Survival Curve Estimation for Informatively Coarsened Discrete Event-Time Data

Michelle Shardell\textsuperscript{1}\textsuperscript{*}, Daniel O. Scharfstein\textsuperscript{2}, and Samuel A. Bozzette\textsuperscript{3}

\textsuperscript{1}Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, Maryland, USA.
\textsuperscript{2}Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA.
\textsuperscript{3}Department of Medicine, San Diego Veterans Affairs Medical Center, San Diego, California, USA.

SUMMARY

Interval-censored, or more generally, coarsened event-time data arise when study participants are observed at irregular time periods and experience the event of interest in between study observations. Such data are often analyzed assuming non-informative censoring, which can produce biased results if the assumption is wrong. This paper extends the standard approach for estimating survivor functions to allow informatively interval-censored data by incorporating various assumptions about the censoring mechanism into the model. We include a Bayesian extension in which final estimates are produced by mixing over a distribution of assumed censoring mechanisms. We illustrate these methods with a natural history study of HIV-infected individuals using assumptions elicited from an AIDS expert.

Copyright © 2000 John Wiley & Sons, Ltd.

\textsuperscript{*}Correspondence to: Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, Maryland 21201-1596, USA. Email: mshardel@epi.umaryland.edu, Phone: 410-706-8563, Fax: 410-706-3808.

Contract/grant sponsor: National Institute of Aging; contract/grant number: T32-AG00247

Contract/grant sponsor: National Institute of Health; contract/grant number: 1-R29-GM48704-04, 5R01A132475, R01CA74112, 1-R01-MH56639-01A1, and 1- R01-DA10184-01A2
Keywords: Interval censoring; Informative censoring; Sensitivity analysis; Coarsened at random; Survival.

1. INTRODUCTION

In many prospective studies, participants are evaluated for events of interest at baseline and several pre-scheduled times after enrollment. Estimating event incidence per unit of time (e.g., monthly) may be of interest to investigators. To accommodate this interest, time may be partitioned into discrete periods motivated by the visit schedule. Since participants often miss visits or attend visits off schedule, the event time is censored into adjacent time periods defined by the partition. For those who attend every visit on schedule, the time period of the event is known (e.g., the month of the event), resulting in discrete event-time data. Often, the event times are only known to have occurred within a subset of adjacent time periods. In addition, the event can be detected at the baseline visit in some patients, and never detected in others. Thus, the observable data are a combination of exact, left, right, and interval censoring, which can all be mathematically formulated under the rubric of interval censoring [1] and even more generally as coarsened data [2–4].

Such data are usually analyzed by assuming non-informative censoring (see, for example, refs [5–7]; and references therein). That is, knowing the outcome was censored due to missed or off-schedule study observations provides no more information about the event process than knowing the interval in which the event occurred. This assumption is a special case of coarsening at random (CAR) described previously [2–4, 8].

There has been some work utilizing auxiliary data and additional assumptions that allows interval-censored data to be coarsened not at random (CNAR) relative to data structures like
that of Turnbull [6] that do not rely on additional information. Others [9] proposed testing CAR using auxiliary data, and Finkelstein, Goggins, and Schoenfeld (hereafter, FGS) [10] developed an estimation procedure for a specific dependent censoring model in which the censoring mechanism is identified by additional assumptions. In particular, they propose a selection model in which the probability of attending a scheduled visit depends on the visit number and whether or not the visit was before or after the event. Visits are assumed to be conditionally independent given the event time (possibly unobserved). Others utilized death information to jointly model the risk of death and vertical transmission of HIV from mother to child. However, they assumed that the HIV-testing process was not informative about infection time [11].

Even when CAR is assumed, the analysis of interval-censored data requires special methods. Turnbull [6] used the EM algorithm [12] to extend the Kaplan-Meier [13] survivor curve estimator to allow general censoring. Others utilized optimization theory for concave programming with linear constraints and analyzed the data by solving the Kuhn-Tucker conditions [7].

While assuming CAR is computationally convenient, the assumption is usually untestable and is often considered implausible by scientific experts. Therefore, a CAR-based analysis of these data could produce biased results. Thus, studies where an absorbing time-to-event outcome is coarsened due to visit non-compliance 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 be flexible enough to incorporate different assumptions about the censoring mechanism in the model. Such an approach would allow investigators to evaluate the sensitivity of conclusions to varying assumed departures from CAR.
Turnbull’s method [6] is commonly encountered in the medical literature in various contexts [14–16]. Therefore, this paper focuses on extending the method. Although Turnbull’s method [6] was extended by FGS [10] to allow CNAR, they did so by making the unverifiable assumption that visits are independent given event time. The authors assume that the probability of missing a visit depends on occurrence of the event, but not on the patient’s knowledge of event occurrence.

We believe that when modeling assumptions are not identifiable, analyses should not be driven by a single set of assumptions. Instead a sensitivity analysis should be performed over a variety of scientifically interpretable assumptions and reported to allow readers to interpret the results in the context of the scientific application. This viewpoint is consistent with those of other authors [17–24]. In this paper, we extend Turnbull’s method [6] by relaxing the CAR assumption through the use of non-identifiable, scientifically interpretable parameters that describe the degree of departure from CAR. This approach facilitates a sensitivity analysis of the results to unverifiable assumptions. In contrast to FGS [10], our method allows the data to be CNAR above and beyond additional assumptions and auxiliary information.

A benefit of the FGS [10] procedure is that they report a single answer. To retain this quality, we also propose a fully Bayesian approach in which we elicit a prior distribution for the magnitude of the departure from CAR from a scientific expert. Priors differ across experts, and our method facilitates scientists interpreting results using their own prior. In addition, the sensitivity of results across priors can be assessed. Using a particular prior, we can summarize our results by displaying the posterior of of the event time distribution. A Gibbs sampler [25] has previously been proposed utilizing data augmentation [26] and Dirichlet priors to analyze non-informatively interval-censored data [27]. In this paper, we extend a discrete-time version
of this algorithm to allow informative coarsening.

As an illustrating example, we will analyze data used by FGS [10] from AIDS Clinical Trials Group (ACTG) Study 181, a natural history study of cytomegalovirus (CMV) infection among HIV-infected patients. In this study the onset of CMV shedding in the blood was determined by laboratory tests scheduled every 12 weeks. However, some patients were evaluated every four weeks (according to the urine CMV test schedule) while others were evaluated irregularly due to missed and off-schedule visits. One goal of this study is to estimate the distribution of time (in months) to CMV shedding in the blood. However, participants often missed visits, thus the shedding times are only known within an interval of adjacent months. As part of the analysis, we will illustrate how to use the interpretable parameters to elicit information from a scientific expert.

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 models that relax the CAR assumption. Section 5 discusses the likelihood and proposed estimation procedures using the EM algorithm and Gibbs sampler. The performance of our frequentist approach is evaluated in Section 6 via simulation studies. Section 7 presents analyses of the ACTG 181 study, including a description of our method for eliciting expert opinion. Finally, Section 8 is a discussion of the proposed extensions of our methodology.

2. DATA STRUCTURE AND NOTATION

Suppose that time is partitioned into $M + 2$ intervals based on some time unit of interest, 
\[ \{(-\infty, u_0], (u_0, u_1], \ldots, (u_{M-1}, u_M], (u_M, \infty)\}, \]
where $u_j < \infty$, for $j = 1, \ldots, M$ and $u_0 = 0$.

Let $T$ denote the time period of the event induced by the partition. So, $E = \{t : t =
\(0, \ldots, M + 1\) is the support of the distribution of \(T\), where \(T = t\) denotes that the event occurred in the \(t\)th time period. Due to possible skipped or off-schedule visits, the observed data for an individual is just a set of adjacent time periods from \(E\). 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 the event happened at or prior to entry, \((-\infty, 0]\), then \(L = R = 0\). If the time of the event is known to occur in a particular period, \((u_{j-1}, u_j], j = 1, \ldots, M\), then \(L = R = j\). If the event did not occur during the planned follow-up period, interval \((u_M, \infty]\), then \(L = R = M + 1\). Lastly, if knowledge about \(T\) is incomplete, then \(L < R\). Those with \(L < R = M + 1\) are drop-outs, and those with \(L < R < M + 1\) are returners. Where necessary, the subscript \(i\) will denote person-specific data. \(T\) will be referred to as the “event time,” and its distribution will be known as the “event process.” Similarly, the distribution of the random variables \(L\) and \(R\) will represent the “coarsening process.”

3. COARSENING AT RANDOM

In this section, we provide a formal definition of CAR in the context of coarsened event-time data. This definition is adapted from the more general form presented in other work [8].

Formally, CAR means that

\[
P(L = l, R = r \mid T = t) \text{ is constant in } t \in [l, r],
\]

for all \([l, r] \in E^* = \{[l, r] : l \leq r, l, r \in E\}\). Equation (1) is equivalent to

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

\(\text{Copyright } \odot 2000 \text{ John Wiley & Sons, Ltd. Statist. Med. 2000; } 00:0-0\)

\(\text{Prepared using simauth.cls}\)
for all \( t \in [l, r] \) whatever \( [l, r] \in E^* \) [8].

The left-hand side of Equation (2) is a function of both the event and coarsening processes, while the right-hand side only depends on the event process (time to CMV shedding). 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.

Heitjan and Rubin [2] introduced the broader notion of coarsening that includes the special cases of censored data and missing data. That missing data are a special case of coarsened data means that missing data mechanisms defined by Rubin [28] are also coarsening processes. The missing data vocabulary can therefore be used to better understand the more general coarsening processes that are being applied to censored data in this paper. Data that are missing completely at random (MCAR) are CAR, and data that are missing at random (MAR) are CAR within levels of completely observed variables. Our method and that of FGS deal with data that are CNAR. In the case of missing data, data that are missing not at random (MNAR) are CNAR because the probability of missing depends on incompletely observed variables. In the context of censored survival data, “missing” is replaced by “censored into an interval.”

Given the \( L \) and \( R \), where \( L \leq T \leq R \), one can always construct a unique event-time random variable \( T \), such that \( T \in [L, R] \) and CAR holds [8]. This result shows that, under CAR, the distribution of \( T \) is uniquely identified for some partition of time and that CAR itself does not place any restrictions on the distribution of the observed data. Methods have been proposed for finding the unique solution [6, 7].

Copyright © 2000 John Wiley & Sons, Ltd. Statist. Med. 2000; 00:0–0

Prepared using simauth.cls
4. CNAR MODELS, IDENTIFIABILITY, SENSITIVITY ANALYSIS

Since CAR is untestable without additional assumptions and often scientifically implausible, it is useful to consider alternative coarsening models. We proceed by positing a class of CNAR 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 uniquely 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 by a small set of interpretable censoring bias parameters.

4.1. CNAR models

To loosen the CAR assumption, we allow the model to be indexed by $q(t, l, r)$, a finite specified function of $t$, $l$, and $r$, which describes the relationship between visit compliance and event time. For each $q$, the posited model is an exponential tilting [29] of the CAR model. In the spirit of models proposed by other authors [21–23, 30]; we posit the following class of pattern-mixture models for the marginal distribution of $T$:

$$P(T = t \mid L = l, R = r) = \frac{P(T = t \mid T \in [l, r]) \exp\{q(t, l, r)\}}{c(l, r; q)},$$

(3)

where $c(l, r; q) = \sum_{s \in [l, r]} P(T = s \mid T \in [l, r]) \exp\{q(s, l, r)\}$. If $q(t, l, r)$ does not depend on $t$, then no tilting is performed, and CAR is satisfied. Without loss of generality, we assume that $q(t, t, t) = 0$, for $t \in E$. Copyright © 2000 John Wiley & Sons, Ltd. Statist. Med. 2000; 00:0–0

Prepared using simauth.cls
Using Bayes’ rule, it is possible to represent (3) as a selection model of the following form:

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

(4)

where \(d(l, r; q) = -\log \{c(l, r; q)\}\). The selection model (4) implies that

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

(5)

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

\(q(t_{ref}, l, r) = 0\). A useful interpretation of \(q(r, l, r)\) can also be seen from the pattern-mixture model perspective. Using Equation (3), we have

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

(6)

Equation (6) reveals that \(q(t, l, r)\) is the difference of log probability ratios of having the event at time \(t\) compared to \(t_{ref}\), conditioned on \(L = l, R = r\) versus conditioning on \(T \in [l, r]\) (i.e., versus CAR). In subsequent sections, we discuss interpretable parameterizations of \(q(t, l, r)\).

### 4.1.1. Interpretable parameterization of the censoring bias function

For ease of interpretation, it is best to assume a low-dimensional parameterization for \(q(t, l, r)\). For example, let \(\phi\) be a scalar censoring bias parameter and let

\[
q(t, l, r; \phi) = \phi \frac{(t - l)}{(r - l)}
\]

(7)
In this case, $t_{ref} = l$ and $\exp\{\phi\}$ is interpreted as the probability ratio of having censoring interval $[l, r]$ comparing those whose true event time is $r$ to those whose true event time is $l$. In this particular parameterization, the length of the interval does not affect the probability ratio. When $\exp\{\phi\} > 1$ ($< 1$), participants whose true event time is later (earlier) in an interval are more likely to have the interval than participants whose true event time is earlier (later) in the interval. In the context of ACTG 181, $\exp\{\phi\} > 1$ means that healthier patients tend to miss visits, while $\exp\{\phi\} < 1$ means that sicker patients tend to miss visits. Using the pattern-mixture perspective, $\exp\{\phi\} > 1$ ($\exp\{\phi\} < 1$) means that the relative probability of having event time $r$ compared to $l$ is higher (lower) for patients censored into the interval $[l, r]$ than what would be assumed under CAR.

Heuristically, $\phi > 0$ ($\phi < 0$) moves mass toward later (earlier) times relative to CAR. To illustrate, consider an example where $P(T = t) = P(T = t')$ for all $t, t' \in E$ so that when data are CAR ($\phi = 0$), $P(T = t \mid L = l, R = r) = (r - l + 1)^{-1}$ for $t \in [l, r]$. Using the censoring bias function in Equation (7), Figure 1 shows the impact of $\phi$ on $P(T = t \mid L = 1, R = 4)$. For this example, $P(T = t \mid L = 1, R = 4) = 0.25$ for all $t \in [1, 4]$ when $\phi = 0$. Large positive (negative) values of $\phi$ result in most mass on $t = r$ ($t = l$). A sensitivity analysis that consists of estimating survival assuming $T = r$ ($T = l$) is a special case of the proposed method with $\phi = \infty$ ($\phi = -\infty$). In contrast, a censoring bias function like that in Equation (7) is flexible enough to allow intermediate departures from CAR by redistributing mass across the censoring interval without putting all mass on one time.'
4.2. Identifiability

To establish sufficient conditions for the distribution of $T$ to be uniquely identified from the distribution of $L$ and $R$ under model (3, 4), we note that for $t \in E$,

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

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

(8)

Under the sufficient conditions given in Appendix 1, the system of equations in (8) can be uniquely solved for the distribution of $T$. If the conditions are not satisfied for a given partition of time, a new partition can be established that fuses adjacent time periods such that the sufficient conditions are satisfied. These conditions are developed by conceptualizing the problem as nonlinear optimization with linear constraints and following the arguments utilized in [7].

4.3. Sensitivity analysis of ACTG 181

The result in Appendix 1 teaches us that, under regularity conditions and given $q$ we can uniquely identity the distribution of $T$ without restricting 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 our analysis of the ACTG 181 data, we consider a function $q(\cdot)$ that contains more features than that in 7. Let $\phi = \{\phi_1, \phi_2\}$ be a two-dimensional censoring bias parameter that differentiates those who return to the study after missing several scheduled appointments and those who drop out of the study with unknown CMV shedding status. In this study, $M = 12$. 

Copyright © 2000 John Wiley & Sons, Ltd. Statist. Med. 2000; 00:0–0

Prepared using simauth.cls
The censoring bias function for ACTG 181 is

\[ q(t, l, r; \phi) = \phi_1 I(r < M + 1) \frac{(t - l)}{(M - 1)} + \phi_2 I(r = M + 1) \frac{(t - l)}{M}. \] (9)

This definition of \( q \) can incorporate the assumption that the probability ratio of having a particular censoring interval comparing participants with two different months of CMV shedding should differ by (1) the length of the censoring interval and (2) the difference in the months of shedding for the participants being compared. It also allows that probability ratios for a drop-out interval (e.g., \([l, M + 1]\) for some \( l < M \)) should be different from an interval with right endpoint \( r \) (\( r \leq M \)).

### 4.3.1. Interpretation of ACTG 181 censoring bias function

The selection-model interpretation of the parameter \( \exp\{\phi_1\} \) is the probability ratio of having censoring interval \([1 \text{ month, 12 months}]\) comparing those who begin CMV shedding in the blood during the twelfth month to those who begin shedding in the first month. If the interval were, say \([3 \text{ months, 9 months}]\), then \( \exp\{\phi_1 \frac{6}{9}\} \) is the probability ratio of having this interval comparing those who begin shedding at nine months versus those who begin shedding at three months. The selection-model interpretation of the parameter \( \exp\{\phi_2\} \) is the probability ratio of dropping out just after baseline comparing those who do not shed CMV in the blood within twelve months to those who begin shedding in the first month. For an interval \([l, M + 1]\) where \( l \in [1, \ldots, M] \), the probability ratio of dropping out at month \( l \) comparing those that shed just after \( l \) to those who do not shed within 12 months is \( \exp\{\phi_2 \frac{M + 1 - l}{M}\} \). 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. Using the pattern-mixture model.
interpretation, \( \exp\{\phi_1\} > 1 \) (\( \exp\{\phi\} < 1 \)) means that the relative probability of shedding CMV at the twelfth month compared to the first month is higher (lower) for patients censored into the interval [1 month, 12 months] than what would be assumed under CAR. Similarly, \( \exp\{\phi_2\} > 1 \) (\( \exp\{\phi\} < 2 \)) means that the relative probability of not shedding CMV within the twelve months compared to shedding during the first month is higher (lower) for patients who drop out just after baseline than what would be assumed under CAR.

5. LIKELIHOOD AND ESTIMATION

In this section, we describe the likelihood and estimation procedures for the frequentist and Bayesian approaches. In discrete time, the complete event-time data are assumed to follow the multinomial distribution. Let \( I_{ij} = I(T_i = j) \), the indicator that person \( i \) fails at time \( j, i = 1, \ldots, n, j = 0, \ldots, M + 1 \). The \( I_{ij} \)'s are the complete data. Let \( \delta_{ij} = I(j \in [L_i, R_i]) \), the indicator that time \( j \) is in person \( i \)'s censoring interval. 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 \( p_j = P(T = j) \) for \( j = 0, \ldots, M + 1 \), and \( p = (p_0, \ldots, p_M, p_{M+1}) \). We assume that time is partitioned such that the sufficient conditions for a unique solution hold.

To conform with this regularity condition, one can discretize the event times to be the smallest disjoint intervals as in Turnbull (1976), or according to some convenient measure of time (months, years, etc.).

5.1. Frequentist approach

To estimate the distribution of \( T \) for fixed \( q \), we utilize the EM algorithm. The complete-data likelihood is \( L(p) = \prod_{i=1}^{n} \prod_{j=0}^{M+1} p_j^{I_{ij}} \), and the log-likelihood is \( \ell(p) = \sum_{i=1}^{n} \sum_{j=0}^{M+1} I_{ij} \log(p_j) - \)
\[ \lambda(\sum_{j=0}^{M+1} p_j - 1) \]

where \( \lambda \) is a Lagrange multiplier that ensures the constraint \( \sum_{j=0}^{M+1} p_j = 1 \) is satisfied.

To perform the E-step at iteration \( s \), \( \sum_{i=1}^{n} I_{ij} \) in the complete-data log-likelihood are replaced by their expected values, given \( \delta_i \) and \( p^{(s-1)} \). When the CNAR model assumption is incorporated into the E-step, the resulting function to be maximized is

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

where \( p^{(s-1)} \) is the estimate of \( p \) at the \((s-1)\)th iteration, and

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

The M-step is performed by differentiating \( Q(p; p^{(s-1)}) \) with respect to \( p_j \), which results in a reweighted version of Turnbull’s self-consistency equation [6],

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

When \( q \) does not depend on \( t \) (e.g., \( q = 0 \)), CAR is assumed, and our estimator simplifies to that of Turnbull [6]. Standard errors can be estimated [32].

5.2. Bayesian approach

We evaluate the posterior distribution of the survival function by performing a Markov Chain Monte Carlo (MCMC) procedure. To do this, we need a prior distribution for \( p \). We assume a Dirichlet distribution for the probability of the event occurring within one of \( M + 2 \) intervals of time, where the time intervals are a partition of continuous time. Let \( B = \{b_0, \ldots, b_{(M+1)}\} \) be a base measure defined on \( E \), interpreted as the prior mean of \( p \). Let \( \alpha^* \) be a precision parameter that determines how concentrated the distribution is around \( B \), where the elements
of \( B \) sum to 1. Let \( \alpha_j = \alpha^* b_j \), for \( j = 0, \ldots, M + 1 \). The Dirichlet distribution is given by:

\[
f(p) = \frac{\Gamma(\alpha_0 + \cdots + \alpha(M+1))}{\Gamma(\alpha_0) \cdots \Gamma(\alpha_{(M+1)})} p_0^{\alpha_0-1} \cdots p_{(M+1)}^{\alpha_{(M+1)}-1},
\]

where \( p_0, \ldots, p_{(M+1)} \geq 0; \sum_{k=0}^{M+1} p_k = 1 \), the components of \( \alpha = \{\alpha_0, \ldots, \alpha_{(M+1)}\} \) are greater than zero, and the \( \alpha_j \)'s are interpreted as the ‘prior sample sizes’ of the event in period \( j \). In addition, we assume that the censoring bias function, \( q \), is indexed by a vector of censoring bias parameters, \( \phi \). We assume that, apriori, \( \phi \) and \( p \) are independent.

Because the data are incomplete, conjugate analyses cannot be performed. Therefore, Gibbs sampling [25] is performed via data augmentation [26] along with a Metropolis-Hastings step [31]. See Appendix 2 for details about the algorithm.

6. SIMULATION STUDY

In our simulation study, we set \( M = 4 \) (baseline visit, four follow-up visits, and right censoring) and used the censoring bias function parameterization (7). The true event time, \( T \), was drawn from a multinomial model with survivor function \{1.000, 0.657, 0.417, 0.254, 0.137, 0.070\}. Censoring intervals were drawn given \( T \). The pattern-mixture restrictions in Section 4.1 and the distribution of \( T \) are not enough to fully identify the distribution of the censoring intervals given \( T \). 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 \( \sum_{l \leq r} P(L = l, R = r) = 1 \).
and \( \sum_{\{(l,r) : l \leq t \leq r\}} P(T = t \mid L = l, R = r) P(L = l, R = r) = P(T = t) \). Values of the true parameters were chosen to produce 86% censoring (the probability of not exactly observing the event).

Let \( \phi_{\text{true}} \) denote the value of \( \phi \) that generated the data, and \( \phi_{\text{modeled}} \) denote the value of \( \phi \) that was assumed in the analysis. The true and modeled censoring bias parameters, \( \phi_{\text{true}} \) and \( \phi_{\text{modeled}} \), respectively, were combinations of \( \{-\log(2), 0, \log(2)\} \). We generated 1000 datasets for each of the nine analyses.

The degree of bias is assessed by comparing the mean of the estimated survivor function at times two and four, \( \hat{S}(2) \) and \( \hat{S}(4) \), respectively, to the true function. Also, the Monte Carlo standard deviation of the estimated survivor function is compared to the mean of the standard error estimates for both times. The results are found in Table 1. The first column is the true \( \phi \) that generated the censoring intervals, and column two is the \( \phi \) assumed in the CNAR model.

The results show little bias when \( \phi_{\text{true}} = \phi_{\text{modeled}} \), with bias increasing as the discrepancy between \( \phi_{\text{true}} \) and \( \phi_{\text{modeled}} \) increases. Also, regardless of the bias, we see good agreement between the empirical standard deviation and the standard error estimate.

7. DATA APPLICATION: 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 blood every 12 weeks, as discussed previously [33]. ACTG 181 is a substudy of ACTG 081, a randomized trial of methods to prevent AIDS-related opportunistic infections [34].

The data consist of 177 patients whose CMV shedding times were partitioned into 12 four-week ("monthly") periods, baseline, and a time that indicates no CMV shedding in the blood.
within the 12 months:

\[ \{(-\infty, 0], (0, 1], (1, 2], \ldots, (10, 11], (11, 12], (12, \infty)\} \].

Therefore, \( P(T = t) \) is the probability of shedding in the \( t \)th month when \( t = 1, \ldots, 12 \), the probability of shedding at or before enrollment when \( t = 0 \), or the probability of shedding after 12 months or never when \( t = 13 \). Using 12 months as the latest follow-up time, 5, 157, and 15 patients were left, right, and interval censored, respectively. The interval lengths for the interval-censored patients were 2, 3, 4, 6, and 8 months for 2, 8, 2, 2, and 1 patients, respectively. One hundred of the 157 right censored patients were known to not be shedding by twelve months. The remaining 57 right-censored patients dropped out before twelve months: 15, 9, 13, and 18 patients dropped out during the first, second, third, and fourth quarter of the year, respectively.

In this example, \( M = 12 \). We used the censoring bias function parameterization (9). Two sensitivity analysis approaches were used to analyze these data. In the first approach, we elicited information from Dr. Sam Bozzette, principal investigator of ACTG 181. As an AIDS expert with experience interacting with patients with advanced HIV, he has first-hand knowledge of the behavior of patients from this population. In addition, we considered a wide range of values for the censoring bias parameters that covers a wide range of assumptions.

We elicited a range of values for \( \exp\{\phi_1\} \) and \( \exp\{\phi_2\} \) from Dr. Bozzette. For both parameters, Dr. Bozzette’s range of values was [1.1, 2]. That is, he believes that those who began shedding during (after) month 12 were 1.1 to 2 times more likely to have the censoring interval \( (0, 12] \) \( ([0, \infty)) \) than those who began shedding during the first month. Another interpretation, using the pattern-mixture model, is that the relative probability of shedding
during (after) the twelfth month compared to the first month for those censored into the interval \((0, 12]\) \(((0, \infty))\) is 1.1 to 2 times higher than what would be assumed under CAR. Based on his experience, those who are healthier are less motivated to attend visits than those who are sicker and will not return until their health deteriorates. We also considered the range \([1/5, 5]\) for both \(\exp\{\phi_1\}\) and \(\exp\{\phi_2\}\). This range translates into the assumptions that those who began shedding during (after) month 12 were five times less to five times more likely to have the censoring interval \((0, 12]\) \(((0, \infty))\) than those who began shedding during the first month. Using the pattern-mixture approach, the relative probability of shedding during (after) the twelfth month compared to the first month for those censored into the interval \((0, 12]\) \(((0, \infty))\) is five times lower to five times higher than what would be assumed under CAR. This analysis allows us to see how much bias can occur over extreme departures from CAR.

A sensitivity analysis using the EM algorithm [12] can be performed with the preceding elicited information, however, prior distributions for the CMV shedding time probabilities and the censoring bias parameters are needed to perform a Bayesian analysis. To elicit the prior CMV shedding distribution, Dr. Bozzette was asked to describe his expected CMV shedding distribution prior to the ACTG 181 Study, and to describe the weight of his prior opinion relative to the combined information from the study and his opinion. He expects that 70% of the study participants will not shed within 12 months, and that the remaining 30% will begin shedding uniformly over the twelve months. ACTG 181 was one of the first studies of AIDS opportunistic infections, therefore Dr. Bozzette gave his prior opinion 20% of the weight. To elicit prior distributions for \(\phi\), we displayed histograms of beta densities, transformed to reflect the elicited ranges of \(\exp\{\phi\}\). For each censoring bias parameter, we asked Dr. Bozzette to choose the histogram that best reflects his prior beliefs. His beliefs correspond to prior modes.
of 1.8 and 1.3 for $\exp\{\phi_1\}$ and $\exp\{\phi_2\}$, respectively.

7.1. Frequentist analysis

In order to explore the sensitivity of the estimated CMV-shedding distribution to the elicited values of $\exp\{\phi\}$, we performed the EM algorithm at $\exp\{\phi\}$ of (1,1) (CAR), (1.1,1.1), and (2,2). The estimated survivor functions, along with those from the FGS method [10] shown in Figure 2 show that, with the exception of the first month, the results are robust to assumptions about the censoring mechanism within the elicited range. Also, the results based on the CAR assumption are more closely aligned with those based on elicited expert opinion than those based on the assumptions made in the FGS method [10]. The results based on extreme departures from CAR in Figure 2 show that the estimates at the first and fifth month are sensitive to $\exp\{\phi\}$, but are robust at every other time. Comparing these results with the FGS estimates suggests that the FGS assumptions correspond to $\exp\{\phi\} = (4,4)$ at months four and five and $\exp\{\phi\} = (1/2,1/2)$ at every other time [10]. The reasons for the overall robustness of results are sample specific. The censoring intervals were short, and more than half of the drop-outs were lost to follow-up after six months. The sensitivity of results during the first month is due to eight of the interval censored patients missing several visits after baseline before returning to the study, and twelve of the drop-outs leaving the study after their baseline visit.

7.2. Bayesian analysis

The Bayesian analysis was performed with point-mass priors at $\exp\{\phi\}$ of (1,1) (CAR), (1.1,1.1) and (2,2), and for random $\phi$. The first three analyses are part of a sensitivity analysis across the elicited values of $\phi$, while the last analysis mixes over the posterior distribution of
exp{\phi} to produce one answer. For each analysis, the MCMC was run for 5000 iterations, plus 500 more for burn-in. A previously published convergence diagnostic scheme was used [35].

The prior and various posterior distributions of \( P(T > 3),\ P(T > 6),\ P(T > 9),\ P(T > 12) \) in Figure 3 show little sensitivity across the range of assumptions. The nearly identical prior and posterior distributions of \( \exp{\phi} \) for the random-\( \phi \) analysis in Figure 4 reflect that \( \phi \) is not identified by the data. Lastly, the mean posterior survivor functions for the elicited-\( \exp{\phi} \) and extreme-\( \exp{\phi} \) are shown in Figure 5. The fixed-\( \exp{\phi} \) results are similar to the analogous frequentist results, due to the small weight of the prior relative to the data. The CAR, random-\( \exp{\phi} \), and elicited fixed-\( \exp{\phi} \) results are nearly identical, showing robust estimates across the elicited assumptions. The mean posterior survivor functions for extreme assumptions show some minor sensitivity when \( \exp{\phi} < (1,1) \). The random-\( \exp{\phi} \) results are between those based on \( \exp{\phi} = (1.1,1.1) \) and \( \exp{\phi} = (2,2) \). The mean posterior survivor functions suggest earlier shedding than the analogous estimated survivor functions from the frequentist analyses, due to some shrinkage to the prior expected survivor function.

8. DISCUSSION

By parameterizing the departure from CAR, our method presents a more honest characterization of uncertainty, which stems from both sampling variability and ignorance of the censoring mechanism, than procedures that rely on a single set of unverifiable assumptions. Our methodology facilitates interactions between statisticians and scientists, which, in our opinion, leads to more transparent inferences. The elicitation approach raises scientists’ awareness of the assumptions associated with standard statistical procedures. Incorporating information from prior scientific work provides a medium through which statisticians obtain
a better understanding of the scientific context of the study. The method is flexible enough to combine information from several experts and past studies. Various approaches for eliciting and combining expert information have been described [36].

The analyses of the ACTG 181 data exemplify several benefits of our methods. First, we verified that the estimated survivor function is robust to minor departures from CAR, based on elicited expert information. Second, we discovered that the estimates are only sensitive to extreme departures from CAR. If the study had longer censoring intervals, or more early drop-outs, the results would have been more sensitive to the assumptions. Also, if more censoring intervals covered later times (i.e., times when fewer patients are at risk), then more sensitivity would have been observed at these later times.

Another benefit of the proposed method is that additional experts who may disagree with the elicited assumptions can verify whether or not the results are robust to their own assumptions. In contrast, the assumptions made in the FGS method [10] do not correspond well, net of sampling variability, to the elicited opinion in this example. Also, while the assumptions allow for CNAR, they may not be plausible, and it is not clear how to perform a low-dimensional sensitivity analysis that explores deviation from their assumption.

The proposed frequentist method is an example of a subjective frequentist procedure [19]. In that paper, the author reminds the reader that objective analyses, while ideal, are not always possible, especially when data are incomplete. Our methods are not meant to replace an objective method, but rather, they allow us to analyze data that require subjective beliefs. Further, in the presence of coarsened data, all inference includes an element of subjectivity, as unidentifiable assumptions must be made to analyze the data. This statement is true for both our approach and that of FGS [10]. Although the approach by FGS has the benefit of
including auxiliary data (visits outside of the censoring interval), they did not allow deviations from their assumption of CAR within levels of auxiliary data [10]. Therefore, while their results are based on a single subjective assumption, ours are based on a range of assumptions.

We recognize that it is more difficult to make decisions based on the results of a sensitivity analysis. The fully Bayesian procedure facilitates decision making by presenting a single summary. Of course, the results will depend on the priors. Decision makers, working with statisticians, can identify plausible priors that would lead to unacceptable levels of risk.

**APPENDIX 1: UNIQUENESS**

In this Appendix, we establish conditions, following the work of Gentleman and Geyer [7], that are sufficient to uniquely solve Equations (8). Throughout, we assume that Model (3,4) holds for a specified function $q$. Let $h_{l,r} = P(L = l, R = r)$, where $[l, r] \in E^*$ and $N_M$ be the number of support points of the distribution of $[L, R]$ for which $h_{l,r} > 0$.

We first show that solving (8) is equivalent to solving a constrained optimization problem.

**Theorem 1.** Solving (8) is equivalent to optimizing

$$
\mathcal{E}^*(p; q) = \sum_{[l,r] \in E^*} h_{l,r} \log \left( \sum_{t \in [l,r]} \exp \{ q(t, l, r) \} p_t \right)
$$

subject to constraints

$$
1 - \sum_{t=0}^{M+1} p_t = 0, \quad \text{(11)}
$$

$$
p_t \geq 0, \quad t = 0, \cdots, M + 1. \quad \text{(12)}
$$

Copyright © 2000 John Wiley & Sons, Ltd. Statist. Med. 2000; 00:0−0
The Equations (8) can be rewritten as

\[
0 = \sum_{[l,r] \in E^*} \frac{h_{l,r} \exp\{q(t,l,r)\}p_t}{\sum_{s \in [l,r]} \exp\{q(s,l,r)\}p_s} - p_t. \tag{13}
\]

Dividing both sides of (13) by \(p_t\) and integrating \(\sum_{[l,r] \in E^*} h_{l,r} \exp\{q(t,l,r)\}p_t\) with respect to \(p_t\) results in the objective function \(\ell^*(p; q) = \sum_{[l,r] \in E^*} h_{l,r} \log \left[ \sum_{t \in [l,r]} \exp\{q(t,l,r)\}p_t \right]\). Therefore, solving (8) is equivalent to optimizing \(\ell^*(p; q)\) subject to the constraints (11) and (12).

**Theorem 2.** The objective function \(\ell^*(p; q)\) is concave in \(p\).

**Proof:** Let \(A(q)\) be the \(N_M \times (M + 2)\) matrix with elements \(I(t \in [l,r]) \exp\{q(t,l,r)\}\) for \([l,r] \in E^*\), and \(h_{l,r} > 0\). Let \(D(p; q)\) be a \(N_M \times N_M\) diagonal matrix with elements \(-h_{l,r}/\sum_{t=0}^{M+1} p_t \exp\{q(t,l,r)\}\), for \(t \in 0, \ldots, M + 1, \ [l,r] \in E^*\), and \(h_{l,r} > 0\). The matrix \(D(p; q)\) is negative definite for all \(p\) since all of its elements are negative. The Hessian of \(\ell^*(p; q), H(p; q)\), is equal to \(A(q)^t D(p; q) A(q)\). \(A(q)^t D(p; q) A(q)\) is negative semi-definite in \(p\) because \(x^t (A(q)^t D(p; q) A(q)) x = (A(q)x)^t D(p; q) A(q)x \leq 0\) for \(x\). This implies that \(\ell^*(p; q)\) is concave in \(p\).
that, for \( t = 0, \ldots, M + 1 \),

\[
\lambda_t p_t^* = 0 \tag{14}
\]

\[
\lambda_t \geq 0 \tag{15}
\]

\[
\frac{\partial}{\partial p_t} \{ \ell^*(p; q) + \sum_{t=0}^{M+1} p_t (\lambda_t - \lambda) \} |_{p^*} = d_t + \lambda_t - \lambda = 0, \tag{16}
\]

where \( d_t = \sum_{[l,r] \in E} \frac{h_{l,r} I(t \in [l,r]) \exp \{ q(t,l,r) \}}{\sum_{s \in [l,r]} \exp \{ q(s,l,r) \}} p_s \).

**Theorem 3.** If \( \text{rank}(A(q)) = M + 2 \), there is a unique maximum, \( p^* \), of \( \ell^*(p; q) \), subject to constraints (11,12).

**Proof:** When \( \text{rank}(A(q)) = M + 2 \), \( A(q)x = 0 \) if and only \( x = 0 \). This implies that \( H(p; q) \) is negative definite in \( p \) since \( x'H(p; q)x = (A(q)x)'D(p; q)(A(q)x) < 0 \) for all \( x \neq 0 \). Thus, \( \ell^*(p; q) \) is strictly concave in \( p \) and the maximum must be unique. \( \bullet \)

If \( \ell^*(p; q) \) is not strictly concave in \( p \), the optimum may still be unique. To determine a sufficient condition for uniqueness, partition the matrix \( A(q) \) into \([A_1(q), A_2(q)]\) where \( A_1(q) \) contains the columns with \( \lambda_t^* > 0 \) and \( A_2(q) \) contains the columns with \( \lambda_t = 0 \).

**Theorem 4.** If \( \text{rank}(A_2(q)) \) is equal to its number of columns, then there is a unique maximum, \( p^* \), of \( \ell^*(p; q) \), subject to constraints (11,12).

**Proof:** Since \( \text{rank}(A_2(q)) \) is equal to its number of columns, we know that \( A_2(q)x_2 \neq 0 \) for all \( x_2 \neq 0 \). Thus, since \( D(p; q) \) is negative definite, we know that \( (A_2(q)x_2)'D(p; q)(A_2(q)x_2) = x_2'A_2(q)'D(p; q)A_2(q)x_2 < 0 \) for all \( x_2 \neq 0 \). This implies that Theorem 9.3.2 of Fletcher [37] is satisfied and \( p^* \) is unique. \( \bullet \)
APPENDIX 2: BAYESIAN ALGORITHM

The algorithm for performing Bayesian analysis begins by selecting starting values for the censoring bias parameters, $\phi^{(0)}$, and the event-time probabilities, $p^{(0)}$. Let $I$ be the complete data and $\delta$ be the observed data for all individuals. The algorithm proceeds in three steps for iteration $s = 1, \ldots, N_{\text{sim}}$:

1. Simulate $I^{(s)}$ from $p(I \mid \delta, p^{(s-1)}, \phi^{(s-1)})$.
2. Simulate $p^{(s)}$ from $p(p \mid \delta, \phi^{(s-1)}, I^{(s)}) = p(p \mid I^{(s)})$.
3. Simulate $\phi^{(s)}$ from $p(\phi \mid \delta, I^{(s)}, p^{(s)})$,

where $p(\cdot)$ denotes the density.

The imputed event times from Step 1 for person $i$ are simulated from

$$
\prod_{j=0}^{M+1} \left[ \frac{-\delta_{ij} p_{ij}^{(s-1)} \exp\{q(j, l, r, \phi^{(s-1)})\}}{\sum_{k=0}^{M+1} \delta_{ik} p_{ik}^{(s-1)} \exp\{q(k, l, r, \phi^{(s-1)})\}} \right]^{I_{ij}}
$$

a truncated multinomial, and are aggregated into frequencies at each time period for each group. Let $n_{ij}^{(s)}$ denote the simulated event frequency during period $j$ at the $s$th iteration, and $n^{(s)} = \{n_{0}^{(s)} \ldots n_{M+1}^{(s)}\}$. Given these frequencies, $p$ is independent of $\phi$ and $\delta$. So, $p^{(s)}$ can be simulated in Step 2 from $p(p \mid I^{(s)}) = p(p \mid n^{(s)})$, which is just a Dirichlet distribution with the $\alpha$ in Equation 10 replaced by $\alpha + n^{(s)}$. The $\phi$ are simulated in Step 3 via a Metropolis-Hastings step [31]. Let $I^{(s)}$ denote the $s$th iteration vector of simulated event indicators. The candidate, $\phi^*$, is simulated from the jumping distribution at iteration $s$, $J_s(\phi^* \mid \phi^{(s-1)})$, and is accepted with probability

$$
\min(1, r_{MH}) = \frac{p(I^{(s)} \mid \delta, p^{(s)}, \phi^*) p(\phi^*) J_s(\phi^{(s-1)} \mid \phi^*)}{p(I^{(s)} \mid \delta, p^{(s)}, \phi^{(s-1)}) p(\phi^{(s-1)}) J_s(\phi \mid \phi^{(s-1)})}
$$

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, 5R01AI132475, R01CA74112, 1-R01-MH56639-01A1, and 1- R01-DA10184-01A2. The authors thank Tom Louis and Zhiqiang Tan for helpful discussions, Rebecca Betensky and Dianne Finkelstein for inspiring this work as well as providing access to data from ACTG 181, David Vlahov and Noya Galai for thoughts and ideas on elicitation, and an anonymous reviewer for thoughts and suggestions that greatly improved this paper.
references


Table I. Simulation results. 1000 iterations, $M = 4$, $n = 200$, 86% censoring. Bold indicates correct values for $\phi_{modeled}$. $S(2) = 0.417$, $S(4) = 0.137$.

<table>
<thead>
<tr>
<th>$\phi_{true}$</th>
<th>$\phi_{modeled}$</th>
<th>$\hat{S}(2)$ (SE)</th>
<th>$\hat{S}(4)$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>0.417 (0.046)</td>
<td>0.137 (0.030)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>0</td>
<td>$-\log(2)$</td>
<td>0.481 (0.050)</td>
<td>0.169 (0.036)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>0</td>
<td>0.516 (0.046)</td>
<td>0.214 (0.033)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>0</td>
<td>0.343 (0.046)</td>
<td>0.118 (0.026)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.043)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>0</td>
<td>$-\log(2)$</td>
<td>0.415 (0.051)</td>
<td>0.137 (0.029)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.050)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>0</td>
<td>0.481 (0.047)</td>
<td>0.154 (0.034)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.047)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>$-\log(2)$</td>
<td>$-\log(2)$</td>
<td>0.284 (0.042)</td>
<td>0.104 (0.023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.041)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>0</td>
<td>$-\log(2)$</td>
<td>0.360 (0.050)</td>
<td>0.116 (0.027)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.050)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>$\log(2)$</td>
<td>0</td>
<td>0.419 (0.050)</td>
<td>0.136 (0.030)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.050)</td>
<td>(0.029)</td>
</tr>
</tbody>
</table>
Figure 1. Impact of $\phi$ on $P(T = t \mid L = 1, R = 4)$ when $P(T = t) = P(T = t')$ for all $t, t' \in E$.
Figure 2. Frequentist: Time to blood CMV shedding.
Figure 3. Densities of blood CMV shedding.
Figure 4. Densities of censoring bias parameters.

Figure 5. Bayesian: Time to blood CMV shedding.