The International Journal of Biostatistics

Volume 8, *Issue* 1 2012 *Article* 4

Estimation of the Mean Frequency Function for Recurrent Events when Ascertainment of Events Is Delayed

T Charles Casper, *University of Utah*
Thomas D. Cook, *University of Wisconsin, Madison*

Recommended Citation:
Casper, T Charles and Cook, Thomas D. (2012) "Estimation of the Mean Frequency Function for Recurrent Events when Ascertainment of Events Is Delayed," *The International Journal of Biostatistics*: Vol. 8: Iss. 1, Article 4.

DOI: 10.1515/1557-4679.1303

Estimation of the Mean Frequency Function for Recurrent Events when Ascertainment of Events Is Delayed

T Charles Casper and Thomas D. Cook

Abstract

In many large clinical trials there are delays between the time at which events occur and the time at which they are reported. Estimators of the mean frequency function for recurrent events that are currently used are inconsistent in these circumstances. We propose two new estimators to be used when events are reported with delay. One method is a basic inverse probability of censoring weighting approach, while the other explicitly estimates the distribution of the reporting delays. The asymptotic properties of these estimators are discussed and variance estimators are given. We examine the results of simulations comparing the new estimators to each other and to existing estimators that do not properly account for the delays. We also calculate some of these quantities using data from TNT, a clinical trial in which there were delays and events of interest were recurrent.

KEYWORDS: counting process, delayed ascertainment, interim analysis, inverse probability weighting, recurrent events

Author Notes: The authors would like to thank Ronald Gangnon and the reviewers of this article for their insightful comments and suggestions.

1 Introduction

1.1 Background

In many medical studies, events of interest are recurrent in nature, possibly including several types such as non-fatal MI, hospitalization, or worsening heart failure. Many authors have discussed estimation of parameters of interest when dealing with recurrent non-fatal events in the presence of death (Ghosh and Lin, 2000; Cook and Lawless, 1997; Li and Lagakos, 1997). In large, multicenter clinical trials, there is usually a delay between the time an event occurs and the time it is reported to the statistical analysis center. At the conclusion of such trials, the final analyses will be performed after collection of all data, giving the most reliable analysis of the whole trial. For interim analyses, however, only data that have been reported at the time of the analysis can be used to provide timely information to DSMBs, steering committees, or sponsors. While hypothesis testing should be considered the primary purpose of a clinical trial and early stopping a primary consideration at interim analyses, estimation of event rates is typically performed and provides useful information. Estimates can be used to check design assumptions, to assess sample size requirements, to predict the time at which an event target will be reached, and to plan for when the trial will be finalized. In the setting of delayed reporting, existing estimators of mean event frequency are inconsistent. Our goal is to use the available, reported information to obtain the most reliable estimates possible.

One of the most commonly used methods for dealing with delayed ascertainment is that of back-censoring. In the usual survival setting, rather than censoring event times for event-free individuals on the date of the analysis, the censoring date is set to 6 months, or some other, often arbitrary, amount of time, in the past. This means that any reported events that have occurred in the most recent 6 months are ignored. When this is done, censoring is purely administrative and no dependence is induced between the survival time and the censoring time. In the recurrent event case, if there is no possibility of any delay surpassing 6 months, then using this convention, the usual Ghosh and Lin (2000) estimator (GL) is consistent. In this situation the estimator is inefficient, however, as many events are discarded. Even worse, no estimation is possible in the tail. In this situation, one must also choose the value to be used for the maximum possible delay. This is a difficult task due to the fact that the delays are right-truncated—events with longer delays are less likely to be observed by the time of the analysis. The distribution of the delays observed in TNT, a clinical trial that will be described in Section 4, is shown in Figure 1. While a large majority of delays were less than 6 months, many were greater and some were many times greater. Figure 2 shows estimated mean frequency of recurrent events for both treatment groups combined in the same trial. It represents

a hypothetical interim estimate at 3 years after the first subject was randomized. The true GL, which could be calculated if there were no delays, is shown for reference. At the time of the analysis, however, the delays make this impossible to calculate. Also shown are the GL that makes no adjustment for delays and the GL that cuts off the data 6 months before the analysis. The GL that ignores the last 6 months is somewhat better, but only provides an estimate through 2.5 years. This illustrates the bias introduced by reporting delays and the motivation for developing new methods.

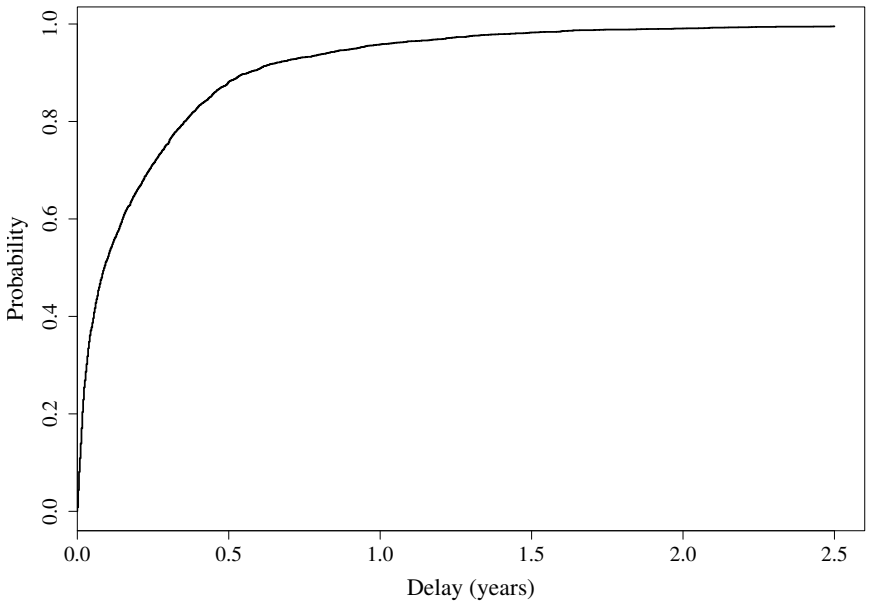

Figure 1: The cumulative distribution of delays in TNT.

Another method is often used in trials that monitor patients at regular visits. At the time of analysis, if a subject is not known to have experienced an event since the most recent clinic visit, the subject is censored at that time. There are several problems with this approach, however. First, the data collected at these visits may not be reported immediately, so it may not adequately address the original problem. Also, events are commonly not reported until after one or more clinic visits have occurred. Finally, and most important, this censoring convention induces dependence between the event times and the censoring times. If an event of interest has not occurred and been reported by the time of analysis, the censoring time is set

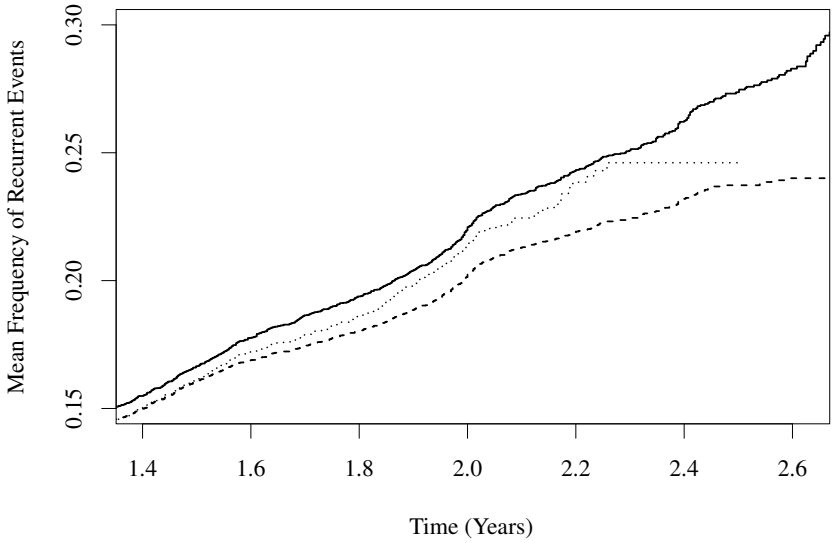

Figure 2: A graph of estimates of mean frequency for an interim analysis at 3 years: the "true" GL estimate (solid), the unadjusted GL estimate (dashed), and the GL estimate that ignores the last 6 months (dotted).

in the past. Otherwise, it is implicitly set to the time of analysis. This violates one of the required assumptions, that event times and censoring times are independent.

For the case of survival estimation for time to a terminal event with delayed ascertainment, Hu and Tsiatis (1996) presented an estimator that correctly accounts for delays when patients are monitored at regular visits. The estimator makes use of both monitoring times and corresponding reporting times, but the authors point out that these times are not usually recorded. Van der Laan and Hubbard (1998) and Hubbard et al. (2000) produced two more general consistent estimators that do not rely on regular monitoring. The first, called the "simple" estimator, is an inverse probability of censoring weighted (IPCW) estimator. The other, the "one-step" estimator, is based on semiparametric efficiency theory, attains optimal efficiency under certain assumptions, and is asymptotically equivalent to the simple estimator in the absence of informative covariates. No methods have been developed to account for delays when the events of interest are of multiple types or recurrent in nature. In this paper, we propose two methods. The first could be considered an extension of the terminal event IPCW concept into a recurrent-event setting. The second is

similar, but involves direct estimation of the distribution of reporting delays.

1.2 Data and Notation

We use a counting process approach to describe the data. We will use as the full data for a subject

$$X = \{N(t), V(t), t \in [0, \tau]\},$$

where $N(t)$ is the process that has a jump of size one at each recurrent event, $V(t)$ is the reporting process, and τ is a maximum follow-up time. If there is an event at time t, $V(t)$ is the reporting time for that event. If there is no event at time t, $V(t)$ is equal to the reporting time of the earliest event after t. In the case that no event occurs after t, we let $V(t) = t$. The definition of $V(t)$ is somewhat arbitrary for non-event times because, in the end, only values at event times will play a role. Let C be the censoring time or time of analysis. We will use an asterisk to denote observed counterparts of the full data, so we have observed data

$$Y = (C, \{N^*(t), V^*(t), t \in [0, C]\}),$$

where $N^*(t)$ has a jump at each observed event and $V^*(t)$ gives the reporting process for observed events. When t is an observed event time, $V^*(t)$ is that event's reporting time. When t is not an observed event time, $V^*(t)$ is the reporting time of the earliest observed event occurring after t. If no event is observed to occur after t, $V^*(t) = t$. Defined in this way, V and V^* are well-defined, left-continuous functions.

We are interested in estimating the parameter $\mu(t)$, defined as the expected number of recurrent events in $(0, t]$, based on n replicates of Y. Ghosh and Lin (2000) discuss the estimation and interpretation of this parameter with no delays. We use two approaches to estimate this parameter when ascertainment of events is delayed. One method is similar to the IPCW method used for terminal events by van der Laan and Hubbard (1998). We discuss the key assumption made when using this estimator and propose another method that relies on more plausible assumptions and explicitly estimates the delay distribution.

1.3 Outline of Paper

In the following section, we discuss nonparametric estimation of the mean frequency function when events are reported immediately and present two new estimators to be used when reporting is delayed. In Section 3, we report the results of a series of simulations. Section 4 contains the description of a clinical trial in which

events of interest were recurrent and reporting was delayed. Our methods are then applied to this dataset. Some final remarks are given in Section 5.

2 Methods

2.1 Estimating Mean Frequency with No Delays

First, we review the estimation of μ when reporting of events occurs immediately, as discussed by Ghosh and Lin (2000). Let T be a subject's time of death. The parameter of interest, $\mu(t) = E[N(t)]$, is defined as a marginal expectation; that is, it implicitly takes into account the termination of the recurrent-event process by death. We assume that C is independent of T and N. With no delays, $V(t) = t$ and, for $t \leq C$, $N^*(t) = N(t)$. Let

$$\Lambda(t) = \int_0^t E\{dN(s) \mid T \geq s\}.$$

We have the following relationship:

$$d\mu(t) = S(t)d\Lambda(t),$$

where $S(t) = P(T \geq t)$. We can estimate $\Lambda(t)$ with

$$\Lambda_n(t) = \int_0^t \frac{d\overline{N}^*(s)}{\overline{Y}(s)},$$

where $d\overline{N}^*(s) = \sum_{i=1}^n dN_i^*(s)$ and $\overline{Y}(s) = \sum_{i=1}^n I(C_i \geq s, T_i \geq s)$ (here, $I(\cdot)$ denotes the indicator function). The resulting estimator of $\mu(t)$ is

$$\mu_n(t) = \int_0^t \widehat{S}(s)d\Lambda_n(s),$$

where $\widehat{S}$ is the Kaplan–Meier estimator of S.

2.2 An IPCW Estimator

We now turn to the situation in which events are reported with delay. We propose an IPCW approach to estimate the mean frequency for the recurrent-event process that can be used both when only recurrent events are possible and when there is a terminating event. We will assume that censoring times for all subjects are observed. This is a reasonable assumption to make for interim analyses of clinical trial data.

An event of interest is observed at time t if and only if (1) there is an event at time t and (2) that event is reported at or before the censoring time. Thus, we have the relationship $dN^*(t) = I\{C \geq V(t)\}dN(t)$, and an IPCW estimator analogous to the simple estimator discussed by van der Laan and Hubbard (1998) and Hubbard et al. (2000) is

$$\tilde{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i^*(s)}{\overline{G}_n\{V_i(s)\}},$$

where $\overline{G}_n(t)$ is the empirical estimate of $\overline{G}(t) \equiv P(C \geq t)$. The assumption here is that $\overline{G}\{V(t)\} > 0$ F_X–almost everywhere so that

$$\begin{aligned} E\left[\int_0^t \frac{dN^*(s)}{\overline{G}\{V(s)\}} \mid X\right] &= \int_0^t \frac{dN(s)}{\overline{G}\{V(s)\}} E\left[I\{C \geq V(s)\} \mid X\right] \\ &= N(t), \end{aligned} \tag{1}$$

almost everywhere, and

$$E\left[\int_0^t \frac{dN^*(s)}{\overline{G}\{V(s)\}}\right] = \mu(t).$$

In the absence of delays and without death, $\overline{G}_n\{V_i(s)\} = \overline{G}_n(s) = 1/n\overline{Y}(s)$, $\widehat{S} = 1$, and $\tilde{\mu}(t)$ is equal to the Ghosh and Lin (2000) estimator. When there are no delays, but mortality is a consideration, the difference between these two estimators is minimal.

The assumption that $\overline{G}\{V(t)\} > 0$ is a strong assumption and unlikely to hold when censoring is purely administrative, which is the case for interim analyses in clinical trials. The assumption implies that any reporting time will be less than the maximum possible censoring time with probability one. This forces the delay times to be compressed for events occurring later in study time. If the delays are bounded above by some time, d_0, then consistent estimation is still achieved for the set of times

$$\left\{t : \overline{G}(t + d_0) > 0\right\}.$$

To see this, we consider equation (1). Now, if t is such that $\overline{G}(t + d_0) > 0$, then $\overline{G}(s + d_0) > 0$ for all $s \leq t$. We also have $V(s) \leq s + d_0$. In summary, if $\overline{G}(t + d_0) > 0$, then $\overline{G}\{V(s)\} > 0$ for all $s \leq t$ and

$$E\left[\int_0^t \frac{dN^*(s)}{\overline{G}\{V(s)\}} \mid X\right] = N(t),$$

as desired. Outside this range of times, the method will likely produce good estimates, having lower bias than the naïve estimator, but consistency is no longer

guaranteed. As the distance from the valid set of times increases, the asymptotic bias should increase.

Of course, even if the parameter d_0 exists, it will not be known in practice. It could be estimated from observed delays, but this estimator may not be reliable because the observed delays are right truncated. That is, longer delay times are associated with later reporting times and are, therefore, less likely to be observed. In this case, it is not only the value of the delay that is missing, but also the knowledge of its existence.

2.2.1 Asymptotic Properties of $\tilde{\mu}$

We now introduce some notation to help describe the asymptotic distribution of $\tilde{\mu}$. Let

$$\begin{aligned} \varphi(t) &= \int_0^t \frac{dN^*(s)}{\overline{G}\{V(s)\}} - \mu(t), \qquad (2) \\ E_t^{(1)}(s) &= E\left[\int_0^t \frac{dN^*(u)}{\overline{G}\{V(u)\}} \mid C > s\right], \\ E_t^{(2)}(s) &= E\left[\int_0^t \frac{dN^*(u)}{\overline{G}\{V(u)\}} \mid C = s\right], \end{aligned}$$

and

$$dM_G(s) = I(C=s) - d\Lambda_C(s) I(C \geq s),$$

where $\Lambda_C(s)$ is the cumulative hazard of C. Note that $dM_G(s)$ is the martingale corresponding to the process that jumps at C. Finally, let

$$\varphi_{\texttt{eff}}(t) = \varphi(t) - \int \left\{ E_t^{(2)}(s) - E_t^{(1)}(s) \right\} dM_G(s). \qquad (3)$$

We prove the following theorem in the appendix.

Theorem 2.1 *Assume that $C_1, \ldots, C_n$ are independent of $N_1(t), \ldots, N_n(t)$. Also, assume that there exists d_0 such that $P\{V(t) - t \leq d_0\} = 1$. For any t such that $\overline{G}(t + d_0) > 0$, $\varphi_{\texttt{eff}}(t)$ is the efficient influence curve and*

$$n^{1/2}\{\tilde{\mu}(t) - \mu(t)\} \;\rightarrow\; N\{0, \sigma^2(t)\},$$

in distribution, where $\sigma^2(t) = var\{\varphi_{\texttt{eff}}(t)\}$. Thus, $\tilde{\mu}(t)$ is asymptotically efficient.

To estimate the variance of $\tilde{\mu}(t)$, we must find consistent estimates of $E_t^{(1)}(s)$ and $E_t^{(2)}(s)$. For the first, we can use the mean of $\int_0^t [\overline{G}_n\{V_i(u)\}]^{-1} dN_i^*(u)$ for all

subjects that satisfy $C_i > s$. Call this estimate $\widehat{E}_t^{(1)}(s)$. To estimate the second, we can assume a regression model for $\int_0^t [\overline{G}_n\{V(u)\}]^{-1} dN^*(u)$ on C:

$$E\left[\int_0^t \frac{dN^*(u)}{\overline{G}_n\{V(u)\}} \mid C\right] = g(C)$$

with an appropriate function g. We then use the corresponding fitted value and call this estimate $\widehat{E}_t^{(2)}(s)$. Similar methods are described by van der Laan and Robins (2003) for other data models. Finally, we must also estimate $dM_G(s)$, but this has been done by estimating $\overline{G}$. Now, a consistent estimate of $\sigma^2(t)$ is

$$\hat{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^t \frac{dN_i^*(s)}{\overline{G}_n\{V_i(s)\}} - \tilde{\mu}(t) - \int\left\{\widehat{E}_t^{(2)}(s) - \widehat{E}_t^{(1)}(s)\right\} dM_{G_n,i}(s)\right]^2.$$

2.3 Estimating the Delay Distribution

The approach that was described in the previous section relies on the assumption that $\overline{G}\{V(t)\} > 0$ almost everywhere, an assumption that is technical, unintuitive, and untenable in most real situations. The result is consistent estimation over only a restricted set of times. An alternative approach is one that nonparametrically estimates the delay distribution. This approach relies on the assumption that the delays are independent, identically distributed, and independent of the event process. Let $H(t \mid C)$ be the probability that an event occurring at time t would be observed, given censoring time C. Then

$$\begin{aligned} E\{dN^*(t)\} &= E[E\{dN^*(t) \mid C\}] \\ &= E[E\{I(\text{event at } t \text{ observed}) dN(t) \mid C\}] \\ &= E\{H(t \mid C)\} d\mu(t). \end{aligned}$$

If H is known, then the value of $E\{H(t \mid C)\}$ can be consistently estimated by $n^{-1}\sum_1^n H(t \mid C_i)$.

To get an estimate of $H(t \mid c)$ for any t and c, we first note that this is equivalent to estimating the delay distribution. We also note that the observed delays are right truncated; we are less likely to observe delays that are large. Let t_j, $j = 1, \ldots, J$, be all event times, $V(t_j)$ the reporting time of event j, and C_j the censoring time of the subject who experienced event j. $D_j = V(t_j) - t_j$ is the reporting delay, which is only observed if $D_j \leq C_j - t_j$. Define $Z_j(s) = I\{D_j \leq s \leq C_j - t_j\}$,

$\overline{Z}(s) = \sum_{j=1}^{J} Z_j(s)$, $K_j(s) = I\{D_j \leq s\}$, and $d\overline{K}(s) = \sum_{j=1}^{J} dK_j^*(s)$. A consistent estimator of H is given by

$$\widehat{H}(t|c) = \prod_{u>c-t} \left(1 - \frac{d\overline{K}(u)}{\overline{Z}(u)}\right).$$

The form of the model can be flexible, but this approach seems to work well in practice. Now, another consistent estimator of μ is

$$\hat{\mu}(t) = \int_0^t \frac{d\overline{N}^*(s)}{\sum_{i=1}^{n} \widehat{H}(s \mid C_i)}.$$

The consistency of this estimator holds for any t less than the maximum possible censoring time, given that the maximum possible delay time, d_0, is less than the maximum possible censoring time.

As with $\tilde{\mu}$ and the Ghosh and Lin (2000) estimator, $\hat{\mu}$ can be used in the presence of a terminating event and it can be shown that, like $\tilde{\mu}$, $\hat{\mu}$ is equal to the Ghosh and Lin (2000) estimator when subjects do not die and events are reported immediately. A remarkable property of both of these estimators is that, when there is a terminating event, there is no need to estimate its survival distribution. This also means that whether or not the terminating event is subject to reporting delay is irrelevant. As in the case of the Ghosh and Lin estimator with no delays, no assumptions (for example, independence) about the relationship between recurrent events and the terminating event are necessary.

It is also possible for some subjects to have two levels of censoring. For example, suppose one of the centers in a multicenter trial shuts down and discontinues follow-up prior to the time of analysis. Events for subjects at this center will only be observed if they occur prior to the time follow-up is discontinued and are reported prior to the time of analysis. The estimator can be easily modified to accommodate such censoring.

2.3.1 Asymptotic Properties of $\hat{\mu}$

Let

$$U(t) = \frac{I\{\sum_{i=1}^{n} H(t \mid C_i) > 0\}}{\sum_{i=1}^{n} H(t \mid C_i)}$$

and

$$M_i(t) = N_i^*(t) - \int_0^t H(s \mid C_i) d\mu(s),$$

$i = 1, \ldots, n$. Now, let

$$\sigma^2_{R,n}(t) = nE\left[\left\{\int_0^t U(s)dM_i(s)\right\}^2\right].$$

In the appendix, we prove the following theorem regarding the asymptotic properties of $\hat{\mu}$.

Theorem 2.2 *Assume that there exists an integer N_{max} such that $P\{N(t) \leq N_{max}\} = 1$ and a value τ such that $\overline{G}(\tau) > 0$. For $t \in (0, \tau)$,*

$$n^{1/2}\left[\hat{\mu}(t) - \int_0^t I\left\{\sum_{i=1}^n \widehat{H}(s \mid C_i) > 0\right\} d\mu(s)\right] \;\rightarrow\; N\{0, \sigma^2_R(t)\},$$

in distribution, where $\sigma^2_R(t) = \lim_{n\to\infty} n\sigma^2_{R,n}(t)$, provided this limit exists.

This result suggests that a good estimator for the variance of $\hat{\mu}$ is

$$\hat{\sigma}^2_{R,n}(t) \equiv \sum_{i=1}^n \left[\int_0^t \widehat{U}(s)\{dN_i^*(s) - \widehat{H}(s \mid C_i)d\hat{\mu}(s)\}\right]^2,$$

where $\widehat{U}(s)$ is obtained by substituting $\widehat{H}$ for H in the expression for U.

3 Simulations

In order to examine the performance of the estimators, we performed a series of simulations, generating data under six different scenarios. In each setting, it is assumed that time of randomization is uniformly distributed on $(0,2)$ and that the time of analysis is 2 (years). The delay times were generated using the uniform distribution on the interval $(0, d_0)$.

In the first three scenarios (I–III), data are generated according to a gamma frailty model. Formally, let θ_i, $i = 1, \ldots, n$, have gamma distribution with mean 1 and variance 1. Given θ_i, the event process for subject i is assumed to be a Poisson process with frequency $\theta_i\mu(t)$. Here, we use $\mu(t) = 5t$. The delays are i.i.d. and the upper bound, d_0, is equal to 0.5, 1, and 1.5 in cases I, II, and III, respectively. Scenario IV represents a homogeneous Poisson process with rate 3 and reporting delays that are i.i.d. uniform with $d_0 = 1$.

For settings V and VI, the recurrent events are generated according to the frailty model used for scenarios I–III. For setting V, the delays generated are uniformly distributed, but not i.i.d. For an event occurring at time t (after the first randomization), the upper bound of the distribution, d_0, now depends on t and is equal

Table 1: Simulations to Compare Mean Frequency Estimators

Scenario	Events	Delays	Death
I	Frailty Model	U(0,0.5)	No
II	Frailty Model	U(0,1)	No
III	Frailty Model	U(0,1.5)	No
IV	Poisson Process (rate= 3)	U(0,1)	No
V	Frailty Model	non-i.i.d.	No
VI	Frailty Model	U(0,1)	Exp(2)

to $1.5 - t/4$. Thus, an event that occurs very early in the "trial" will have an expected delay in reporting of about 9 months, while an event occurring near the two-year mark, will have an expected delay of about 6 months. This scenario assumes that event reporting improves with time. For scenario VI, with i.i.d. uniform$(0,1)$ delays, subjects can also experience death. Time of death is simulated for each individual according to an exponential distribution with mean 2 years and terminates the recurrent-event process. A summary of these six situations is given in Table 1.

For each simulated set of data, several estimators are used to estimate the mean frequency function. The appropriate analysis in the case without delays uses the Ghosh and Lin (2000) estimator. Three versions of this estimator are calculated using the simulated data. The first is the "true" version that would be calculated if there were no delays, denoted TGL. The second (GL) is the version that would be calculated at the time of analysis, using only reported events. The last (GL6M) is the version that censors all subjects 6 months prior to the time of analysis. The two estimators introduced in this paper, $\tilde{\mu}$ and $\hat{\mu}$, are calculated. All estimators are evaluated at four times: 0.4, 0.8, 1.2, and 1.6 years.

The sample sizes used were 100 and 500. For each scenario and sample size 1000 simulations were performed. All simulations and computations in this section were carried out with the R Language and Environment (R Development Core Team, 2009) and the Condor system (Condor Team, 2007).

Estimated mean squared errors for all estimators are shown in Tables 2 and 3. The naïve estimator (GL) performs very poorly. Although the estimator that ignores the past 6 months shows modest improvement, in every case, the new estimators show superior performance.

When the delays are i.i.d., the performance of $\tilde{\mu}$ seems to be equivalent to that of $\hat{\mu}$ for early times. For later times, especially when the delays are longer, the

performance of $\tilde{\mu}$ is not as good. The source of this trouble is seen in Table 4, which gives bias. A sample size of 500 is sufficiently large to observe the asymptotic bias in these estimators. As t moves out of the valid range, given by

$$\left\{t : \overline{G}(t+d_0) > 0\right\},$$

the asymptotic bias of $\tilde{\mu}$ increases. $\hat{\mu}$, however, is asymptotically unbiased for all times of interest. In fact, the performance of $\hat{\mu}$ comes close to that of the estimator that "knows" event times when they happen, rather than when they are reported.

In scenario V, when the delays are not i.i.d., $\tilde{\mu}$ performs best. The formulas for the variance of $\tilde{\mu}$ and $\hat{\mu}$ were also evaluated in each setting and gave satisfactory results.

4 A Clinical Trial Example: TNT

In this section, data from a real clinical trial will be used to illustrate the methods. The Treating to New Targets (TNT) trial (LaRosa et al., 2005) was a double-blind, randomized clinical trial designed to assess the safety and efficacy of reducing low-density lipoprotein (LDL) cholesterol levels in patients with coronary heart disease. Several cardiovascular events were of interest to the study investigators: death, non-fatal MI, resuscitation after cardiac arrest, hospitalization for congestive heart failure or unstable angina, and stroke. A common practice, especially in cardiology, is to consider the composite of a set of related components as an outcome of interest, in addition to the individual component outcomes. A total of 10,001 patients were randomly assigned to either 10 mg or 80 mg of atorvastatin.

The available trial data include event dates and reporting dates, although some reporting dates were not recorded early in the trial. Since reporting times and delays are crucial in our discussion, we simulated delays for approximately 1000 events for which reporting dates were missing. We did this by randomly picking delays that were recorded, based on event type. This should be a good approximation of what actually happened in the trial and allows successful illustration of the methods. We consider the mean frequency function for the composite outcome process, combining all non-terminal event types.

Figure 3 represents the same hypothetical interim analysis that was shown in Figure 2. The goal is to estimate $\mu(t)$, the expected number of non-fatal events in $(0,t]$. As a standard for comparison, the "true" Ghosh and Lin estimator is shown. This is the estimator that would be used if all events were reported immediately. We are now able to compute this "true" estimator by using information that was unavailable at the analysis time. The estimator that ignores delays is also shown, as is the estimator that cuts the data off 6 months in the past. Finally, the two

Table 2: MSE of Mean Frequency Estimators. Sample Size $n = 100$. TGL is the "true" Ghosh and Lin estimator, GL is the Ghosh and Lin estimator, and GL6M is the Ghosh and Lin estimator, ignoring the last 6 months.

Scenario	t	$\mu(t)$	TGL	GL	GL6M	$\widetilde{\mu}$	$\widehat{\mu}$
I	0.4	2.0	0.1	0.1	0.1	0.1	0.1
	0.8	4.0	0.2	0.6	**0.3**	**0.3**	**0.3**
	1.2	6.0	0.5	1.8	0.9	**0.7**	**0.7**
	1.6	8.0	1.1	4.9	2.3	1.6	**1.5**
II	0.4	2.0	0.1	0.4	**0.1**	**0.1**	**0.1**
	0.8	4.0	0.2	1.8	0.5	**0.3**	**0.3**
	1.2	6.0	0.6	5.5	1.7	1.0	**0.8**
	1.6	8.0	1.2	14	5.9	2.4	**1.8**
III	0.4	2.0	0.1	0.7	0.3	**0.1**	**0.1**
	0.8	4.0	0.2	3.7	1.7	0.5	**0.4**
	1.2	6.0	0.5	11	5.4	1.6	**1.1**
	1.6	8.0	1.1	23	14	5.6	**2.5**
IV	0.4	1.2	0.0	0.1	**0.0**	**0.0**	**0.0**
	0.8	2.4	0.0	0.6	**0.1**	**0.1**	**0.1**
	1.2	3.6	0.1	1.9	0.4	0.2	**0.1**
	1.6	4.8	0.1	4.9	1.9	0.6	**0.3**
V	0.4	2.0	0.1	0.5	0.2	**0.1**	0.2
	0.8	4.0	0.3	2.6	0.9	**0.6**	0.7
	1.2	6.0	0.6	7.5	3.1	**1.1**	2.2
	1.6	8.0	1.2	18	9.5	**2.9**	5.5
VI	0.4	1.8	0.1	0.3	**0.1**	**0.1**	**0.1**
	0.8	3.3	0.2	1.2	0.4	**0.3**	**0.3**
	1.2	4.5	0.4	2.9	1.0	**0.7**	**0.7**
	1.6	5.5	0.8	5.7	2.5	**1.3**	**1.3**

Table 3: MSE of Mean Frequency Estimators. Sample Size $n = 500$. TGL is the "true" Ghosh and Lin estimator, GL is the Ghosh and Lin estimator, and GL6M is the Ghosh and Lin estimator, ignoring the last 6 months.

Scenario	t	$\mu(t)$	TGL	GL	GL6M	$\widetilde{\mu}$	$\widehat{\mu}$
I	0.4	2.0	0.01	0.09	**0.02**	**0.02**	**0.02**
	0.8	4.0	0.0	0.4	**0.1**	**0.1**	**0.1**
	1.2	6.0	0.1	1.4	0.2	**0.1**	**0.1**
	1.6	8.0	0.2	4.2	0.7	**0.3**	**0.3**
II	0.4	2.0	0.01	0.32	0.05	**0.02**	**0.02**
	0.8	4.0	0.0	1.6	0.3	**0.1**	**0.1**
	1.2	6.0	0.1	5.2	1.1	**0.2**	**0.2**
	1.6	8.0	0.2	14	4.8	1.2	**0.4**
III	0.4	2.0	0.01	0.71	0.28	**0.02**	**0.02**
	0.8	4.0	0.1	3.7	1.6	**0.1**	**0.1**
	1.2	6.0	0.1	11	5.2	0.9	**0.2**
	1.6	8.0	0.2	24	14	4.5	**0.5**
IV	0.4	1.2	0.00	0.11	0.02	**0.00**	**0.00**
	0.8	2.4	0.0	0.6	0.1	**0.0**	**0.0**
	1.2	3.6	0.0	1.9	0.4	**0.0**	**0.0**
	1.6	4.8	0.0	4.9	1.7	0.4	**0.1**
V	0.4	2.0	0.01	0.47	0.13	**0.04**	0.05
	0.8	4.0	0.0	2.4	0.8	**0.1**	0.3
	1.2	6.0	0.1	7.3	2.6	**0.2**	0.9
	1.6	8.0	0.2	17	8.4	**1.5**	2.5
VI	0.4	1.8	0.01	0.27	0.05	**0.02**	**0.02**
	0.8	3.3	0.0	1.1	0.2	**0.1**	**0.1**
	1.2	4.5	0.1	2.8	0.6	0.2	**0.1**
	1.6	5.5	0.2	5.6	1.8	0.5	**0.3**

Table 4: Bias of Mean Frequency Estimators. Sample Size $n = 500$. GL is the Ghosh and Lin estimator and GL6M is the Ghosh and Lin estimator, ignoring the last 6 months.

Scenario	t	$\mu(t)$	GL	GL6M	$\widetilde{\mu}$	$\widehat{\mu}$
I	0.4	2.0	–0.3	**0.0**	**0.0**	**0.0**
	0.8	4.0	–0.6	**0.0**	**0.0**	**0.0**
	1.2	6.0	–1.1	**0.0**	**0.0**	**0.0**
	1.6	8.0	–2.0	–0.5	–0.1	**0.0**
II	0.4	2.0	–0.6	–0.2	**0.0**	**0.0**
	0.8	4.0	–1.3	–0.5	**0.0**	**0.0**
	1.2	6.0	–2.3	–1.0	–0.1	**0.0**
	1.6	8.0	–3.7	–2.1	–0.9	**0.0**
III	0.4	2.0	–0.8	–0.5	**0.0**	**0.0**
	0.8	4.0	–1.9	–1.3	–0.2	**0.0**
	1.2	6.0	–3.2	–2.2	–0.8	**0.0**
	1.6	8.0	–4.8	–3.7	–2.0	**0.0**
IV	0.4	1.2	–0.3	–0.1	**0.0**	**0.0**
	0.8	2.4	–0.8	–0.3	**0.0**	**0.0**
	1.2	3.6	–1.4	–0.6	–0.1	**0.0**
	1.6	4.8	–2.2	–1.3	–0.6	**0.0**
V	0.4	2.0	–0.7	–0.3	**0.1**	0.2
	0.8	4.0	–1.5	–0.8	**0.2**	0.4
	1.2	6.0	–2.7	–1.6	**–0.1**	0.8
	1.6	8.0	–4.2	–2.9	**–1.1**	1.4
VI	0.4	1.8	–0.5	–0.2	**0.0**	**0.0**
	0.8	3.3	–1.0	–0.4	**0.0**	**0.0**
	1.2	4.5	–1.7	–0.7	–0.1	**0.0**
	1.6	5.5	–2.4	–1.3	–0.5	**0.0**

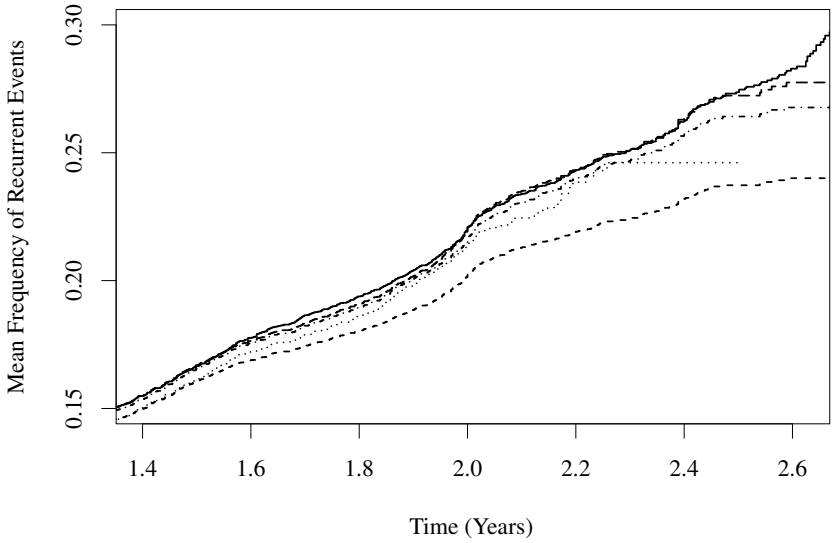

Figure 3: Estimates of mean frequency for an interim analysis at 3 years: the "true" GL estimate (solid), the unadjusted GL estimate (short dashes), the GL estimate that ignores the last 6 months (dotted), $\tilde{\mu}$ (dot-dash), and $\hat{\mu}$ (long dashes).

new estimators, $\tilde{\mu}$ and $\hat{\mu}$, are also displayed. In this example, the estimator that has the greatest error, of course, is the one that makes no adjustment for the delays. Ignoring 6 months gave a better result, but not as good as that of the new estimators.

In this analysis, we have combined all nonfatal event types into a single recurrent-event outcome. While each event type will have a distinct frequency function and reporting delay distribution, $\mu(t)$ retains a valid interpretation as the mean frequency of the composite recurrent outcome. If delay time is dependent on event type, this can be incorporated into the estimation of the delay distribution.

Although the frequency of events in this example is low, the data are fundamentally different from single-event survival data. Prior to the 3-year mark, 442 subjects had experienced multiple events, over 30 had experienced at least 5 events, and one individual had experienced 18 events. This example depicts only one possible monitoring time. We performed the calculations at other times and observed results similar to those shown.

5 Discussion

We have proposed estimators that can be used to estimate the mean frequency function for a recurrent-event process when event reporting is delayed. These include an inverse probability of censoring weighted estimator that is the equivalent of the van der Laan and Hubbard simple estimator in the new setting and an estimator that uses explicit estimation of the delay distribution. Both can be applied to situations with or without a terminating event, such as death.

The properties of the IPCW estimator are based on semiparametric theory. This type of estimator relies heavily on the assumption that $\overline{G}\{V(t)\} > 0$ F_X–almost everywhere or that t is small enough and the delays are bounded. Some of the effects of violating this assumption can be seen in the simulation results. The other estimator, $\hat{\mu}$, does not have this restriction. Instead, the assumption is made that the delays are independent and identically distributed. When dealing with recurrent events, it is quite possible that this assumption does not hold. For example, two or more of a subject's events could be reported at the same time. In these situations, it is possible for $\tilde{\mu}$ to perform better. In most cases, however, slight violations of the independent and identically distributed assumption will probably not have an adverse effect on the estimation of H. The estimator $\hat{\mu}(t)$ has an additional advantage over $\tilde{\mu}(t)$. When reporting dates are missing, as is the case for early data in the TNT example, if we can assume that they are missing at random, the delay distribution can be estimated from the remaining events and the estimate $\hat{\mu}(t)$ computed as usual. On the other hand, $\tilde{\mu}(t)$ cannot be calculated when reporting times are missing.

Appendix

Proof of Theorem 2.1. We will use semiparametric efficiency theory as discussed by van der Laan and Robins (2003). A similar discussion can be found in Tsiatis (2006). Let $P_{F_X,G}$ represent the distribution of Y, the observed data, where F_X is the distribution of the full data, X. The nuisance "parameter" in this case is G, the distribution function of C. The nuisance tangent space for the observed recurrent-event process is

$$S\left(P_{F_X,G}\right) = \overline{\left\{\int U(s)dM_G(s) : U\right\} \cap L_0^2(F_X)},$$

where $dM_G(s) = I(C = s) - d\Lambda_C(s)I(C \geq s)$ and $L_0^2(F_X)$ is the space of functions of X with mean zero and finite variance. The projection of a function $h(Y)$ onto this

space is

$$\prod\left\{h(Y) \mid S\left(P_{F_X,G}\right)\right\} = \int \left[E\left\{h(Y) \mid C = s\right\} - E\left\{h(Y) \mid C > s\right\}\right] dM_G(s).$$

This allows us to find the efficient influence function for estimation of μ.

Assuming that G is known, the influence function of $\tilde{\mu}(t)$ is (2). We have already shown that, for all t such that $\overline{G}(t+d_0) > 0$, $E\{\varphi(t) \mid X\} = N(t) - \mu(t)$. This last quantity is the efficient influence function for estimation of μ when the full data, X, are observed. We calculate the projection of this function onto the nuisance tangent space:

$$\prod\left(\varphi(t) \mid S\left(P_{F_X,G}\right)\right) = \int \left(E\left[\varphi(t) \mid C = s\right] - E\left[\varphi(t) \mid C > s\right]\right) dM_G(s).$$

The efficient influence function in the observed data model is found by subtracting from $\varphi(t)$ its projection onto the nuisance tangent space. This is $\varphi_{\mathtt{eff}}(t)$ in (3).

Consider $\tilde{\mu}(G)$ as a functional on the space of distribution functions. We have that

$$\begin{aligned}\tilde{\mu}(G_n) - \mu &= \tilde{\mu}(G) - \mu + \tilde{\mu}(G_n) - \tilde{\mu}(G) \\ &= \tilde{\mu}(G) - \mu + \Phi(G_n) - \Phi(G) + o_P(n^{-1/2}),\end{aligned}$$

where

$$\Phi(f) \equiv E\left[\int_0^t \frac{dN(s)}{1 - f\{V(s)\}}\right],$$

for any function f satisfying $0 \le f < 1$. The second equality is due to the uniform convergence of G_n, the empirical distribution function. We recall also that G_n is an efficient estimator of G, so that $\Phi(G_n)$ is an efficient estimator of $\Phi(G)$. Then, by Theorem 2.3 in van der Laan and Robins (2003), $\tilde{\mu}(G_n)$, is a regular, asymptotically linear estimator with influence function given by $\varphi_{\mathtt{eff}}(t)$. This completes the proof.

Proof of Theorem 2.2. Although $M_i(t)$ $i = 1,\ldots,n$, defined in Section 2.3 has the form of a martingale, in general it is not, due to likely within-subject dependence of events. Let N^*_{ij}, $j = 1,\ldots,N_{\max}$, be the counting process of the jth observed event for subject i. Now, define N^{**}_{il}, $l = 1,\ldots,N_{\max}$, such that

$$\left\{N^{**}_{i1}, N^{**}_{i2}, \ldots, N^{**}_{i,N_{\max}}\right\}$$

is a random permutation of

$$\{N^*_{i1}, N^*_{i2}, \ldots, N^*_{i,N^*_i(t)}, \underbrace{0,0,\ldots,0}_{N_{\max} - N^*_i(t)}\}.$$

So we have, marginally, that

$$N_{il}^{**}(t) = \begin{cases} N_{i1}^{*}(t), & \text{with prob } 1/N_{\max} \\ \vdots & \vdots \\ N_{i,N_i^*(t)}^{*}(t), & \text{with prob } 1/N_{\max} \\ 0, & \text{with prob } 1 - N_i^*(t)/N_{\max} \end{cases}.$$

We have now defined a collection of counting processes. These are not mutually independent, but they do have identical marginal distributions with expected value

$$E\{N_{il}^{**}(t)\} = \frac{1}{N_{\max}} \int_0^t E\{H(s \mid C_i)\} d\mu(s).$$

It follows that each

$$M_{il}(t) = N_{il}^{**}(t) - \frac{1}{N_{\max}} \int_0^t H(s \mid C_i) d\mu(s),$$

$l = 1, \ldots, N_{\max}$, is a martingale with respect to its history, $\mathscr{F}_{t,il} = (C, \{N_{il}^{**}(s), s \leq t\})$. Note that

$$\sum_{l=1}^{N_{\max}} N_{il}^{**}(t) = \sum_{j=1}^{N_i(t)} N_{ij}^{*}(t) = N_i(t).$$

Thus, $M_i(t) = \sum_{l=1}^{N_{\max}} M_{il}(t)$ is the sum of dependent martingales that all have the same compensator, $N_{\max}^{-1} \int_0^t H(s \mid C_i) d\mu(s)$.

Defined in this way, the data fall into a category of clustered survival data similar to that discussed by Gangnon and Kosorok (2004). We note that

$$\sup_t \left| n^{-1} \sum \widehat{H}(t \mid C_i) - E\{H(t \mid C_i)\} \right| \to 0, \tag{4}$$

almost surely, by standard probability results and the boundedness of H. Over the interval $(0, \tau)$, $nU(t)$ has bounded variation and

$$\sup_{t \in K} \left| n\{\widehat{U}(t) - U(t)\} \right| \ \to \ 0,$$

in probability, for any closed subinterval K of $(0, \tau)$. Let

$$M_U(t) = \int_0^t U(s) \sum_{i=1}^{n} dM_i(s)$$

and note that $\text{var}\{M_U(t)\} = \sigma_{R,n}^2(t)$. We also have

$$M_{\widehat{U}}(t) - M_U(t) = \int_0^t n\{\widehat{U}(s) - U(s)\} \frac{1}{n} \sum_{i=1}^{n} dM_i(s). \tag{5}$$

Arguments similar to those given by Gangnon and Kosorok (2004) can be used to show that

$$n^{1/2} M_U(t) \;\rightarrow\; N\{0, \sigma_R^2(t)\},$$

and by (4) and (5) the same result holds when $M_U(t)$ is replaced by $M_{\widehat{U}}(t)$. This gives the desired result.

References

Condor Team (2007): Condor, Computer Sciences Department, University of Wisconsin–Madison, Madison, Wisconsin, USA, URL `http://www.cs.wisc.edu/condor`.

Cook, R. J. and J. F. Lawless (1997): "Marginal analysis of recurrent events and a terminating event," Statist. Med., 16, 911–924.

Gangnon, R. and M. Kosorok (2004): "Sample-size formula for clustered survival data using weighted log-rank statistics," Biometrika, 91, 263–275.

Ghosh, D. and D. Lin (2000): "Nonparametric analysis of recurrent events and death," Biometrics, 56, 554–562.

Hu, P. H. and A. A. Tsiatis (1996): "Estimating the survival function when ascertainment of vital status is subject to delay," Biometrika, 83, 371–380.

Hubbard, A. E., M. J. van der Laan, W. Enanoria, and J. M. Colford Jr. (2000): "Nonparametric survival estimation when death is reported with delay," Lifetime Data Anal, 6, 237–250.

LaRosa, J., S. Grundy, D. Waters, C. Shear, P. Barter, J. Fruchart, A. Gotto, H. Greten, J. Kastelein, J. Shepherd, and Wenger, N. for the Treating to New Targets (TNT) Investigators (2005): "Intensive lipid lowering with atorvastatin in patients with stable coronary disease," New Engl. J. Med., 352, 1425–1435.

Li, Q. and S. Lagakos (1997): "Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event," Statist. Med., 16, 925–940.

R Development Core Team (2009): R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Tsiatis, A. (2006): Semiparametric Theory and Missing Data, Springer, New York.

van der Laan, M. J. and A. E. Hubbard (1998): "Locally efficient estimation of the survival distribution with right-censored data and covariates when collection of data is delayed," Biometrika, 85, 771–783.

van der Laan, M. J. and J. M. Robins (2003): Unified Methods for Censored Longitudinal Data and Causality, Springer-Verlag, New York.