify functions $q_{t1}(Y, \bar{V}_t, X)$ and $q_{t2}(Y, \bar{V}_t, X)$ and postulate that, for $j = 1, 2$,

$$f\left(Y \mid R_t - R_{(t-1)} = j, \bar{V}_t, X\right) = \frac{f\left(Y \mid R_t = 0, \bar{V}_t, X\right) \exp\left\{q_{tj}\left(Y, \bar{V}_t, X\right)\right\}}{\mathrm{E}\left[\exp\left\{q_{tj}\left(Y, \bar{V}_t, X\right)\right\} \mid R_t = 0, \bar{V}_t, X\right]}. \tag{2}$$

Note that (2) implicitly assumes that $f(Y \mid R_t = 0, \bar{V}_t, X)$ and $f(Y \mid R_t - R_{(t-1)} = j, \bar{V}_t, X)$, $j = 1, 2$, have the same supports. The expectation $\mathrm{E}[\exp\{q_{tj}(Y, \bar{V}_t, X)\} \mid R_t = 0, \bar{V}_t, X]$ in (2) exists because $CD4$ count, and hence Y, has bounded support. In the Appendix, we show by reverse induction that assumption (2) identifies $f(Y \mid X)$.

Using Bayes' rule, it is easy to verify that relationship (3) is equivalent to assuming that

$$\log\left\{\frac{P\left(R_t = j \mid R_{(t-1)} = 0, Y, \bar{V}_t, X\right)}{P\left(R_t = 0 \mid R_{(t-1)} = 0, Y, \bar{V}_t, X\right)}\right\} = h_{tj}\left(\bar{V}_t, X\right) + q_{tj}\left(Y, \bar{V}_t, X\right), \tag{3}$$

where $h_{tj}(\bar{V}_t, X)$, $t = 4, \ldots, 1$, $j = 1, 2$, are defined via backward recursion as follows:

$$h_{tj}\left(\bar{V}_t, X\right) = \log\left\{\frac{P\left(R_t = j \mid R_{(t-1)} = 0, \bar{V}_t, X\right)}{m_t\left(\bar{V}_t, X\right)}\right\}, \tag{4}$$

with

$$m_t\left(\bar{V}_t, X\right) = \mathrm{E}\left[\frac{I(R_4 = 0)\bar{\pi}^{-1}_{(t+1)0}\left(Y, \bar{V}_4; h\right)\exp\left\{q_{tj}\left(Y, \bar{V}_t, X\right)\right\}}{\bar{\pi}_{(t+1)0}\left(Y, \bar{V}_4; h\right)}\middle| R_{(t-1)} = 0, \bar{V}_t, X\right] = \prod_{k=t+1}^{4}\left[1 + \sum_{l=1}^{2}\exp\left\{h_{kl}\left(\bar{V}_k, X\right) + q_{kl}\left(Y, \bar{V}_k, X\right)\right\}\right]^{-1},$$

and $\Pi_{k=5}^{4}\{\cdot\} = 1$.

Equation (3) shows that specifying the functions q_{tj}, $t = 1, \ldots, 4$ and $j = 1, 2$, is tantamount to quantifying the influence of Y on the odds that patients drop out at clinic visit t and the odds that doctors prescribe prophylaxis therapy between visits $t - 1$ and t, after adjusting for all the recorded prognostic factors up to visit $t - 1$. Thus, q_{tj} determines how Y enters into a polytomous logistic regression model for the cause-specific probabilities of censoring. We refer to the functions q_{tj} as cause-specific selection-bias functions. Because by specifying the functions q_{tj} we do not place any restrictions on the distribution of the observed data $\mathbf{O}$, restriction (2) with given q_{tj}'s determines a nonparametric model for the law of the observables. Models that make explicit assumptions

about the censoring mechanism are referred to in the missing data literature as selection models (Little and Rubin, 1987). In contrast, models that relate the distribution of the outcome in the uncensored population to that of the censored population are referred to as pattern-mixture models (Little, 1993, 1995). We have shown that a model that specifies just the selection-bias functions q_{tj} is a nonparametric model that has interpretation both as a selection model and as a sequential pattern-mixture model.

The following two key remarks provide the formal justification for the sensitivity analysis philosophy of this article. First, the functions q_{tj} are not identified because all choices of these functions are compatible with the law of the observed data. Thus, no statistical test can reject any specific choice of q_{tj}. Second, specification of the functions q_{tj}, $t = 1, \ldots, 4$ and $j = 1, 2$, identifies the treatment–arm-specific distribution of Y.

The first remark is in line with intuition. Because the selection-bias functions quantify the influences of unmeasured prognostic factors on the decisions of subjects to drop out and start prophylaxis, it would be scientifically unreasonable that they would be identified in the absence of further information about these factors. The second remark says that indeed we can identify the treatment-arm means once we postulate the selection-bias functions. Since the data contain no evidence about the selection-bias functions, we recommend repeating the analysis estimating the treatment means under a range of selection-bias functions judged plausible by the field expert. Our recommendation is not new. The importance of conducting several inferences rather than a single one has been noted by many authors (Little and Rubin, 1987; Little, 1994; Copas and Li, 1997). Sensitivity analyses conducted by making different unverifiable assumptions have been discussed in the missing-data literature by Baker, Rosenberger, and DerSimonian (1992), Nordheim (1984), Robins (1997), Rotnitzky, Robins, and Scharfstein (1998), Scharfstein, Rotnitzky, and Robins (1999), and Robins, Rotnitzky, and Scharfstein (2000) and in the competing risk literature by Klein and Moeschberger (1988), Slud and Rubinstein (1983), and Zheng and Klein (1995, 1996). Our proposal differs from these articles in that we allow for selection-bias functions that are cause specific. This methodology is particularly attractive in the ACTG 002 trial because there is no *a priori* reason to believe that unmeasured prognostic factors for Y influence identically the decision of patients to drop out of the study and of doctors to prescribe prophylaxis therapy.

Specifying that q_{tj} is equal to zero for all t and j is tantamount to assuming that the recorded data constitute all the prognostic factors for CD4 count that physicians use to prescribe prophylaxis therapy and that patients use to decide whether or not to return to the next clinic visit. Accordingly, when $q_{tj} = 0$ for all t and j, we say that censoring is explainable; otherwise, in accordance with the missing data terminology, we say that censoring is nonignorable. Explainable censoring is related to the assumption of no unmeasured confounders of Robins et al. (1992) and to a sequential version of Rosenbaum and Rubin's (1983) strong ignorability assumption. Technically, explainable censoring is not equivalent to the assumption that the data are missing at random (MAR), as defined by Rubin (1976). The MAR assumption says that subjects uncensored at time t and subjects censored at time t with the same recorded past have the same distribution of the entire current and future variables $(Y, \underline{V}_t)$, where $\underline{V}_t = (V_t, \ldots, V_4)$. Explainable censoring, on the other hand, specifies that, among subjects with the same recorded past, the population of subjects censored due to each specific cause at time t has the same distribution of the outcome of interest Y as that of the population of uncensored subjects at time t. We note however that it would be rare in practice that explainable censoring would hold but the MAR would not.

4. The Curse of Dimensionality and the Need for Additional Modeling

Inference about $\mathrm{E}(Y \mid X = x)$ under a model that assumes only that the functions q_{tj} are known requires estimation of unknown functions of $\bar{V}_t$. For example, since $\mathrm{E}(Y \mid X = x)$ is the functional of the law of the observed data,

$$\mathrm{E}(Y \mid X = x) = \mathrm{E}\left\{ \frac{I(R_4 = 0)Y}{\prod_{t=1}^{4}\left[1 + \sum_{j=1}^{2} \exp\left\{h_{tj}\left(\bar{V}_t, x\right) + q_{tj}\left(Y, \bar{V}_t, x\right)\right\}\right]} \,\middle|\, X = x \right\}, \quad (4)$$

then, when V_t is discrete for all t, its nonparametric maximum likelihood estimator (NPMLE) is calculated by evaluating the expectation under the empirical law of the observed data and replacing $h_{tj}(\bar{V}_t, x)$ with its NPMLE, $\hat{h}_{tj}(\bar{V}_t, x)$, obtained by replacing in (4) the population distributions by the empirical distributions. When $\bar{V}_t$ has at least one continuous component, then one needs to estimate h_{tj} using smoothing techniques. Unfortunately, when $\bar{V}_t$ has two or more continuous components, impractically large sample sizes will be required for $\hat{\mathrm{E}}(Y \mid X = x)$ to have an approximately centered normal sampling distribution with variance small enough to be of substantive use. This is so because, with moderate sample sizes, essentially no two units will have $\bar{V}_t$-vectors close enough to one another to allow the borrowing of information necessary for smoothing. This phenomenon is usually referred to as the curse of dimensionality (Huber, 1985; Robins and Ritov, 1997).

The curse of dimensionality brings an important practical limitation: in order to obtain well-behaved estimators of $\mathrm{E}(Y \mid X = x)$, we must place restrictions on the law of the observed data, **O**. Little (1995) and Hogan and Laird (1997a) review currently available methods placing strong paramet-

ric restrictions on the observed data law. These methods do not allow for cause-specific censoring mechanisms and can be broadly classified as being based on selection models or on pattern-mixture models. Here we summarize the modeling strategies and contrast the features of sensitivity analysis under the various approaches.

Parametric selection models. Diggle and Kenward (1994), Baker (1995), Molenberghs, Kenward, and Lesaffre (1997), and Fitzmaurice, Laird, and Zahner (1996) assume (i) a parametric model for the conditional distribution of C given $\bar{V}_5$ and X indexed by τ and (ii) a parametric model for the conditional distribution of $\bar{V}_5$ given X indexed by η, where τ and η are variation independent. A sensitivity analysis then requires that one vary both the assumed parametric forms on the censoring mechanism and on the full data law. The semiparametric selection models described below specify only models for the censoring probabilities, thus reducing the number of assumptions to vary in a sensitivity analysis.

Parametric or semiparametric pattern-mixture models. Rubin (1977), Herzog and Rubin (1983), Little (1993, 1994), and Little and Wang (1996) assume (i) a parametric, semiparametric, or nonparametric model for the conditional distribution of C given X indexed by a finite dimensional parameter τ and possibly an infinite dimensional parameter λ, (ii) parametric models for the conditional distributions of $\bar{V}_5$ given X and $C = t$, indexed by ξ_t, where (λ, τ) and $(\xi_1, \ldots, \xi_5)$ are variation independent parameters, and (iii) restrictions on $(\xi_1, \ldots, \xi_5)$ sufficient for identification of $f(Y \mid X)$. In most approaches, the identifying restrictions in (iii) are tantamount to parameterizing either (a) the ratios $f(Y \mid \bar{V}_t, X, C = t)/f(Y \mid \bar{V}_t, X, C = t+1)$ of the densities of subjects dropping out at two consecutive occasions and with the same recorded past or (b) the ratios $f(Y \mid \bar{V}_t, X, C = t)/f(Y \mid \bar{V}_t, X, C = 5)$ of the density of subjects dropping out at time t to the density of the completers with the same recorded past. These parameterizations have an important limitation for conducting sensitivity analysis. Specifically, there is no natural parameter value associated with explainable censoring (e.g., the ratios in either (a) or (b) being all equal to one does not imply explainable censoring). In contrast, under the sequential pattern-mixture model (2), the functions $q_{tj} = 0$ correspond to explainable censoring, therefore allowing the study of sensitivity of inferences to local departures from this assumption.

Random effects models. Random effects models make either the selection model assumptions (i) and (ii) (Follmann and Wu, 1995; Wu and Carroll, 1988; Schluchter, 1992; DeGruttola and Tu, 1994; Tsiatis, DeGruttola, and Wulfsohn, 1995) or the pattern-mixture assumptions (i)–(iii) (Wu and Bailey, 1988, 1989; Hogan and Laird, 1997b, 1998) except that they replace X with $X^* = (X, \beta)$, where β is a random effect. A sensitivity analysis then requires that one also vary the assumptions on the random effects distribution.

Semiparametric selection models. These models place parametric restrictions only on the functions h_{tj}, i.e., they assume $h_{tj}(\bar{V}_t, X) = h_{tj}(\bar{V}_t, X; \boldsymbol{\eta}^*)$, where $h_{tj}(\bar{V}_t, X; \boldsymbol{\eta})$ is a known function and $\boldsymbol{\eta}$ is an unknown finite dimensional parameter vector with true value $\boldsymbol{\eta}^*$. Thus, in contrast with the parametric selection model approach, the semiparametric selection model approach makes no assumptions about the law of $\bar{V}_5$. Because the semiparametric selection model restricts the law of the observed data, then, at least in principle, we can check our assumptions on h_{tj} and q_{tj}. However, since the functional form of q_{tj} is not identified when the functions h_{tj} are left unrestricted, our ability to check the validity of the assumed functional form of q_{tj} relies entirely upon the correct specification of h_{tj}. Thus, any model checking of q_{tj} is indeed tantamount to checking the assumed parametric form for h_{tj}. Therefore, rather than checking the assumed model for q_{tj}, we recommend that inference about the treatment-arm CD4 count means be conducted under various plausible selection-bias functions q_{tj} (still regarded as known in the analysis). In addition, since inference can be sensitive to misspecification of the model for h_{tj}, we recommend choosing a flexible parametric model for h_{tj}. In the Appendix, we show how to compute estimators of the treatment–arm-specific means under the semiparametric selection model when the selection-bias functions q_{tj} are regarded as known.

5. Analysis of the ACTG 002 Study

Using the estimating equations given in the Appendix in this section, we estimate the treatment–arm-specific CD4 means in trial 002 at week 32, say μ_x $x = 0, 1$, under the model that assumes that $q_{tj}(Y, \bar{V}_t, x) = \tau_j^{(x)} Y$ and $h_{tj}(\bar{V}_t, x) = h_{tj}(\bar{V}_t, x; \boldsymbol{\eta}^*)$, where

$$\begin{aligned} h_{tj}(\bar{V}_t, x; \boldsymbol{\eta}) = {} & \eta_{0tj}^{(x)} + \eta_{1j}^{(x)} \log\left(CD4_{(t-1)} + 1\right) \\ & + \eta_{2j}^{(x)} CD4_{(t-1)} + \eta_{3j}^{(x)} WBC_{(t-1)} \\ & + \eta_{4j}^{(x)} PCP_{(t-1)} + \eta_{5j}^{(x)} AZT_{(t-1)}, \end{aligned} \tag{5}$$

$\boldsymbol{\eta} = (\boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)})$, $\boldsymbol{\eta}^{(x)} = (\eta_{0tj}^{(x)}, \eta_{1j}^{(x)}, \eta_{2j}^{(x)}, \eta_{3j}^{(x)}, \eta_{4j}^{(x)}, \eta_{5j}^{(x)} : t = 1, \ldots, 5, j = 1, 2)$, and $\boldsymbol{\eta}^*$ denotes the true value of $\boldsymbol{\eta}$; i.e., we assume that, in each treatment arm, the cause-specific censoring probabilities follow separate polytomous logistic regression models with occasion-specific intercepts. The parameter $\tau_1^{(x)}$ ($\tau_2^{(x)}$) has the interpretation as the increase in the conditional log odds of dropping out (initiating prophylaxis) versus remaining on study between subjects in treatment arm x who differ by one in CD4 count at week 32. The values $\tau_j^{(x)} = 0$, $j = 1, 2$, correspond to explainable censoring in treatment arm x.

In our analyses, we regard the selection-bias parameters $\tau_j^{(x)}$, $j = 1, 2$ and $x = 0, 1$, as fixed and examine how our inferences change as they vary over a plausible range of values. In Figure 1, we present point estimates of the CD4 count means in the high-dose arm for values of $\tau_j^{(0)}$, $j = 1, 2$, ranging from -0.05 to 0.05. This range was chosen for illustration purposes only. In practice, we recommend that this range be selected in consultation with the field expert. In Figure 1, the estimated treatment-specific means are increasing functions of $\tau_1^{(0)}$ and $\tau_2^{(0)}$. This reflects the fact that larger values of these parameters imply that more subjects with large values of the (possibly) unobserved $CD4$ at week 32 either drop out or initiate prophylaxis therapy. Furthermore, for fixed values of the prophylaxis parameter $\tau_2^{(0)}$, the treatment-specific means are very sensitive to the drop-out parameter $\tau_1^{(0)}$. In contrast, for fixed values of $\tau_1^{(0)}$, the treatment-specific means are not

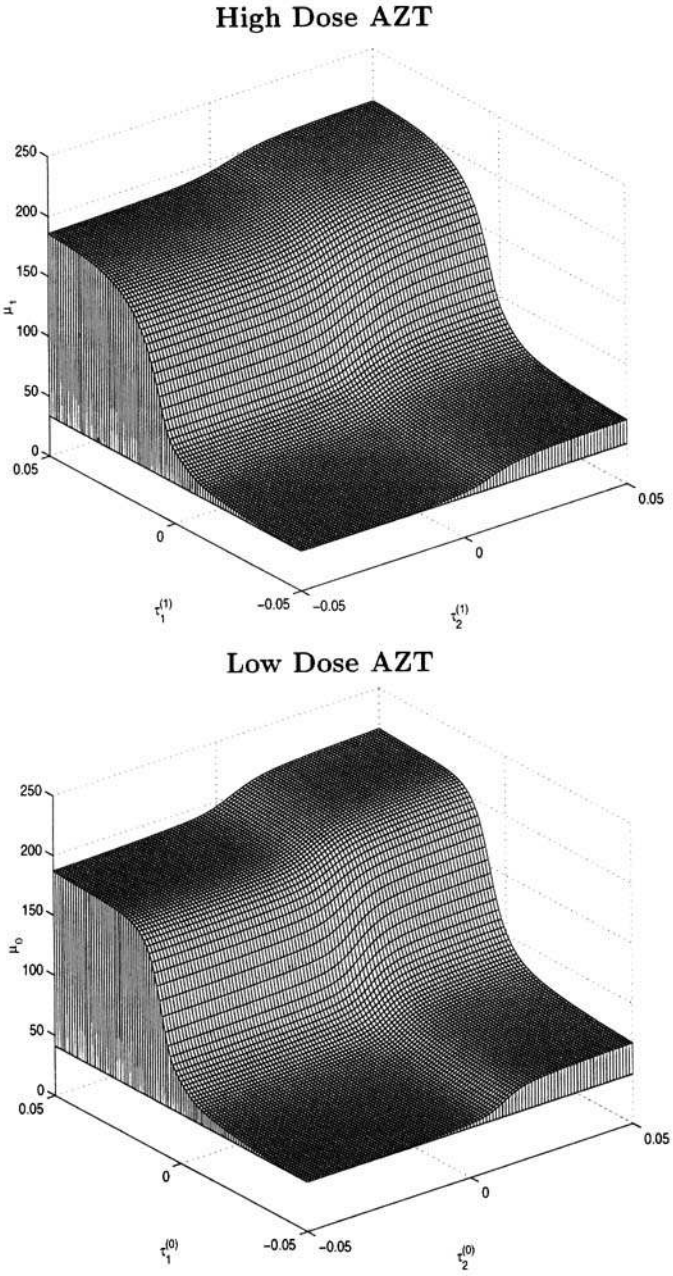

Figure 1. Estimated treatment-specific mean CD4 count at week 32, $\hat{\mu}_x$, as a function of the drop-out selection-bias parameter, $\tau_1^{(x)}$, and the prophylaxis drop-out parameter, $\tau_2^{(x)}$, under model (5).

very sensitive to $\tau_2^{(0)}$. This reflects the fact that, as shown in Table 1, drop-out is much more prevalent than initiation of prophylaxis.

In Figure 2, we investigate the effect of differential cause-specific selection bias in the two treatment arms. In this figure, on each plot we fixed the prophylaxis and drop-out parameters in the low-dose arm at one of the values $\tau_1^{(0)} = -0.02$, 0, or 0.02 and $\tau_2^{(0)} = 0$ or 0.02. On each plot, on the horizontal axis we varied the levels of the prophylaxis parameter $\tau_2^{(1)}$ and on the vertical axis we varied the levels of the drop-out parameter $\tau_1^{(1)}$ in the high-dose group. For each combination of selection-bias parameters, we performed a test (at the 0.05-level) of the null hypothesis H_0: $\mu_0 = \mu_1$ with the selection bias parameters regarded as fixed. Each plot in Figure 2 is a contour plot of the Z-statistic $(\hat{\mu}_1 - \hat{\mu}_0)/[\widehat{\text{var}}(\hat{\mu}_1 - \hat{\mu}_0)]^{1/2}$ as a function of the two selection-bias parameters in the high-dose arm. ($\hat{\mu}_j, j = 0, 1$, and $\widehat{\text{var}}$ were computed using the methods described in the appendix.) The lines in each plot represent the combination of selection-bias parameter values that lead to Z-statistics equal to 1.96 and -1.96. Above the 1.96 line, we would conclude that high-dose AZT was preferred to low-dose AZT; and below the -1.96 line, we would conclude that low-dose AZT was preferred to high-dose AZT. Between the two lines, there is insufficient evidence to reject the null hypothesis H_0 at the 0.05-level. The dot in each plot corresponds to explainable censoring in the high-dose arm, i.e., $\tau_1^{(1)} = \tau_2^{(1)} = 0$. The dot in the middle plot of the first column corresponds to the assumption of explainable censoring in the two treatment arms. The estimated CD4 count means in the two treatment arms under this assumption were 64.90 and 69.25 for the high- and low-dose AZT arms, respectively, but the difference did not reach statistical significance (p-value = 0.25). In contrast, the means among subjects uncensored at week 32 were 72.67 and 76.22 in the high- and low-dose arms, respectively (see Table 2). The lower means under explainable censoring are possibly reflective of the fact that censored subjects (i.e., drop-outs and prophylaxis users) tend to be sicker than those who remain uncensored. The plots in Figure 2 demonstrate that significant differential drop-out biases would have to occur in order to conclude that one AZT dose is preferred to the other. For example, the plots on the top row correspond to substantial nonignorable drop-out in the low-dose arm, with subjects with lower CD4 counts at week 32 being more likely than others to drop out. Under this scenario, we would reject H_0 and conclude that the high-dose arm is preferred only under the assumption that in the high-dose arm there is also substantial nonignorable drop-out but with opposite directionality to that in the low-dose arm.

To assess the dependence of the sensitivity analysis on the assumed model for h_{tj}, we have separately repeated the analysis under each of the two following more restrictive models for them:

$$h_{tj}(\bar{\boldsymbol{V}}_t, x; \boldsymbol{\eta}) = \eta_{0tj}^{(x)} + \eta_{1j}^{(x)} \log\left\{CD4_{(t-1)} + 1\right\} + \eta_{3j}^{(x)} WBC_{(t-1)} + \eta_{4j}^{(x)} PCP_{(t-1)} + \eta_{5j}^{(x)} AZT_{(t-1)}, \quad (6)$$

$$h_{tj}(\bar{\boldsymbol{V}}_t, x; \boldsymbol{\eta}) = \eta_{0tj}^{(x)} + \eta_{2j}^{(x)} CD4_{(t-1)} + \eta_{3j}^{(x)} WBC_{(t-1)} + \eta_{4j}^{(x)} PCP_{(t-1)} + \eta_{5j}^{(x)} AZT_{(t-1)}. \quad (7)$$

Figures 3 and 4 display the contour plots of the Z-statistics $(\hat{\mu}_1 - \hat{\mu}_0)/[\widehat{\text{var}}(\hat{\mu}_1 - \hat{\mu}_0)]^{1/2}$ under (6) and (7). The conclusions of the analysis under (7) are nearly identical as those under the less restrictive model (5). In contrast, the conclusions of the analysis under (6) differ from those under (5) in many regions of the selection-bias parameter space. Assuming that the larger model (5) is correct, for each value of the selection-bias parameters, we subsequently conducted a test that the

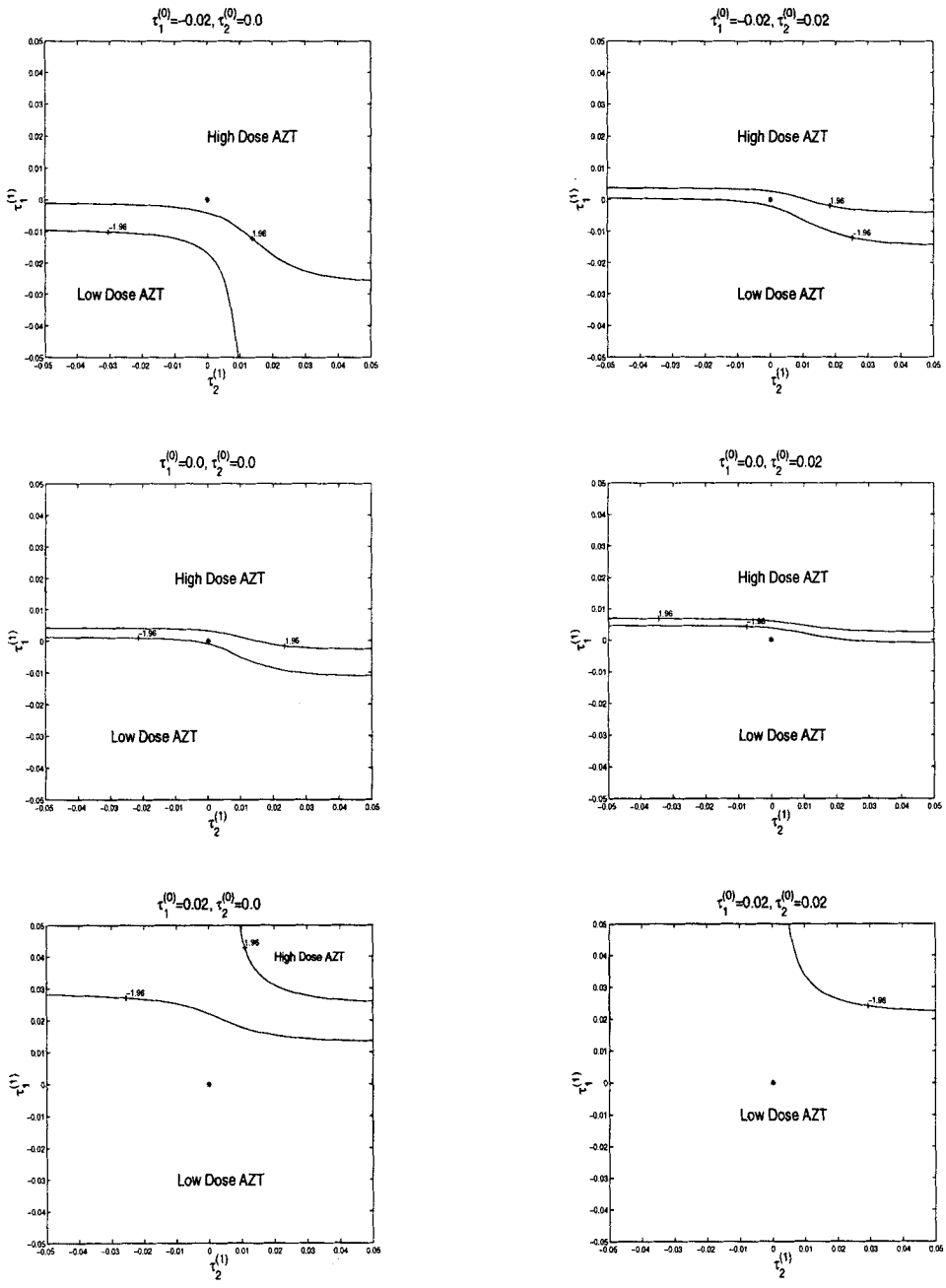

Figure 2. Contour plots of Z-statistic (under model (5)) comparing low-dose versus high-dose mean CD4 at week 32 as a function of $\tau_1^{(1)}$ and $\tau_2^{(1)}$ (high-dose cause-specific selection-bias parameters) for fixed values of $\tau_1^{(0)}$ and $\tau_2^{(0)}$ (low-dose cause-specific selection-bias parameters) ranging from -0.02 to 0.02 and 0.0 to 0.02, respectively.

smaller model (6) was true, i.e., a test that $\eta_{21}^{(1)} = \eta_{22}^{(1)} = \eta_{21}^{(2)} = \eta_{22}^{(2)} = 0$. Not surprisingly, the test rejected (results not shown) for most values of the selection-bias parameters in which the conclusions of the two analyses differed. In contrast, a test that model (7) is true did not reject over the entire selection-bias parameter region. This is in agreement with the point made in Section 4 that inference can be sensitive to misspecification of the model for the function h_{tj} and in line with our recommendation of choosing a flexible model for h_{tj}.

6. Final Remarks

In this article, we considered methods for conducting sensitivity analysis when two reasons for censoring are present. The semiparametric polytomous logistic regression selection models of this article can be trivially extended to settings with more than two causes for nonresponse.

In our analysis, we have disregarded data obtained on a subject subsequent to the first missed clinic visit or after the

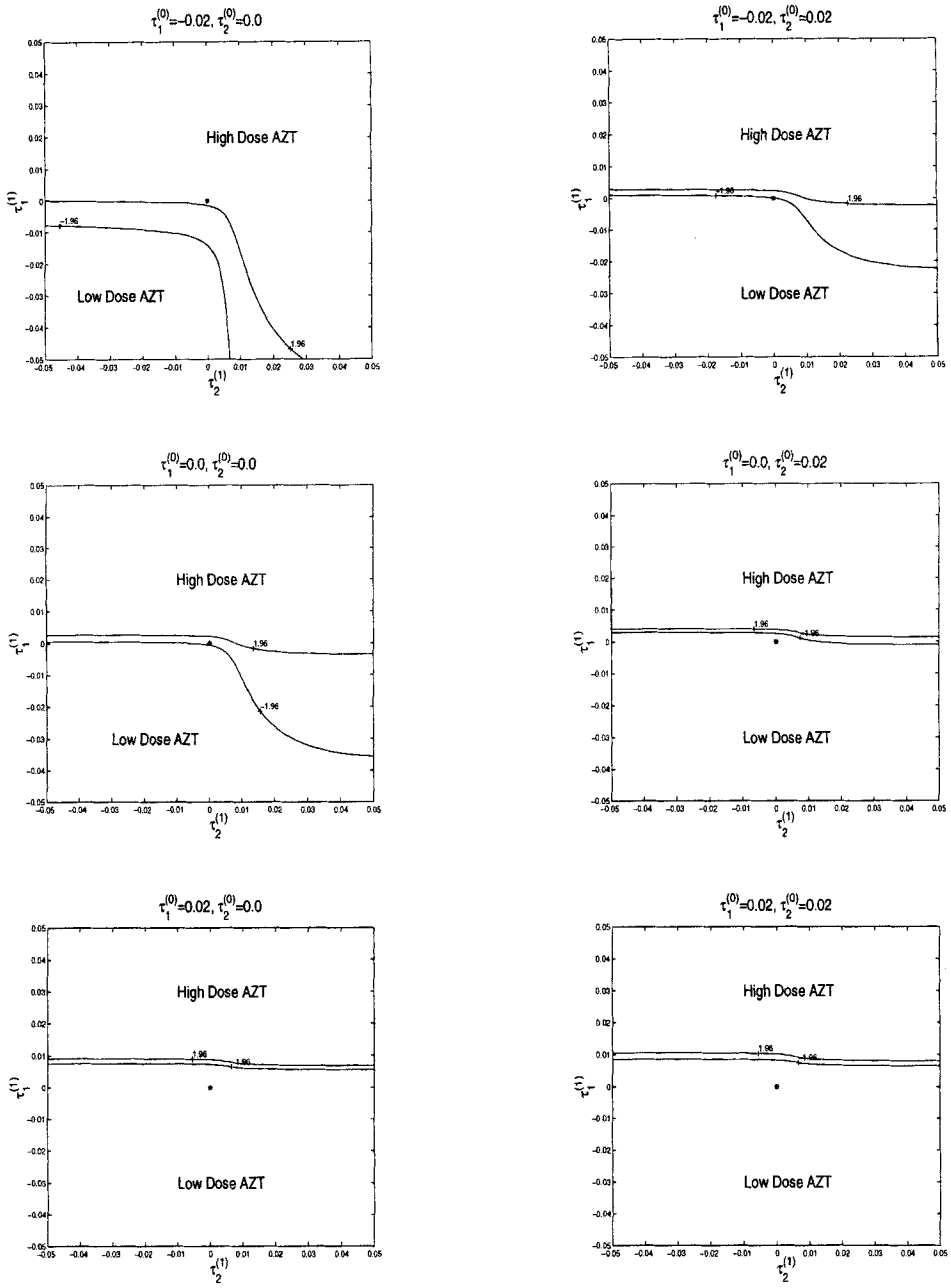

Figure 3. Contour plots of Z-statistic (under model (6)) comparing low-dose versus high-dose mean CD4 at week 32 as a function of $\tau_1^{(1)}$ and $\tau_2^{(1)}$ (high-dose cause-specific selection-bias parameters) for fixed values of $\tau_1^{(0)}$ and $\tau_2^{(0)}$ (low-dose cause-specific selection-bias parameters) ranging from -0.02 to 0.02 and 0.0 to 0.02, respectively.

patient initiates prophylaxis. Indeed, we have done so for robustness reasons. Specifically, arguing as in Rotnitzky et al. (1998), it can be shown that, if the conditional probabilities of not returning to the clinic and of continuing prophylaxis therapy, $P(R_t = j \mid R_{(t-1)}, \ldots, R_1, X, \bar{V}_t, Y)$, $j = 1, 2$, are left unrestricted for all $(R_{(t-1)}, \ldots, R_1) \neq (0, \ldots, 0)$ and are strictly less than one, then the data obtained on subjects after missing a visit or after initiating prophylaxis does not asymptotically provide information about the treatment-specific means. Since the above probabilities are not identified from the clinical trial data, we have chosen not to impose additional nonidentifiable modeling assumptions on these probabilities and therefore have disregarded the data obtained subsequent to a missed clinic visit.

Acknowledgements

Research was partially supported by National Institutes of Health grants: 1-R29-GM48704-0, 5-R01-A132475-07, R01-

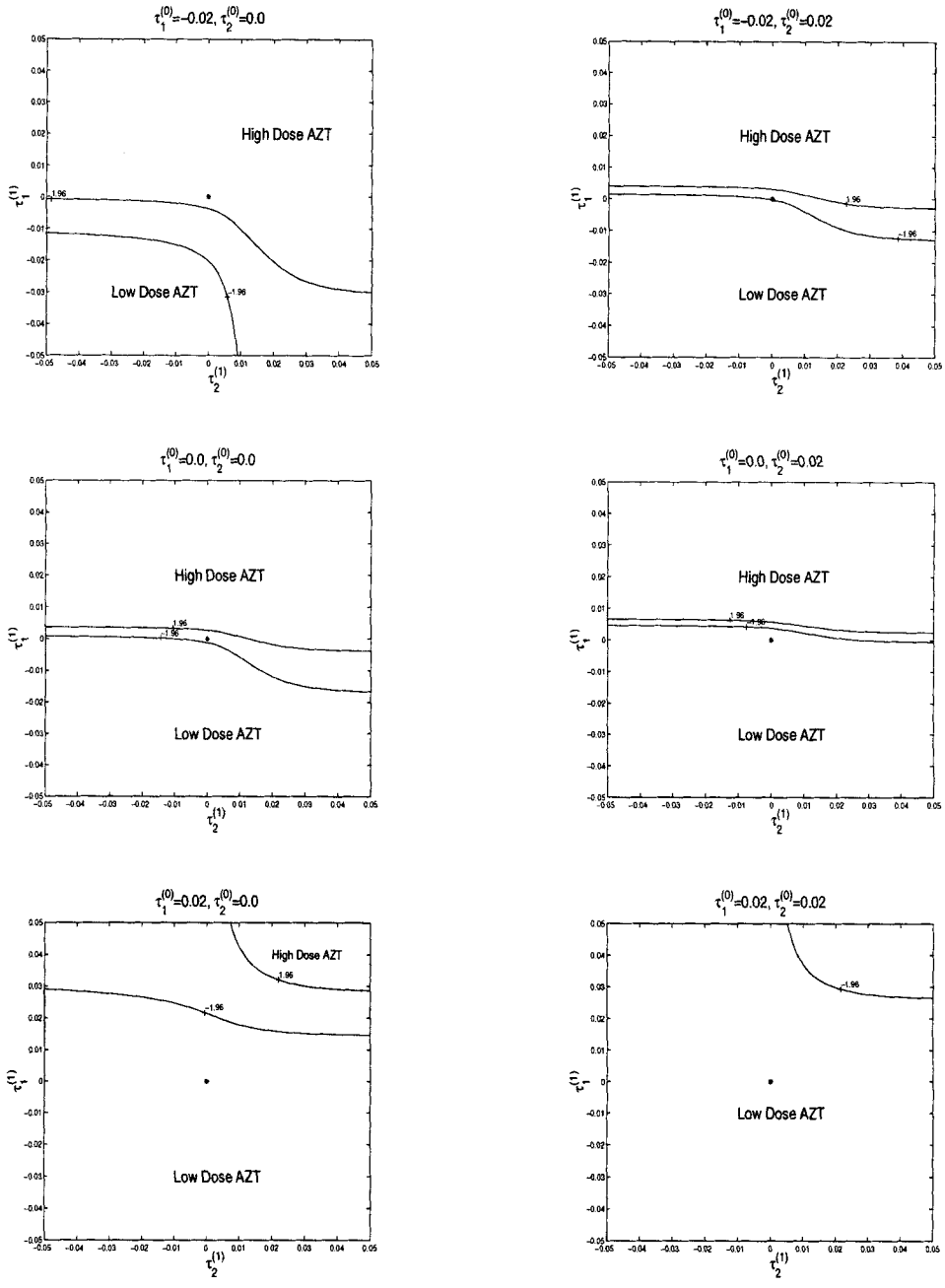

Figure 4. Contour plots of Z-statistic (under model (7)) comparing low-dose versus high-dose mean CD4 at week 32 as a function of $\tau_1^{(1)}$ and $\tau_2^{(1)}$ (high-dose cause-specific selection-bias parameters) for fixed values of $\tau_1^{(0)}$ and $\tau_2^{(0)}$ (low-dose cause-specific selection-bias parameters) ranging from -0.02 to 0.02 and 0.0 to 0.02, respectively.

CA74112, 1-R01-MH56639-01A1, R01-HD-38209-01, 1-R01-DA10184-01A2.

Résumé

Nous nous intéressons au test d'une différence moyenne entre groupes de traitement randomisés pour une réponse qui, si il n'y avait pas eu de patient ayant débuté un traitement autre que le traitement randomisé ou encore de patients sortis d'étude, aurait été observée à la fin de l'essai. Nous affirmons que la différence moyenne ne peut pas être estimée à moins de faire des suppositions invérifiables, qui reviennent à postuler des relations entre les composantes d'un modèle à mélange de populations, mais qui peuvent aussi être interprétées comme l'imposition, dans un modèle de sélection, de contraintes sur les probabilités de censure relatives à chaque motif de censure. Nous affirmons ensuite que, bien que suffisantes pour effectuer une estimation, ces suppositions ne permettent pas l'inférence proprement dite, en raison de la dimension du problème. Nous proposons de réduire la dimension du problème par la spécifi-

cation semi-paramétrique de modèles de censure, modèles relatifs à chaque motif possible de censure. Ces modèles peuvent être utilisés pour conduire des analyses de robustesse permettant d'étudier comment les variations, dans un intervalle plausible, du biais de sélection lié à un motif de censure affectent le test de la différence moyenne entre groupes de traitement. Nous décrivons la méthodologie d'une telle analyse de robustesse et l'illustrons sur les données d'un essai dans le sida (AIDS Clinical Trial Group étude 002).

References

Baker, S. G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* **51,** 1042–1052.

Baker, S. G., Rosenberger, W. F., and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* **11,** 643–657.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society, Series B* **59,** 55–77.

DeGruttola, V. and Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50,** 1003–1014.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43,** 49–73.

Fischl, M. A., Parker, C. B., Pettinelli, C., et al. (1990). A randomized controlled trial of a reduced daily dose of Zidovudine in patients with the acquired immunodeficiency syndrome. *The New England Journal of Medicine* **323,** 1009–1014.

Fitzmaurice, G. M., Laird, N. M., and Zahner, G. E. P. (1996). Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association* 91, 99–108.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51,** 151–168. Corrigendum: (1997), **53,** 384.

Herzog, T. N. and Rubin, D. B. (1983). Using multiple imputation to handle nonresponse in sample surveys. In *Incomplete Data in Sample Surveys: Theory and Bibliography,* Volume 2, W. G. Madow, H. Nisselson, I. Olkin (eds). New York: Academic Press.

Hogan, J. W. and Laird, N. M. (1997a). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16,** 259–272.

Hogan, J. W. and Laird, N. M. (1997b). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16,** 239–257.

Hogan, J. W. and Laird, N. M. (1998). Increasing efficiency from censored survival data using random effects to model longitudinal covariates. *Statistical Methods in Medical Research* **7,** 28–48.

Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics* **13,** 435–475.

Klein, J. P. and Moeschberger, M. L. (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics* **44,** 529–538.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88,** 125–134.

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81,** 471–483.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90,** 1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Little, R. J. A. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52,** 98–111.

Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* **84,** 33–44.

Nordheim, E. V. (1984). Inference from nonrandomly missing categorical data: An example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association* **79,** 772–780.

Robins, J. M. (1997). Non-response models for the analysis of non-monotone, non-ignorable missing data. *Statistics in Medicine* **16,** 21–38.

Robins, J. M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16,** 285–319.

Robins, J. M., Blevins, D., Ritter, G., and Wulfsohn, M. (1992). *g*-Estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology* **3,** 319–336.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials,* Elizabeth M. Halloran and Donald Berry (eds), 1–94. New York: Springer-Verlag.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93,** 1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63,** 581–590.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72,** 538–543.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94,** 1096–1146.

Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11,** 1861–1870.

Slud, E. V. and Rubinstein, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika* **70,** 643–649.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90,** 27–37.

Wu, M. C. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7,** 337–346.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45,** 939–955. Corrigendum: (1989), **46,** 889.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44,** 175–188. Corrigenda, (1989), **45,** 1347 and (1991), **47,** 357.

Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82,** 127–138.

Zheng, M. and Klein, J. P. (1996). Identifiability and estimation of marginal survival functions for dependent competing risks assuming the copula is known. In *Lifetime Data Models in Reliability and Survival Analysis,* N. P. Jewell (ed), 401–408. Dordrecht: Kluwer Academic.

Received May 1998. *Revised June* 2000.
Accepted June 2000.

Appendix

Augmented Inverse Probability of Cause-Specific Censoring Estimating Equations

In order to obtain Figures 1–4, for each fixed value of $\tau^{(x)} = (\tau_1^{(x)}, \tau_2^{(x)})$, we estimated $\eta^{(x)}$ and $\mu_x = \mathrm{E}(Y \mid X = x)$ separately in each treatment arm x. Thus, without loss of generality, we eliminate the superscript and subscript x from the parameters and consider the one-sample problem of estimating the CD4 count mean $\mathrm{E}(Y)$ in one treatment arm. We used the following extension of the methods proposed by Rotnitzky et al. (1998) and Scharfstein et al. (1999). For each fixed value of τ, the estimators $\hat{\eta}$ and $\hat{\mu}$ solve the joint system of estimating equations, $\Sigma_{i=1}^{n} \boldsymbol{U}(\boldsymbol{O}_i; \eta, \mu, \tau) = 0$, where the summation is over the subjects in the specific treatment arm being considered, $\boldsymbol{U}(\mathbf{O}; \eta, \mu, \tau) = (\boldsymbol{U}_1(\boldsymbol{O}; \eta, \tau)', U_2(\boldsymbol{O}; \eta, \mu, \tau)')'$,

$$U_2(\boldsymbol{O}; \eta, \mu, \tau) = \frac{I(R_4 = 0)}{\bar{\pi}_{10}(Y, \bar{\boldsymbol{V}}_4; \eta, \tau)}(Y - \mu),$$

$$\boldsymbol{U}_1(\boldsymbol{O}; \eta, \tau) = \sum_{j=1}^{2}\sum_{t=1}^{4} I(R_{(t-1)} = 0) \times \left\{ I(R_t = j) - \frac{I(R_4 = 0)\,\pi_{tj}(Y, \bar{\boldsymbol{V}}_t; \eta, \tau)}{\bar{\pi}_{t0}(Y, \bar{\boldsymbol{V}}_4; \eta, \tau)} \right\} \times \hat{\boldsymbol{\phi}}_{tj}(\bar{\boldsymbol{V}}_t),$$

$\hat{\boldsymbol{\phi}}_{tj}(\bar{\boldsymbol{V}}_t; \eta)$ is a possibly data-dependent vector function of the same dimension as η whose definition is given later in this section, $\bar{\pi}_{t0}(Y, \bar{\boldsymbol{V}}_4; \eta, \tau) = \Pi_{k=t}^{4} \pi_{k0}(Y, \bar{\boldsymbol{V}}_k; \eta, \tau)$, and $\log\{\pi_{kj}(Y, \bar{\boldsymbol{V}}_k; \eta, \tau)/\pi_{k0}(Y, \bar{\boldsymbol{V}}_k; \eta, \tau)\} = h_{kj}(\bar{\boldsymbol{V}}_k; \eta) + q_{kj}(Y, \bar{\boldsymbol{V}}_k; \tau)$. Under regularity conditions, $(\hat{\eta}, \hat{\mu})$ are consistent and asymptotically normal. A key to the consistency of $(\hat{\eta}, \hat{\mu})$ is the fact that $\boldsymbol{U}(\boldsymbol{O}; \eta, \mu, \tau)$ is an unbiased estimating function, i.e., $\mathrm{E}_{(\eta,\mu,\tau)}\{U(\boldsymbol{O}; \eta, \mu, \tau)\} = 0$ where the subscript (η, μ, τ) indicates that the expectation is taken with respect to the distribution with parameters (η, μ, τ).

We computed consistent estimators of the asymptotic variance of $(\hat{\eta}, \hat{\mu})$ using the sandwich variance estimator

$$n\left\{\sum_{i=1}^{n} \frac{\partial \boldsymbol{U}(\boldsymbol{O}_i; \hat{\eta}, \hat{\mu}, \tau)}{\partial(\eta', \mu)'}\right\}^{-1} \times \left\{\sum_{i=1}^{n} \boldsymbol{U}(\boldsymbol{O}_i; \hat{\eta}, \hat{\mu}, \tau)\,\boldsymbol{U}(\boldsymbol{O}_i; \hat{\eta}, \hat{\mu}, \tau)'\right\} \times \left\{\sum_{i=1}^{n} \frac{\partial \boldsymbol{U}(\boldsymbol{O}_i; \hat{\eta}, \hat{\mu}, \tau)}{\partial(\eta', \mu)}\right\}^{-1}.$$

In our analyses, we used for all values of τ_1 and τ_2 the functions

$$\hat{\boldsymbol{\phi}}_{tj}(\bar{\boldsymbol{V}}_t) = \boldsymbol{\varphi}_{tj}(\bar{\boldsymbol{V}}_t; \tilde{\eta}) - \sum_{m=1}^{t}\sum_{l=1}^{2} \pi_{ml}(Y, \bar{\boldsymbol{V}}_m; \tilde{\eta}; 0)\,\boldsymbol{\varphi}_{ml}(\bar{\boldsymbol{V}}_m; \tilde{\eta}),$$

where $\boldsymbol{\varphi}_{tj}(\bar{\boldsymbol{V}}_t; \eta) = \partial h_{tj}(\bar{\boldsymbol{V}}_t; \eta)/\partial \eta$ and $\tilde{\eta}$ is the maximum likelihood estimate for η when $\tau_1 = \tau_2 = 0$. This choice gives the semiparametric efficient estimator for μ when $\tau_1 = \tau_2 = 0$.

Proof That Relationship (2) Identifies $f(Y \mid X)$

We argue by reverse induction. The law $f(Y \mid R_4 = 0, \bar{\boldsymbol{V}}_4, X)$, being a law in an uncensored population, is identified. Now suppose that $f(Y \mid R_t = 0, \bar{\boldsymbol{V}}_t, X)$ is identified. Then $f(Y \mid R_t - R_{(t-1)} = j, \bar{\boldsymbol{V}}_t, X)$ is also identified by relationship (2). But then $f(Y \mid R_{(t-1)} = 0, \bar{\boldsymbol{V}}_{t-1}, X)$ is identified because it equals $\Sigma_{j=0}^{2} P(R_t = j \mid R_{(t-1)} = 0, \bar{\boldsymbol{V}}_{t-1}, X) \int f(Y \mid R_t - R_{(t-1)} = j, \bar{\boldsymbol{V}}_t, X) f(\boldsymbol{V}_t \mid R_t - R_{(t-1)} = j, \bar{\boldsymbol{V}}_{t-1}, X) d\boldsymbol{V}_t$ and $P(R_t = j \mid R_{(t-1)} = 0, \bar{\boldsymbol{V}}_{t-1}, X)$ and $f(\bar{\boldsymbol{V}}_t \mid R_t - R_{(t-1)} = j, \bar{\boldsymbol{V}}_{t-1}, X)$ are identified by the observed data. Finally, $f(Y \mid X)$ is identified by equation (1).