Summary

In many prospective studies, subjects are evaluated for the occurrence of an absorbing event of interest (e.g., HIV infection) at baseline and at a common set of pre-specified visit times after enrollment. Since subjects often miss scheduled visits, the underlying visit of first detection may be interval censored, or more generally, coarsened. Interval-censored data are usually an-
alyzed using the non-identifiable coarsening at random (CAR) assumption. In some settings, the visit compliance and underlying event time processes may be associated, in which case CAR is violated. To examine the sensitivity of inference, we posit a class of models that express deviations from CAR. These models are indexed by non-identifiable, interpretable parameters, which describe the relationship between visit compliance and event times. Plausible ranges for these parameters require eliciting information from scientific experts. For each model, we use the EM algorithm to estimate marginal distributions and proportional hazards model regression parameters. The performance of our method is assessed via a simulation study. We also present analyses of two studies: AIDS Clinical Trial Group (ACTG) 181, a natural history study of cytomegalovirus shedding among advanced AIDS patients, and AIDS Link to the Intravenous Experience (ALIVE), an observational study of HIV infection among intravenous drug users. A sensitivity analysis of study results is performed using information elicited from substantive experts who worked on ACTG 181 and ALIVE.

Key Words: Survival Analysis; Informative Censoring; Interval Censoring; Coarsening at Random; Sensitivity Analysis.

1 Introduction

In many prospective studies, participants are evaluated for the occurrence of an absorbing event of interest (e.g., HIV infection) at baseline and at a common set of pre-specified visit times after enrollment. Since the exact time
of occurrence is never known, it is useful to focus inference on the distribution of the underlying visit time of first detection. This distribution has support at each of the planned visit times as well as a point indicating non-occurrence over the follow-up period.

Since participants often miss scheduled visits, the underlying visit of first detection may be known (1) exactly, (2) to potentially have occurred before the initial scheduled visit, (3) to occur either after the last attended visit or not at all, or (4) to fall between two visit times. Thus, the observable data are a combination of exact observance, left, right, and interval censoring, which can all be mathematically formulated under the rubric of interval censoring (Lindsay and Ryan, 1998) and even more generally as coarsened data (Heitjan, 1993).

Interval-censored data are usually analyzed by assuming that the interval that includes the true underlying visit of first detection arises from a random mechanism independent of censoring (see, for example, Peto, 1973; Turnbull, 1976; Finkelstein and Wolfe, 1985, 1986; Finkelstein, 1986; Tu et al., 1993; Gentleman and Geyer, 1994; Alioum and Commenges, 1996; Sun, 1997; Goggins et al., 1998; Pan, 2000; and references therein). This non-informative censoring assumption means that knowing the outcome was censored due to missed study observations provides no more information about the event process than knowing the interval in which the event occurred. Such an assumption is a special case of the coarsening at random (CAR) assumption discussed in Heitjan and Rubin (1991); Heitjan (1993, 1994); and Gill, van der Laan, and Robins (1997). There has been some limited inferential work for testing CAR using auxiliary data (Betensky and Finkelstein, 2002) and
estimation where CAR is not assumed (NCAR) (Finkelstein, Goggins, and Schoenfeld, 2002).

When CAR is assumed, the analysis of interval-censored data requires special methods. Turnbull (1976) used the EM algorithm (Dempster, Laird, and Rubin, 1977) to extend the Kaplan-Meier (1958) survivor curve estimator to allow for general censoring. Tu, Meng, and Pagano (1993) used the EM algorithm to extend discrete-time proportional hazards regression analysis using the discrete Cox model (Cox, 1972; Prentice and Gloeckler, 1978) to allow general censoring and truncation, while Sun (1997) used maximum likelihood to extend the continuation ratio model (CRM) (McCullagh, 1980) to allow interval censoring.

While assuming CAR is computationally convenient, the assumption is usually untestable and is often considered implausible by scientific experts. In this paper, we consider two observational studies, AIDS Clinical Trial Group (ACTG) 181 and AIDS Link to Intravenous Experience (ALIVE). Co-author Samuel A. Bozzette was the principal investigator of ACTG 181, and co-authors David Vlahov and Noya Galai were principal investigator and lead statistician of ALIVE, respectively. In these studies, these experts believe that the nature of missed clinic visits relates to the outcome of interest. A CAR-based analysis of these data would produce biased results. Thus, studies like ACTG 181 and ALIVE, where an absorbing time-to-event outcome is interval censored due to subject non-response, require statistical methods that can relax the CAR assumption. Since the true departure from CAR cannot be identified from the data, the statistical methodology should incorporate scientific expertise into the model. This approach allows investigators...
to evaluate the sensitivity of conclusions to varying assumptions about the relationship between the visit compliance process and the outcome under investigation.

The first goal of this paper is to extend Turnbull’s (1976) one-sample estimation approach by relaxing the CAR assumption through the use of expert opinion about the degree to which the visit compliance process is associated with the underlying study visit of first detection. Similarly, our second goal is to estimate the regression parameters of the discrete-time proportional hazards CRM under relaxations of CAR.

The paper is organized as follows. Section 2 introduces the data structure. Section 3 discusses CAR in the context of interval censoring, while Section 4 introduces a class of ”exponential tilt” models, which are interpretably indexed by non-identifiable parameters that measure the degree of departure from CAR. Section 5 discusses the one-sample and CRM likelihoods and proposed estimation procedures using the EM algorithm, respectively. The performance of our approach is evaluated in Section 6 via simulation studies. Section 7 presents analyses of the ACTG 181 and ALIVE studies. For each study, we describe our method for eliciting expert opinion about the exponential tilt parameters. The final section presents a discussion, including extensions of our approach and directions for future research.

2 Data Structure

Suppose that there are $M + 1$ planned study visits at times $0 = t_0 < t_1 < t_2 < \ldots < t_M$. Let $T$ denote the underlying study visit of first detection if the
absorbing event occurs before or during the study, otherwise let \( T = t_{M+1} \), where \( \infty > t_{M+1} > t_M \). So \( E = \{t_j : j = 0, \ldots, M + 1\} \) is the support of the distribution of \( T \). Due to possible skipped visits, the observed data for an individual is just a subset of \( E \). In particular, observed data can be represented as \([L, R] = \{t \in E : L \leq t \leq R\}\). The set \([L, R]\) is a coarsening of \( T \) because \( T \in [L, R] \). Note that if \( L = R = t_0 \), then the event happened prior to entry, if \( L = R < t_M \) then the visit of first detection is observed, if \( L = R = t_{M+1} \) then the event did not occur during the planned follow-up period, and if \( L < R \) then there is incomplete knowledge about \( T \). Those with \( L < R = t_{m+1} \) are drop-outs, and those with \( L < R < t_{m+1} \) are returners. For the regression setting, we define \( Z \) to be a fully-observed low-dimensional covariate. We assume that, within values of \( Z \), we observed \( n_z \) i.i.d. copies of the observed data. Where appropriate, the subscript \( i \) will denote subject-specific data. Hereafter, \( T \) and its distribution will be referred to as the “event time” and “event process,” respectively. Similarly, the distribution of \( \{L, R\} \) will be the “coarsening process.”

3 Coarsening at Random

In the one-sample context, CAR means that

\[
P(L = l, R = r|T = t) \text{ is constant in } t \in [l, r]
\]
for all \([l, r] \in E^* = \{[l, r] : l \leq r, l, r \in E\}\). Gill, van der Laan, and Robins (1997) show that CAR is equivalent to the following two statements:

\[
P(T = t | L = l, R = r, T \in [l, r]) = P(T = t | T \in [l, r])
\]

and

\[
P(T = t | L = l, R = r) = P(T = t | T \in [l, r]) \quad \text{for all } t \in [l, r],
\]

whatever be \([l, r] \in E^*\).

Notice that the left-hand sides of Equations (1) and (2) are functions of both the event and coarsening processes, while the right-hand sides of the equations only depend on the event process (the process of scientific interest). Therefore, CAR means that the coarsening process provides no information about the event process beyond knowing that the true event time is in a particular interval.

Gill, van der Laan, and Robins (1997) show that, given the random set \([L, R]\) \((L \leq T \leq R)\), one can always construct a unique event-time random variable \(T\), such that \(T \in [L, R]\) and CAR holds. This result shows that, under CAR, the distribution of \(T\) is uniquely identified and that CAR itself does not place any restrictions on the distribution of the observed data. The unique solution can be found by using the self-consistency equation of Turnbull (1976) in the one-sample setting. The preceding result holds within levels of \(Z\) in the regression setting, and estimation can be performed using the approaches of Tu, Meng, and Pagano (1993) or Sun (1997).
4 NCAR Models, Identifiability, Sensitivity Analysis

Since CAR is usually untestable and often scientifically implausible, it is useful to consider alternative coarsening models. We proceed by positing a class of NCAR models, indexed by a non-identifiable censoring bias function that quantifies the relationship between the event and coarsening processes. For each model in our class, we show (under regularity conditions) that the marginal distribution of $T$ is identified from the observed data. As the censoring bias function is not identifiable, we recommend that one draw inference over a range of these models. To make such a sensitivity analysis feasible, we suggest that the censoring bias function be parameterized via a small set of interpretable censoring bias parameters.

4.1 NCAR Models

In the spirit of Rotnitzky, Scharfstein, Su, and Robins (2001); Scharfstein, Daniels, and Robins (2003); and Birmingham, Rotnitzky, and Fitzmaurice (2003); we posit the following class of pattern-mixture models for the marginal case:

$$P(T = t|L = l, R = r) = \frac{P(T = t|T \in [l, r]) \exp\{q(t, l, r)\}}{c(l, r; q)},$$  \hspace{1cm} (3)$$

where

$$c(l, r; q) = \sum_{s \in [l, r]} P(T = s|T \in [l, r]) \exp\{q(s, l, r)\},$$
and \( q(t, l, r) \), the model index, is a specified function of \((t, l, r)\). For each \( q \), the model is an exponential tilting (Barndorff-Nielsen and Cox, 1989) of the CAR model. If \( q(t, l, r) \) does not depend on \( t \), then no tilting is performed, and CAR is satisfied.

Using Bayes’ rule, it is possible represent (3) as a selection model of the following form:

\[
\log \left\{ \frac{P(L = l, R = r | T = t)}{P(L = l, R = r | T \in [l, r])} \right\} = d(l, r; q) + q(t, l, r) \text{ for } t \in [l, r], \quad (4)
\]

where

\[
d(l, r; q) = -\log\{c(l, r; q)\}.
\]

The selection model (4) implies that

\[
\log \left\{ \frac{P(L = l, R = r | T = t)}{P(L = l, R = r | T = t')} \right\} = q(t, l, r) - q(t', l, r) \text{ for } t, t' \in [l, r]. \quad (5)
\]

From (5), we see that \( q(t, l, r) \) is the difference in the log probability of having interval \([l, r]\) comparing a subject with \( T = t \) to a subject with \( T \) equal to some reference value, \( t_{ref} \), such that \( q(t_{ref}, l, r) = 0 \).

### 4.2 Identifiability

To show that the distribution of \( T \) is uniquely identified in the one-sample setting (from the population distribution of the observed data \( P(L = l, R = r) \))
under model (3,4), we note that for $t \in E$,

$$P[T = t] = \sum_{[l,r] \in E^*} P(T = t|L = l, R = r)P(L = l, R = r)$$

$$= \sum_{[l,r] \in E^*} \frac{P(T = t|T \in [l,r]) \exp\{q(t, l, r)\}}{c(l, r; q)} P(L = l, R = r)$$

$$= \sum_{[l,r] \in E^*} \frac{P(T = t) \exp\{q(t, l, r)\}}{\sum_{s \in [l,r]} P(T = s) \exp\{q(s, l, r)\}} P(L = l, R = r). \quad (6)$$

The system of $M + 2$ equations represented by (6) are linearly dependent. By removing one equation, the resulting system of $M + 1$ equations will be linearly independent. Thus, we have a system of $M + 1$ linearly independent equations and unknowns. The existence of a unique solution, which is a proper probability mass function for $T$, is proved in the appendix.

### 4.3 Sensitivity Analysis

Theorem 1 in the appendix tells us that given $q$ and model (3,4), we can non-parametrically identify the distribution of $T$ from the distribution of the observed data. That is, knowledge of $q$ in the context of model (3,4) does not place any restrictions on the distribution of the observed data. Thus, the function $q$ is not identifiable, and the best that can be achieved is to perform a sensitivity analysis with respect to $q$. In later sections, we will discuss parameterizations of $q$ and a method for eliciting the parameters from subject-matter experts.
4.4 Low-Dimensional Covariate

In the presence of a covariate, $Z$, we would like to estimate $P(T = t \mid Z = z)$ for all $t \in E$. We start by noting that

$$P(T = t \mid Z = z) = \sum_{[l, r] \in E^*} P(T = t \mid L = l, R = r, Z = z)P(L = l, R = r \mid Z = z).$$

(7)

Typically, CAR is assumed within levels of $Z$. That is, Equations (1) and (2), conditioned on $Z = z$, hold.

Under CAR, $P(T = t \mid Z = z)$ is identified by replacing $P(T = t \mid L = l, R = r, Z = z)$ with $P(T = t \mid T \in [l, r], Z = z)$ in Equation 7. In order to loosen the assumption of CAR within levels of $Z$, we posit a level-specific class of pattern-mixture models as in Section 4.1. In addition, $q$ may also be a function of $z$.

The result in the appendix implies that the distribution of $T$ given $Z = z$ is identifiable. Since the proportional hazards assumption reduces the number of free parameters, the data contain enough information to estimate the regression parameters.

5 Likelihood and Estimation

In this section, we describe the likelihoods and estimation procedures for two different objectives. Section 5.1 focuses on estimating marginal survival curves, while Section 5.2 describes estimation for proportional hazards regression. In both settings, estimation is performed via the EM algorithm.
Standard errors are estimated using Louis’s (1982) method.

In discrete time, the complete event-time data are assumed to follow the multinomial distribution. The following notation will be useful in both the marginal and regression settings. Let \( I_{ij} = I(T_i = t_j) \). The \( I_{ij} \)'s are the complete data. Let \( \delta_{ij} = I(t_j \in [L_i, R_i]) \). The \( \delta_{ij} \)'s are the observed data, and the vector \( \delta_i = \{\delta_{i0}, \ldots, \delta_{iM}, \delta_{i(M+1)}\} \) contains the same information as \( \{L_i, R_i\} \).

Let \( p_j = P(T = t_j) \) for \( j = 0, \ldots, M + 1 \) and \( p = (p_0, \ldots, p_M, p_{M+1}) \). We assume that each of the event times \( j \) have at least one exactly observed event occurring at that time. To conform with this regularity condition, one can discretize the event times to be the smallest disjoint intervals as in Turnbull (1976), but can also be done according to some convenient measure of time (months, years, etc.).

### 5.1 Marginal Distribution Estimation

In the one-sample setting, the complete-data likelihood is

\[
L(p) = \prod_{i=1}^{n} \prod_{j=0}^{M+1} p_j^{I_{ij}}.
\]

The E-step at the \( s \)th iteration is

\[
Q(p; p^{(s-1)}) = \sum_{i=1}^{n} \sum_{j=0}^{M+1} I_{ij}^{(s-1)} \log(p_j) - \lambda \left( \sum_{j=0}^{M+1} p_j - 1 \right),
\]

where \( \lambda \) is a Lagrange multiplier, \( p^{(s-1)} \) is the estimate of \( p \) at the \((s - 1)\)th iteration, and
\[ I_{ij}^{(s-1)} = E \left( I_{ij} | \delta_i, P^{(s-1)} \right) \]
\[ = P(T_i = t_j | \delta_i, P^{(s-1)}) \]
\[ = \frac{\delta_{ij} p_j^{(s-1)} \exp\{q(t_j, l, r)\}}{\sum_{k=0}^{M+1} \delta_{ik} p_k^{(s-1)} \exp\{q(t_k, l, r)\}}. \]

The M-step results in a reweighted version of Turnbull’s (1976) self-consistency equation:

\[ p_j^{(s)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{ij} p_j^{(s-1)} \exp\{q(t_j, l, r)\}}{\sum_{k=0}^{M+1} \delta_{ik} p_k^{(s-1)} \exp\{q(t_k, l, r)\}}. \]

When \( q \) does not depend on \( t \) (ie, \( q = 0 \)), CAR holds, and our estimator simplifies to that of Turnbull (1976).

### 5.2 Proportional Hazards Regression Estimation

Let \( \rho_{ij} = P(T_i = t_j | T_i \geq t_j) \) for \( j = 0, ..., M + 1 \). Using a variation of the logit link, the continuation ratio model is

\[ \log \left( \frac{\rho_{ij}}{1 - \rho_{ij}} \right) = \theta_j + \beta Z_i, \quad j = 0, \cdots, M, \]

where \( Z_i \) is a covariate for person \( i \), and \( \beta \) is the log hazard ratio comparing those whose covariate differs by one unit. The modeled relative hazard is constant over time. \( \theta = \{\theta_0, \ldots, \theta_M\} \) are the log hazards when \( Z_i = 0 \). Let \( R_{ij} = \sum_{k=0}^{j} I_{ik} \), where \( R_i(M+1) \equiv 1 \). The complete-data likelihood can be expressed as
L(\(\theta, \beta\)) = \prod_{i=1}^{n} \left\{ (\rho_{i0})^{I_{i0}} (1 - \rho_{i0})^{1 - R_{i0}} \right\} \ldots \left\{ (\rho_{iM})^{I_{iM}} (1 - \rho_{iM})^{1 - R_{iM}} \right\},

a product of binomials. The E-step at the sth iteration is

\[
Q(\theta, \beta; \theta^{(s-1)}, \beta^{(s-1)}) = \sum_{i=1}^{n} \sum_{j=0}^{M+1} I_{ij}^{(s-1)} \log(\rho_{ij}) + (1 - R_{ij}^{(s-1)}) \log(1 - \rho_{ij}),
\]

where

\[
I_{ij}^{(s-1)} = \mathbb{E} \left( I_{ij} | \delta_i, \theta^{(s-1)}, \beta^{(s-1)}, Z_i = z \right)
\]

\[
= P(T_i = t_j | \delta_i, \theta^{(s-1)}, \beta^{(s-1)}, Z_i = z)
\]

\[
= \frac{\delta_{ij} p_{ij}^{(s-1)} \exp\{q(t_j, l, r, z)\}}{\sum_{k=0}^{M+1} \delta_{ik} p_{ik}^{(s-1)} c_i^{(s-1)}(l, r, z; q)},
\]

and

\[
R_{ij}^{(s-1)} = \mathbb{E} \left( R_{ij} | \delta_i, \theta^{(s-1)}, \beta^{(s-1)}, Z_i = z \right) = \sum_{k=0}^{M+1} I_{ij}^{(s-1)}.
\]

For this model, \(p_{i0}^{(s-1)} = \rho_{i0}^{(s-1)}\) and \(p_{ij}^{(s-1)} = \rho_{ij}^{(s-1)} \prod_{k=1}^{j-1} (1 - \rho_{ik}^{(s-1)}) \), \(j = 1, \ldots, M + 1\), where

\[
\rho_{ij}^{(s-1)} = \frac{\exp(\theta^{(s-1)}_j + \beta^{(s-1)} z_i)}{1 + \exp(\theta^{(s-1)}_j + \beta^{(s-1)} z_i)}.
\]

The M-step for finding \(\{\theta^{(s)}, \beta^{(s)}\}\) requires a numerical technique such as Newton-Raphson.

6 Simulation Study

Simulations were performed in the regression setting with a variety of true and modeled censoring bias parameters. In each analysis, \(M = 4\) (the baseline visit, four follow-up visits, and right censoring/no event). A single binary
covariate, $Z$, was considered where $n_z = 100$, $z = 0, 1$. The censoring bias function used in these simulations is

$$q(\phi, t, l, r, z) = \phi^z \frac{(t - l)}{(r - l)} , \ z = 0, 1.$$  

The parameters $\phi^z$ are interpreted as the log probability ratio of having censoring interval $[l, r]$ comparing those whose true event time is $r$ to those whose true event time is $l$, among those with $Z = z$.

True event times, $T$, were drawn given $Z$, with $\beta \in \{0, 0.75\}$ and $\theta = \{-0.65, -0.55, -0.45, -0.15, -0.05\}$. Censoring intervals were drawn given $T$ and $Z$. The pattern-mixture restrictions in Section 4.1 and the distribution of $T$ given $Z$ are not enough to fully identify the distribution of the censoring intervals given $T$ and $Z$. The number of free parameters in this distribution is $\frac{(M+2)(M+1)}{2}$, the number of intervals minus the number of event times. These parameters (interval probabilities) were fixed at values that satisfy the constraints $P(T = t) > P(T = t, L = l, R = r)$. The strict inequality allows positive probability for each interval in which $t$ lies. The remaining $M + 2$ interval probabilities were identified from the constraints $\sum_{t \leq r} P(L = l, R = r) = 1$ and

$$\sum_{\{t, r\} : l \leq t \leq r} P(T = t | L = l, R = r) P(L = l, R = r) = P(T = t).$$  

Values of the true parameters were chosen to produce between 86% and 97% censoring (the probability of not exactly observing the event), depending on $\phi_{true}$ and $\beta$. 

15
Let $\phi = \{\phi^z : z = 0, 1\}$ be the censoring bias parameters. The true and modeled censoring bias parameters, $\phi_{\text{true}}$ and $\phi_{\text{modeled}}$, respectively, were combinations of $\{-\log(2), 0, \log(2)\}$ where $\phi_{\text{true}}^0 < \phi_{\text{true}}^1$ and $\phi_{\text{modeled}}^0 < \phi_{\text{modeled}}^1$. Analyses were performed with $\beta = 0$ and $\beta = 0.75$. We generated 1000 datasets for each of the 72 analyses.

Tables 1 and 2 present simulation results. The degree of bias is assessed by comparing the mean of the estimated $\beta$’s to the true value. The performance of 95% Wald-type confidence intervals is explored by determining the actual coverage probability. Also, the Monte Carlo standard deviation of the estimate of $\beta$, $\hat{\beta}$, is compared to the mean of the standard error estimates. The results show excellent coverage and good agreement between the empirical standard deviation and the standard error estimate when the censoring bias parameters are correctly specified. As expected, coverage and standard error estimates are sensitive to misspecification of these parameters. When $\beta = 0.75$ (Table 2), $\hat{\beta}$ shows moderate bias when $\phi_{\text{true}} = \log(2)$. This specification was rerun with $n_0 = n_1 = 200$, which resulted in a bias of 0.02 and coverage of 0.96.

7 Data Examples

7.1 ACTG 181

ACTG 181 is a natural history study of advanced HIV disease. Patients in this study were scheduled to be monitored for CMV shedding in the urine every four weeks and in the blood every 12 weeks, as discussed in Betensky and Finkelstein (1999). ACTG 181 is a substudy of ACTG 081, which is
described in Bozzette et al. (1995).

An important scientific question is to determine the relative risk of CMV shedding comparing those with high baseline CD4 counts (at least 200 cells/mm$^3$) to those with low baseline CD4 counts. Censoring due to missed visits is thought to be informative, as healthier patients may behave differently from sicker patients. In addition, behavior may depend on baseline CD4 count.

The data consist of 204 patients whose CMV shedding time was discretized into four three-month quarters, baseline, and a time that indicates no shedding within the 12 months: \( \{0, 3, 6, 9, 12, >12\} \). At baseline, 69 patients had high CD4 counts, while the remaining 135 had low CD4 counts. Among those with high CD4 counts, 23\%, 10\%, 57\%, and 10\% were left, interval, and right censored, and exactly observed, respectively. The percentages among those with low CD4 counts are 36\%, 12\%, 41\%, and 11\%. There were no deaths in this population prior to 12 months.

Define \( Z = I(CD4 \geq 200) \). Here, \( M = 4 \). We defined the following censoring bias function for this analysis:

\[
q(\phi, t, l, r, z) = \phi_z I(r < M + 1) \frac{(t - l)}{(M - 1)} + \phi_z I(r = M + 1) \frac{(t - l)}{M}, \quad z = 0, 1
\]

where \( \phi = \{\phi_0, \phi_1 : z = 0, 1\} \). Using Bayes’ rule as in Section 4.1,

- \( \exp\{\phi_1\} \) is the CD4-specific probability ratio of having interval [3 months, 12 months] comparing those who begin shedding at 12 months to those who begin shedding at 3 months.

- \( \exp\{\phi_2\} \) is the CD4-specific probability ratio of dropping out just after baseline comparing those who do not begin shedding within 12 months
to those who begin shedding within 3 months from baseline.

When $\exp\{\phi_1\} > 1$ ($< 1$), returners are more (less) likely to shed late than shed early. When $\exp\{\phi_2\} > 1$ ($< 1$), drop-outs are more (less) likely to shed late or not at all than shed early.

To elicit the values of the censoring bias parameters, the schematic in Figure 1 was presented to Dr. Bozzette and he was asked, “Among those with CD4 counts at least 200 cells/mm$^3$, who is more likely to test negative for CMV shedding at baseline, miss several visits, then return at 12 months with CMV shedding: someone who began CMV shedding just after baseline or someone who began CMV shedding at 12 months?” Next, we referred to the schematic in Figure 1b and asked, “Among those with CD4 counts at least 200 cells/mm$^3$, who is more likely to test negative for CMV shedding at baseline, then drop out: someone who began CMV shedding just after baseline or someone who did not begin CMV shedding within 12 months?” The questions were repeated for those with CD4 counts less than 200 cells/mm$^3$. Dr. Bozzette believes that, among those with high baseline CD4 counts, those who begin shedding within 3 months of testing negative at baseline are more likely to drop out (be interval-censored) than those who begin shedding after (at) 12 months. He believes that those with high baseline CD4 counts have likely managed their HIV well, and the most healthy of this group will continue to do so. However, he thinks that the least healthy of this group are more likely to miss visits and shed earlier. Furthermore, he believes that, among those with low baseline CD4 counts, those who begin shedding just after testing negative at baseline are less likely to drop out (be interval-censored) than those who begin shedding after (at) 12 months. He thinks
that those with low CD4 counts have probably not managed their HIV well in the past, and that the least healthy members of this group have the greatest motivation to make visits. The most healthy members of this group are likely to shed later and may feel less motivated to comply with the visit schedule. When CAR is assumed ($\exp\{\phi\} = 1$) in the CRM, the estimated hazard ratio (SE) of CMV shedding comparing those with high CD4 counts to those with low CD4 counts is 0.58 (0.14). Therefore, we see that patients with high CD4 counts have a significantly lower estimated risk of CMV shedding than those with low CD4 counts. Our goal is to determine how sensitive these conclusions are to departures from CAR.

Analyses were performed with $\exp\{\phi_1^0\}, \exp\{\phi_2^0\} \in \{1, 2, 3\}$ and $\exp\{\phi_1^1\}, \exp\{\phi_2^1\} \in \{\frac{1}{3}, \frac{1}{2}, 1\}$. These values are consistent with Dr. Bozzette’s beliefs that censored individuals with low CD4 counts tend to shed late in their interval, but censored individuals with high CD4 counts tend to shed early in their interval. The hazard ratio estimate (SE) ranges from that under CAR to 0.66 (0.16), when $\exp\{\phi_1^0\} = \exp\{\phi_2^0\} = 3$ and $\exp\{\phi_1^1\} = \exp\{\phi_2^1\} = \frac{1}{3}$. Redistributing those with low (high) CD4 counts to the right (left) diluted the the estimated protective effect of having high baseline CD4 count. An example of this phenomenon in the one-sample setting for those with low CD4 counts is displayed for CAR and extreme departures from CAR in Figure 2. When CAR is assumed, a censored individual’s mass is distributed according to estimates of event-times probabilities. Larger values of censoring bias parameters redistribute censored individuals’ mass further to the right. When the relative probability of dropping out just after baseline (being interval censored) is 50 times more likely for someone who did not shed
within 12 months (began shedding at 12 months) compared to someone who began shedding just after baseline, an individual with low CD4 count and interval \([3, > 12]\) has an estimated 90.6% chance of not shedding within 12 months, compared to 52.4% under CAR.

Estimated survival curves within CD4 count status are shown in Figure 3 under CAR and the most extreme departures from CAR considered. For those with high CD4 counts, we see that small values of censoring bias parameters result in lower estimates of survival. That is, individuals were assumed to shed early in the observed intervals, so estimated survival probabilities are lower than those under CAR. In contrast, for those with low CD4 counts, large values of censoring bias parameters were considered. Individuals in this group were assumed to fail late in their observed intervals, which resulted in higher survival probabilities, relative to CAR.

The hazard ratio estimates and P-values for the hypothesis \(H_0: \beta = 0\) across values of \(\phi^0_1\) and \(\phi^1_1\) (the returners) when \(\exp\{\phi^0_2\} = 2\) (drop-outs with low CD4 count are redistributed to the right) and \(\exp\{\phi^1_2\} = \frac{1}{2}\) (drop-outs with high CD4 count are redistributed to the left) are shown in Figure 4. Over all of the 81 different analyses, the null is rejected in favor of those with high CD4 count when CAR is assumed for drop-outs with high or low CD4 count \((\exp\{\phi^0_2\} = 1\) or \(\exp\{\phi^1_2\} = 1\), or when \(\exp\{\phi^0_2\} = 2\), \(\exp\{\phi^1_2\} = \frac{1}{2}\), and \(\exp\{\phi^0_1\} = 1\) (CAR is assumed for returners with low CD4 count). Overall, the hazard ratio estimates were not sensitive to the range of values chosen for the censoring bias parameters. However, testing conclusions were more sensitive to assumptions about drop-outs than those of returners, because the former outnumbered the latter. Also, conclusions were more sensitive to
assumptions about those with low CD4 counts than those with high CD4 counts, because those with low CD4 counts are the largest group and were censored most often.

### 7.2 ALIVE

ALIVE is an observational study of risk factors for HIV infection among injection drug users (IDUs). Participants were semiannually tested for HIV and interviewed about potential risk factors (Vlahov et al. 1991, Strathdee et al. 2001, and Nelson et al. 2002).

An important scientific question is to determine the 10-year relative risk of seroconversion, comparing those who self-reported needle sharing at baseline to those who did not. Censoring due to missed visits is thought to be informative and may depend on baseline needle sharing status.

The competing risk of death in these data needs to be addressed. An additional complication of this competing risk, which is known for all participants, is that participants’ HIV status may be unknown at death. As before, $M$ denotes the number of follow-up visits. In this case, time $M + 1$ is interpreted as “did not seroconvert while at risk during the study.” Those who drop out and die with unknown HIV status are censored into time $M + 1$ and an interval of missed visits after dropping out until death. When participants are censored by death, the interval $[L, R]$ has a different interpretation, because $R$ now denotes the last missed scheduled visit before death. The possible event times for someone with $L = l$ and $R = r$ are \{l, \ldots, r, M + 1\}.

Also, the equations in Sections 3 to 5 are now conditioned on death status. This conditioning reveals whether or not $M + 1$ is a possible ‘event time’ in
addition to \( \{l, \ldots, r\} \). The relationship between needle-sharing and seroconversion may also be biased due to differential death rates between the groups. However, Figure 5 shows that the mortality process is not significantly different between the two groups. Those who died during the study may have a different relationship between visit and seroconversion processes than those who remained alive at the end of the study, so additional parameters in the censoring bias function are included.

The data consist of 2205 patients who were HIV negative at baseline and whose time to seroconversion was discretized into 10 yearly observations, and > 10 years/no event. At baseline, 1527 participants reported sharing needles, while the remaining 678 did not. Among those who reported sharing needles, 12%, 74%, 9%, and 4% were censored by death, right-censored by drop-out or the end of study, interval censored, and exactly observed, respectively. The percentages among those who did not report needle sharing were 11%, 77%, 8%, and 4%. Among the needle sharers, 242 (16%) died during the study: 52 (21%) died after seroconverting while the remaining 190 (79%) died with unknown HIV status. Among those who did not report needle sharing, 100 (15%) died during the study: 27 (27%) died after seroconverting while the remaining 73 (73%) died with unknown HIV status.

Define \( Z = I(\text{Reported Needle-Sharing}) \), and let \( \Delta \) indicate whether or not \( R \) is the last missed scheduled visit before death. In this case, \( M = 10 \). Since no one is left-censored in this analysis \( (p_0 = 0) \), the equations in Section 5 are adjusted to exclude \( t_0 \). The censoring bias function used in this analysis is
\[
q(\phi, \Delta, t, l, r, z) = \frac{9}{4} \phi_1^z I(r < M + 1) \frac{(t - l)}{(M - 1)} + \phi_2^z I(r = M + 1)(1 - \Delta) \frac{(t - l)}{M} \\
+ \phi_3^z I(r = M + 1)(\Delta) \frac{(t - l)}{M}, \quad z = 0, 1.
\]

where \(\phi = (\phi^1, \phi^0)\) and \(\phi^z = \{\phi_1^z, \phi_2^z, \phi_3^z\}\). Using Bayes’ rule as in Section 4.1,

- \(\exp\{\phi_1^z\}\) is the needle sharing-specific probability ratio of having interval \([1\text{ year, 5 years}]\) comparing those who seroconverted at 5 years to those who seroconverted within one year after baseline.

- \(\exp\{\phi_2^z\}\) is the needle sharing-specific probability ratio of dropping out after baseline comparing those who did not seroconvert within 10 years to those who seroconverted within one year after baseline, among those who remained alive throughout the study.

- \(\exp\{\phi_3^z\}\) is the needle sharing-specific probability ratio of dropping out after baseline comparing those who did not seroconvert while alive to those who seroconverted within one year after baseline, among those who died during the study.

The factor \(\frac{9}{4}\) accounts for the fact that the study lasted 10 years, but the investigators were more comfortable stating beliefs for a 5-year interval as compared to a 10-year interval. When \(\exp\{\phi_1^z\} > 1 (< 1)\), returners are assumed to be more (less) likely to seroconvert late than seroconvert early. When \(\exp\{\phi_2^z\} > 1 (< 1)\), drop-outs who remain alive are assumed to be
more (less) likely to seroconvert late or not at all than seroconvert early. 

When \( \exp\{\phi_2^b\} > 1 \) (< 1), drop-outs who later die are assumed to be more (less) likely to seroconvert late or not at all than seroconvert early.

To elicit the values of the censoring bias parameters used in our analysis, Drs. Galai and Vlahov were separately shown the schematic in Figure 6a, and they were asked, “Among those who self-reported needle sharing at baseline, who is more likely to test negative for HIV at baseline, miss visits, then return at 5 years infected with HIV: someone who seroconverted just after baseline or someone who seroconverted at 5 years? How many times more likely?”

Next, using the schematic in Figure 6b, they were asked, “Among those who self-report needle sharing at baseline and who remained alive throughout the study, who is more likely to test negative for HIV at baseline, then drop out: someone who seroconverted just after baseline or someone who did not seroconvert within 10 years? How many times more likely?” Lastly, referring the schematic in Figure 6c, they were asked, “Among those who self-reported needle sharing at baseline, who is more likely to test negative for HIV at baseline, then drop out and die with unknown HIV status: someone who seroconverted just after baseline or someone who did not seroconvert while at risk? How many times more likely?” The questions were repeated for those who did not report needle sharing at baseline.

The elicited censoring bias parameters varied between the two experts, so they reached a consensus about the range of plausible values. The experts believe that, among needle sharers, those who seroconvert at 5 years are 1.75 times less to 2.75 times more likely to be censored into the interval [1 year, 5 years] than those who seroconvert within one year after baseline.
The range for non-sharers is 1.15 times less to 2.50 times more likely. The experts are unsure about the direction of this relationship because those who seroconvert earlier may either behave irresponsibly and miss visits, but return when their health diminishes, or may acknowledge their high-risk status and feel motivated to participate in the study, compared to those who seroconvert later. For those who remained alive at the end of the study, the experts believe that among needle sharers, those who did not seroconvert within 10 years are 1.50 to 3.00 times more likely to drop out after baseline than those who seroconvert within one year from baseline. Among non-needle sharers, the range was 1.75 to 2.50 times more likely. Among those who die during the study (within 10 years from baseline), the experts believe that those who do not seroconvert while at risk are 2.00 to 2.50 more likely to drop out than those who seroconvert within one year from baseline with the same baseline needle-sharing status. They believe that those who seroconvert early would eventually return to the study as their condition worsens, while those who do not seroconvert while at risk during the study would not be motivated to return. Those who die without having seroconverted are likely to die from other reasons, such as drug overdose and homicide.

When CAR is assumed to hold, the estimated relative hazard (SE) of seroconversion from the CRM comparing those with baseline self-reported needle-sharing to those without baseline self-reported needle-sharing is 1.05 (0.14). Therefore, we see that needle-sharers are estimated to be more likely to seroconvert than non-sharers, but not significantly so. The hazard ratio estimate (SE) ranges from 0.94 (0.12) when \( \exp\{\phi^0\} \) are the minimum elicited values (non-sharers distributed toward early seroconversion times) and \( \exp\{\phi^1\} \) are
the maximum (needle-sharers distributed toward late seroconversion times) to 1.24 (0.17) when \( \exp\{\phi^0\} \) are the maximum elicited values (non-sharers distributed toward late seroconversion times) and \( \exp\{\phi^1\} \) are the minimum (needle-sharers distributed toward early seroconversion times). While the estimated hazard ratio changes direction over the range of values, it is never significantly different from 1. Redistributing (non) needle sharers to the (right) left increases the estimated hazard ratio, while redistributing (non) needle sharers to the (left) right decreases the estimated hazard ratio. An example of this phenomenon in the one-sample setting is displayed in Figure 7 for needle-sharers who remain alive at the end of the study. When CAR is assumed, a censored individual’s mass is distributed according to estimates of seroconversion probabilities. Larger values of censoring bias parameters redistribute censored individuals’ mass further to the right, while smaller values distribute mass further to the left. When the relative probability of dropping out just after baseline (being interval censored) is 5 times less likely for someone who did not seroconvert within 10 years (seroconverted at 10 years) compared to someone who seroconverted just after baseline, a needle-sharer with interval \([1, > 10]\) has an estimated 38.5% chance of seroconverting within one year of baseline and 19.4% chance of not seroconverting within 10 years. However, when the relative probability of dropping out just after baseline (being interval censored) is 5 times more likely for someone who did not seroconvert within 10 years (seroconverted at 10 years) compared to someone who seroconverted just after baseline, a needle-sharer with interval \([1, > 10]\) has an estimated 0.6% chance of seroconverting within one year of baseline and 92.9% chance of not seroconverting within 10 years. The CAR
assumption produces intermediate values. Estimated survival curves within needle-sharing status are shown in Figure 8 for CAR and the most extreme values of the censoring bias parameters elicited. In general, the results were not sensitive to the elicited range of values of the censoring bias parameters. We see that despite the change in direction of extreme values for returners, the estimated survival probabilities are higher than those under CAR. This result is due to the large values of censoring bias parameters for drop-outs, who are the bulk of the study sample. Regardless of needle-sharing status, the range of elicited values for drop-outs (relative probabilities of 2 to 2.5) means that drop-outs are believed to be more likely to not seroconvert within 10 years, than fail at any one time in their interval. However, when survival is estimated at the maximum elicited values for needle-sharers and minimum elicited values for non-sharers, the survival curves for the groups are reversed. Under CAR, non-sharers have better estimated survival than needle-sharers, but under these extreme assumptions, needle-sharers have better estimated survival than non-sharers. Similarly, direction of the estimated hazard ratios from the CRM were more sensitive to drop-outs than to returners. Also, results were more sensitive to assumptions about needle sharers than non-sharers, as the former group is larger and was censored more often than the latter.

8 Discussion

By parameterizing the departure from CAR, our method is more flexible than the sensitivity analysis approach of calculating bounds by imputing event
times to be the interval endpoints. An additional benefit of our method is the use of scientific information for sensitivity analysis. The elicitation approach raises scientists’ awareness of the assumptions associated with standard statistical procedures and facilitates discussion between the scientist and statistician. The results of the ALIVE and ACTG 181 data analyses exemplify the sensitivity of the results to assumptions about the censoring process.

One difficulty of this approach is finding a low-dimensional, scientifically interpretable function to capture several characteristics that are potentially related to the censoring process. Different versions of the censoring bias function were mentioned in this paper, from a two-parameter function that accounted for a binary covariate and interval length (as in the simulation study), to a six-parameter function that accounted for a binary covariate, drop-out, interval length, and death (as in ALIVE).

Our method extends discrete-time survival analysis using the CRM by loosening the CAR assumption. The complete-data likelihood we used in our analyses is a special case of a broad class of generalized linear models using a multinomial distribution. This class also includes the discrete-time Cox model, which can be extended to handle informatively interval-censored data by generalizing the method of Tu, Meng, and Pagano (1993) via exponential tilt models. This approach is equivalent to performing our method where the cumulative complementary log-log link replaces the continuation ratio logit link. Our method can also be used to analyze interval-censored counts of a binary outcome by using the cumulative logit link. Complications of this case will be explored in future research. Other future research of in-

28
formatively interval-censored data includes alternative estimation procedures for the models presented in this paper, as the EM algorithm converges slowly and not necessarily to a unique value. We will also explore a Bayesian version of these models, non-parametric methods for comparing survival curves, continuous-time methods, and the inclusion of high-dimensional covariates. Lastly, we will pursue computer-based interactive methods for eliciting expert opinion.

Appendix

**Theorem 1** Suppose Model (3,4) holds for a specified function $q$. If $P(L = t, R = t) > 0$ for all $t = 0, \ldots, M + 1$, then there exists a unique solution to (6), which is a proper probability mass function for $T$. Furthermore, this solution in conjunction with the selection model (4) yields a joint distribution of $(L, R, T)$, which marginalizes to the population distribution of $(L, R)$.

To shorten notation, let $p_t = P(T = t) \geq 0$, $h_{l,r} = P(L = l, R = r)$, and $p = (p_0, \ldots, p_M, p_{M+1})$.

**Proof:** Based on the selection model representation (4), the log of the observed data distribution can be written as

$$
\ell(p; q) = \sum_{[l,r] \in E^*} h_{l,r} \log \left[ \sum_{t=l}^{r} P(L = l, R = r \mid T = t) P(T = t) \right]
= \sum_{[l,r] \in E^*} h_{l,r} \log \left[ f(l, r, p; q) \sum_{t \in [l,r]} \exp\{q(t, l, r)\} p_t \right],
$$

where $f(l, r, p) = P(L = l, R = r \mid T \in [l, r]) \exp\{d(l, r; q)\}$. 

29
Ignoring \( f(l, r, p) \) in the above equation, consider maximization of the resultant non-likelihood based objective function subject to the constraints \( \sum_{t=0}^{M+1} p_t = 1 \) and \( p_t \geq 0 \), for \( t = 0, \ldots, M + 1 \). Given our assumption that \( h_{t,t} > 0 \) for all \( t \), it follows the objective function is \(-\infty\) when any of the \( p_t \)'s equal 0. Thus, we restrict attention to the interior region where all \( p_t > 0 \) and maximize the objective function with respect to the constraint that the sum of the \( p_t \)'s equal 1. To do this, we maximize (over \( p > 0 \)) the Lagrangian function,

\[
\ell^*(p, \lambda; q) = \sum_{[l,r] \in E^*} h_{l,r} \log \left[ \sum_{t \in [l,r]} \exp\{q(t, l, r)\} p_t \right] - \lambda \left\{ \sum_{t=0}^{M+1} p_t - 1 \right\}.
\]

The score equations for \( \ell^*(p, \lambda; q) \) with respect to \( p \) are

\[
\frac{\partial \ell^*(p, \lambda; q)}{\partial p_j} = \sum_{[l,r] \in E^*} h_{l,r} \frac{I(j \in [l, r]) \exp\{q(j, l, r)\}}{\sum_{t \in [l,r]} \exp\{q(t, l, r)\} p_t} - \lambda = 0 \quad j = 0, \ldots, M + 1 \tag{8}
\]

\[
\frac{\partial \ell^*(p, \lambda)}{\partial \lambda} = \sum_{t=0}^{M+1} p_t - 1 = 0. \tag{9}
\]

We can solve for \( \lambda \) by summing (8) over \( j \). This result tells us that \( \lambda = \{\sum_{t=0}^{M+1} p_t\}^{-1} \). By (9), we see that \( \lambda = 1 \). By plugging \( \lambda = 1 \) into (8) and multiplying by \( p_j \), we obtain our identifiability Equations (6).

Now, it can be shown that \( \ell^*(p, 1; q) \) is strictly concave, which will imply that there is a unique solution, \( p^* \), to the identifiability equations. To see this, we will show that the Hessian of \( \ell^*(p, 1; q) \) is negative definite for all \( p > 0 \). In order to show that the Hessian, \( H(p; q) \), is negative definite, we must show that

\[
x^t H(p; q) x < 0 \quad \text{for all} \quad x \neq 0,
\]

30
where \( x = (x_0, \ldots, x_{M+1})' \) is a real vector of length \( M + 2 \). The components of \( H(p; q) \) are:

\[
\frac{\delta^2 \ell^*(p, 1)}{\delta p_j^2} = -\sum_{r=j}^{M+1} \sum_{l=0}^{j} h_{l,r} \left( \frac{\exp\{q(j, l, r)\}}{\sum_{i=l}^{r} \exp\{q(i, l, r)\} p_i} \right)^2,
\]

\[
\frac{\delta^2 \ell^*(p, 1)}{\delta p_j \delta p_{j'}} = -\sum_{r=\max(j, j')}^{M+1} \sum_{l=0}^{\min(j, j')} h_{l,r} \frac{\exp\{q(j, l, r)\} \exp\{q(j', l, r)\}}{(\sum_{i=0}^{r} \exp\{q(i, l, r)\} p_i)^2},
\]

for all \( j, j' \in \{0, \ldots, M + 1\}, j \neq j' \). Now, \( x' H(p; q)x \) becomes

\[
-\sum_{t=0}^{M+1} h_{t,t} \left( \frac{x_t}{p_t} \right)^2 - \sum_{r=1}^{M+1} \sum_{l=0}^{r-1} h_{l,r} \left( \frac{\sum_{i=l}^{r} \exp\{q(i, l, r)\} x_i}{\sum_{k=l}^{r} \exp\{q(k, l, r)\} p_k} \right)^2,
\]

which is negative since \( h_{t,t} > 0 \) for all \( t \).

Using (4) and the solution \( p^* \), the joint distribution of \((T, L, R)\) is given by

\[
P(T = t, L = l, R = r) = \frac{P(L = l, R = r | T = t)p_t^*}{\sum_{s \in [l, r]} \exp\{q(s, l, r)\} p_s^* p_t^*}. \]

We obtain that \( P(L = l, R = r) = h_{l,r} \) by summing over \( t \in [l, r] \).

**Acknowledgements**

The research of Michelle Shardell was supported by National Institute of Aging grant T32 AG00247. The research of Daniel Scharfstein was partially supported by National Institute of Health grants 1-R29-GM48704-04,
5R01A132475, R01CA74112, 1-R01-MH56639-01A1, and 1- R01-DA10184-01A2. The research of Noya Galai and David Vlahov was supported by National Institute on Drug Abuse grant DA 04334. The authors thank Tom Louis and Zhiqiang Tan for helpful discussions. They also thank Rebecca Betensky and Dianne Finkelstein for inspiring this work as well as providing access to data from ACTG 181.

References


methods and characteristics of participants, *NIDA Research Monographs*, 109, 75-100.
Table 1: Simulation results. 1000 iterations, $\beta = 0$, $M = 4$, $n_0 = n_1 = 100$, 86% censoring in both groups. $\alpha = 0.05$. Bold indicates correct values for $\phi_0^{\text{modeled}}$ and $\phi_1^{\text{modeled}}$.

<table>
<thead>
<tr>
<th>$\phi_0^{\text{true}}$</th>
<th>$\phi_1^{\text{true}}$</th>
<th>$\phi_0^{\text{modeled}}$</th>
<th>$\phi_1^{\text{modeled}}$</th>
<th>Mean $\beta - \beta$</th>
<th>Mean $\tilde{SE}(\hat{\beta})$</th>
<th>Empirical $\tilde{SD}(\hat{\beta})$</th>
<th>MSE$^2$</th>
<th>Empirical Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>-0.01 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>-0.16 0.22 0.21 0.26</td>
<td>0.21</td>
<td>0.26</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0</td>
<td>log(2)</td>
<td>-0.36 0.22 0.22 0.42</td>
<td>0.22</td>
<td>0.24</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.02 0.22 0.23 0.24</td>
<td>0.23</td>
<td>0.24</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>-0.19 0.22 0.23 0.30</td>
<td>0.24</td>
<td>0.24</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0.16 0.22 0.22 0.27</td>
<td>0.22</td>
<td>0.27</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>-0.001 0.22 0.21 0.28</td>
<td>0.21</td>
<td>0.28</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0</td>
<td>log(2)</td>
<td>-0.18 0.22 0.21 0.28</td>
<td>0.21</td>
<td>0.28</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.16 0.22 0.23 0.28</td>
<td>0.23</td>
<td>0.28</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>-0.02 0.22 0.23 0.29</td>
<td>0.23</td>
<td>0.29</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0.16 0.22 0.22 0.37</td>
<td>0.22</td>
<td>0.27</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>-0.04 0.22 0.23 0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0</td>
<td>log(2)</td>
<td>0.30 0.22 0.22 0.37</td>
<td>0.22</td>
<td>0.27</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.13 0.22 0.23 0.27</td>
<td>0.23</td>
<td>0.27</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>-0.02 0.22 0.23 0.29</td>
<td>0.23</td>
<td>0.29</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0.002 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>-0.13 0.22 0.22 0.26</td>
<td>0.22</td>
<td>0.26</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0</td>
<td>log(2)</td>
<td>-0.14 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>log(2)</td>
<td>log(2)</td>
<td>-0.31 0.22 0.22 0.38</td>
<td>0.22</td>
<td>0.38</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.005 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0.13 0.22 0.22 0.26</td>
<td>0.22</td>
<td>0.26</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>-0.19 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>log(2)</td>
<td>0</td>
<td>log(2)</td>
<td>0.01 0.22 0.22 0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.14 0.22 0.22 0.26</td>
<td>0.22</td>
<td>0.26</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>-0.04 0.22 0.23 0.24</td>
<td>0.23</td>
<td>0.24</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>log(2)</td>
<td>0.12 0.22 0.22 0.25</td>
<td>0.22</td>
<td>0.25</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Simulation results. 1000 iterations, $\beta = 0.75$, $M = 4$, $n_0 = n_1 = 100$, 97% censoring when $Z = 1$ and 86% censoring when $Z = 0$. $\alpha = 0.05$.

Bold indicates correct values for $\phi_0^{\text{modeled}}$ and $\phi_1^{\text{modeled}}$.

<table>
<thead>
<tr>
<th>$\phi_0^{\text{true}}$</th>
<th>$\phi_1^{\text{true}}$</th>
<th>$\phi_0^{\text{modeled}}$</th>
<th>$\phi_1^{\text{modeled}}$</th>
<th>Mean $\hat{\beta} - \beta$</th>
<th>Mean $\widehat{\text{SE}}(\hat{\beta})$</th>
<th>Empirical $\text{SD}(\hat{\beta})$</th>
<th>MSE$^{\frac{1}{2}}$</th>
<th>Empirical Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.03$</td>
<td>$0.26$</td>
<td>$0.26$</td>
<td>$0.27$</td>
<td>$0.27$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.28$</td>
<td>$0.26$</td>
<td>$0.27$</td>
<td>$0.38$</td>
<td>$0.27$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.72$</td>
<td>$0.26$</td>
<td>$0.27$</td>
<td>$0.77$</td>
<td>$0.22$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.09$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.28$</td>
<td>$0.48$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.53$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.60$</td>
<td>$0.48$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.34$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.43$</td>
<td>$0.38$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.33$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.44$</td>
<td>$0.79$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.04$</td>
<td>$0.27$</td>
<td>$0.26$</td>
<td>$0.27$</td>
<td>$0.95$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.39$</td>
<td>$0.26$</td>
<td>$0.26$</td>
<td>$0.47$</td>
<td>$0.67$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.20$</td>
<td>$0.27$</td>
<td>$0.29$</td>
<td>$0.35$</td>
<td>$0.86$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.21$</td>
<td>$0.26$</td>
<td>$0.27$</td>
<td>$0.34$</td>
<td>$0.86$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.01$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.95$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.99$</td>
<td>$0.35$</td>
<td>$0.34$</td>
<td>$0.32$</td>
<td>$0.47$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.59$</td>
<td>$0.34$</td>
<td>$0.38$</td>
<td>$0.38$</td>
<td>$0.71$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-0.01$</td>
<td>$0.31$</td>
<td>$0.32$</td>
<td>$0.32$</td>
<td>$0.95$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.77$</td>
<td>$0.34$</td>
<td>$0.41$</td>
<td>$0.87$</td>
<td>$0.41$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$0.17$</td>
<td>$0.31$</td>
<td>$0.34$</td>
<td>$0.38$</td>
<td>$0.93$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.35$</td>
<td>$0.31$</td>
<td>$0.32$</td>
<td>$0.42$</td>
<td>$0.74$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.16$</td>
<td>$0.27$</td>
<td>$0.28$</td>
<td>$0.33$</td>
<td>$0.92$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.13$</td>
<td>$0.27$</td>
<td>$0.26$</td>
<td>$0.30$</td>
<td>$0.91$</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-0.53$</td>
<td>$0.26$</td>
<td>$0.25$</td>
<td>$0.58$</td>
<td>$0.47$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.03$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.27$</td>
<td>$0.96$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-0.37$</td>
<td>$0.26$</td>
<td>$0.26$</td>
<td>$0.46$</td>
<td>$0.69$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-0.21$</td>
<td>$0.26$</td>
<td>$0.28$</td>
<td>$0.35$</td>
<td>$0.87$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.73$</td>
<td>$0.34$</td>
<td>$0.42$</td>
<td>$0.85$</td>
<td>$0.44$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.42$</td>
<td>$0.34$</td>
<td>$0.38$</td>
<td>$0.56$</td>
<td>$0.80$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$0$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-0.18$</td>
<td>$0.30$</td>
<td>$0.31$</td>
<td>$0.36$</td>
<td>$0.89$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.61$</td>
<td>$0.34$</td>
<td>$0.43$</td>
<td>$0.74$</td>
<td>$0.62$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-0.01$</td>
<td>$0.31$</td>
<td>$0.32$</td>
<td>$0.32$</td>
<td>$0.96$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.19$</td>
<td>$0.32$</td>
<td>$0.32$</td>
<td>$0.32$</td>
<td>$0.95$</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>$\log(2)$</td>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>$0.59$</td>
<td>$0.34$</td>
<td>$0.43$</td>
<td>$0.73$</td>
<td>$0.64$</td>
</tr>
</tbody>
</table>
Figure 1: ACTG 181. Schematic used to elicit expert information.
Figure 2: ACTG 181. Sensitivity of CMV shedding-time probabilities to departures from CAR, \( P(T = t \mid [L, R] = [3, >12], Z = 0) \).
Figure 3: ACTG 181. Sensitivity analysis for CD4-specific survival curves for CAR and extreme values of censoring bias parameters.
Figure 4: ACTG 181: Sensitivity analysis for hazard ratios CD4 ≥ 200 v. CD4 < 200 when $\exp\{\phi_0^0\} = 2$, and $\exp\{\phi_1^1\} = \frac{1}{2}$. Hazard Ratio = 0.623 and P-value = 0.047 when $\exp\{\phi_0^0\} = 1$ and $\exp\{\phi_1^1\} = 1$. 
Figure 5: ALIVE. Mortality distribution by needle-sharing status.
Figure 6: ALIVE. Schematic used to elicit expert information.

a.

\[\text{negative tests} \quad \text{baseline} \quad \text{missed visits} \quad \text{positive tests} \quad 5 \text{ years} \]

b.

\[\text{negative tests} \quad \text{baseline} \quad 5 \text{ years} \quad 10 \text{ years} \quad \text{end of study} \quad \text{seroconversion after 10 years or never} \quad \text{death} \quad \text{drop out} \]

c.

\[\text{negative tests} \quad \text{baseline} \quad 10 \text{ years} \quad \text{end of study} \quad \text{death} \quad \text{seroconversion?} \quad \text{drop out} \]

43
Figure 7: ALIVE. Sensitivity of seroconversion-time probabilities to departures from CAR, $P(T \leq t \mid [L, R] = [1, >10], Z = 1, \Delta = 0)$

\[ \exp(\phi_1) = \exp(\phi_2) = \exp(\phi_3) = 1/5 \]

Seroconversion Time t (years from baseline)

Seroconversion Probability

CAR

Seroconversion Time t (years from baseline)

Seroconversion Probability

exp(\phi_1) = exp(\phi_2) = exp(\phi_3) = 5

Seroconversion Time t (years from baseline)

Seroconversion Probability

44
Figure 8: ALIVE. Sensitivity Analysis for needle sharing-specific Survival Curves for CAR and extreme elicited values of censoring bias parameters.