ESTIMATION OF MENARCHEAL AGE DISTRIBUTION
FROM IMPERFECTLY RECALLED DATA

Sedigheh Mirzaei Salehabadi¹, Debasis Sengupta² and Rahul Ghosal³

¹ National Institute of Child Health and Human Development, Bethesda, USA
² Indian Statistical Institute, Kolkata, India
³ North Carolina State University, Raleigh, USA

Abstract: In a cross-sectional study, pubertal females were asked to recall the time of menarche, if experienced. Some respondents recalled the date exactly, some recalled only the month or the year of the event, and some were unable to recall anything. We consider estimation of the menarcheal age distribution from this interval censored data. A complicated interplay between age-at-event and calendar time, together with the evident fact of memory fading with time, makes the censoring informative. We propose a model where the probabilities of various types of recall would depend on the time since menarche, through a multinomial regression function. Establishing consistency and asymptotic normality of the parametric MLE requires a bit of tweaking of the standard asymptotic theory, as the data format varies from case to case. We also provide a non-parametric MLE, propose a computationally simpler approximation, and establish the consistency of both of these estimators under mild conditions. We study the small sample performance of the parametric and non-parametric estimators through Monte Carlo simulations. Moreover, we provide a graphical check of the assumption of
the multinomial model for the recall probabilities, which appears to hold for the menarcheal data set. Its analysis shows that the use of the partially recalled part of the data indeed leads to smaller confidence intervals of the survival function.

Key words and phrases: Interval censoring, Informative censoring, Maximum likelihood estimator, Retrospective study, Current status data, Self consistency.

1. Introduction

In a recent survey conducted by the Indian Statistical Institute (ISI) in and around the city of Kolkata (Dasgupta, 2015), over four thousand randomly selected individuals, aged between 7 and 21 years, were sampled. In this retrospective and cross-sectional study, the subjects were interviewed on or around their birthdays. The data set on female subjects contains age, menarcheal status, some physical measurements and information on some socioeconomic variables. If a subject had already experienced menarche, she was asked to recall the time of the onset of her menarche. Among the 2195 females represented in the data set, 775 individuals did not have menarche, 443 individuals recalled the exact date of the onset of menarche, 276 and 209 individuals recalled the calendar month and the calendar year of the onset, respectively, and 492 individuals could not recall any range of dates. Thus, the data are interval-censored. A major goal of this study is to estimate the distribution of the age at onset of menarche.
This problem should be of interest to anyone working with incompletely recalled time-to-event data, of which there are many examples in the literature. The key variables in these studies include age at onset of menarche in adolescent and young adult females (Koo and Rohana, 1997), time-to-pregnancy (Joffe et al., 1995), time-to-weaning from breastfeeding (Gillespie et al., 2006), time-to-injury for victims injured during a year (Harel et al., 1994), time-to-employment (Mathiowetz and Ouncanb, 1988), and so on. In all these studies, estimation of the time-to-event distribution is important for building a standard for individuals, comparing two populations or assessing the effect of a covariate. There is a possibility that the recalled time-to-event is inaccurate (Koo and Rohana, 1997; Mathiowetz and Ouncanb, 1988). In the ISI study, there was an attempt to circumvent this problem by allowing the respondents to report an interval in lieu of the exact age-at-menarche. The recalled intervals generally happened to be in terms of calendar months and years, rather than age intervals. We refer to this special type of incompleteness as partial recall.

Figure 1 shows the cumulative proportions of successively less precise recall in different groups of ages at interview, for the respondents of the ISI study. It is seen that the lines do not cross and the age group order is preserved. Also, there is greater precision of recall at lower age group, i.e., memory fades with time. Thus, two subjects interviewed at the same age
Figure 1: Cumulative proportion of decreasing degrees of recall for different age ranges in menarcheal data

would have different chances of recalling their age at menarche, depending on which of them had experienced the event earlier. In other words, the censoring mechanism underlying such recall-based data is inherently informative. The natural question is: how can one model the different degrees of partial recall, so that the distribution of menarcheal age can be estimated?

There is no suitable model and method for estimating the time-to-event distribution from partially recalled data, though such data abound in various fields. Apart from the informative nature of censoring, the problem is complicated by the mismatch of the time scales of the partial recall information (expressed through calendar time) and the time-to-event (measured from a respondent-specific starting time, e.g., birth). Mirzai et al. (2015)
and Mirzaei and Sengupta (2016) addressed the first issue by proposing a model for this type of informative censoring, but they bypassed the second issue by clubbing all the cases of partial recall with the cases of no recall.

In this paper, we propose a realistic censoring model for estimating the time-to-event distribution. After presenting this model in Section 2, we derive the appropriate likelihood under the proposed model. In Section 3, we express the likelihood as a product of densities in an appropriate space, and derive conditions that would ensure consistency and asymptotic normality of a parametric maximum likelihood estimator (MLE). In Section 4, we derive the non-parametric maximum likelihood estimator (NPMLE) and an approximate MLE (AMLE) and establish consistency of both the estimators. In Sections 5 and 6, we report the results of Monte Carlo simulations of small sample performance of the MLE and the AMLE, and present some diagnostic checks of adequacy of the model. We analyze the real data set in Section 7. The proofs of the theorems are given in supplementary material.

2. Model and Likelihood

Consider a set of $n$ subjects having ages at occurrence of landmark events $T_1, \ldots, T_n$, which are samples from the distribution $F$, with density $f$. Let these subjects be interviewed at ages $S_1, \ldots, S_n$, respectively. Suppose the $S_i$s are samples from another distribution and are independent of the $T_i$s.
Let $\delta_i$ be the indicator of $T_i \leq S_i$. This inequality means that the event for the $i$th subject had occurred on or before the time of interview.

In the case of current status data, one only observes $(S_i, \delta_i), i = 1, 2, \ldots, n$. The corresponding likelihood, conditional on the times of interview, is

$$\prod_{i=1}^{n} [F(S_i)]^{\delta_i}[\bar{F}(S_i)]^{1-\delta_i}, \quad (2.1)$$

where $\bar{F}(S_i) = 1 - F(S_i)$. For properties of the MLE based on the above likelihood, see [Lee and Wang (2003)].

The structure of recalled data is generally more complicated. [Mirzaei et al. (2013)] proposed a simplistic model, where the subject may either recall the time of the event exactly or not remember it at all. They used an indicator, $\varepsilon_i$, to record whether an exact recall is possible. As the chance of recall may depend on the time elapsed since the event, they modeled the non-recall probability as a function of this time. According to this model,

$$P(\varepsilon_i = 0 | S_i = s, T_i = t) = \pi(s - t) \quad \text{for} \quad 0 < t < s,$$

for some non-recall probability function $\pi$. Thus, the likelihood is

$$\prod_{i=1}^{n} \left[ \left( \int_{0}^{S_i} f(u)\pi(S_i - u)du \right)^{1-\varepsilon_i} [f(T_i)(1 - \pi(S_i - T_i))]^{\varepsilon_i} \right]^{\delta_i} [F(S_i)]^{1-\delta_i}. \quad (2.2)$$

Let us now consider the possibility that the $i$th subject can recall the date of the event only up to a calendar month or a calendar year, and define
the recall status variable $\varepsilon_i$ for the $i$th subject as

$$
\varepsilon_i = \begin{cases} 
0 & \text{if the exact date is recalled}, \\
1 & \text{if the date is recalled up to the calendar month}, \\
2 & \text{if the date is recalled up to the calendar year}, \\
3 & \text{if the event has not happened or the date is not recalled}.
\end{cases}
$$

(2.3)

We regard the four scenarios outcomes of a multinomial selection, where allocation probabilities depend on the time elapsed since the occurrence of the event. Thus, for $0 < t < s$, we model the allocation probabilities as

$$
P(\varepsilon_i = 0|S_i = s, T_i = t) = \pi^{(0)}(s - t),
$$

$$
P(\varepsilon_i = 1|S_i = s, T_i = t) = \pi^{(1)}(s - t),
$$

$$
P(\varepsilon_i = 2|S_i = s, T_i = t) = \pi^{(2)}(s - t),
$$

$$
P(\varepsilon_i = 3|S_i = s, T_i = t) = \pi^{(3)}(s - t).
$$

(2.4)

where $\sum_{k=0}^{3} \pi^{(k)}(s - t) = 1$.

We refer to the set-up described in the first paragraph of this section, together with (2.3) and (2.4) as the proposed model. According to this model, contributions to the likelihood in different cases are as follows.

**Case (i)** When $\delta_i = 0$ (the event has not occurred till the time of observation), the contribution of the $i$th individual to the likelihood is $\tilde{F}(S_i)$.

**Case (ii):** When $\delta_i = 1$ and $\varepsilon_i = 0$ (the event has occurred and the $i$th
individual can remember the time), the contribution of the individual to the likelihood is
\[ f(T_i)\pi^{(0)}(S_i - T_i). \]

**Case (iii):** When \( \delta_i = 1 \) and \( \varepsilon_i = 1 \) (the event has occurred but the \( i \)th individual can only recall the calendar month of the event), the contribution of the individual to the likelihood is
\[ \int_{M_{i1}}^{M_{i2}} f(u)\pi^{(1)}(S_i - u)du, \]
where \( M_{i1} \) and \( M_{i2} \) are the ages of the individual at the beginning and the end of the calendar month recalled by the individual.

**Case (iv):** When \( \delta_i = 1 \) and \( \varepsilon_i = 2 \) (the event has occurred but the \( i \)th individual can only recall the calendar year of the event), the contribution of the individual to the likelihood is
\[ \int_{Y_{i1}}^{Y_{i2}} f(u)\pi^{(2)}(S_i - u)du, \]
where \( Y_{i1} \) and \( Y_{i2} \) are the ages of the individual at the beginning and the end of the calendar year recalled by the individual.

**Case (v):** When \( \delta_i = 1 \) and \( \varepsilon_i = 3 \) (the event has occurred but the \( i \)th individual cannot recall the time at all), the contribution of the individual to the likelihood is
\[ \int_{0}^{S_i} f(u)\pi^{(3)}(S_i - u)du. \]

Therefore, the overall likelihood is
\[
\prod_{i=1}^{n} \left[ F(S_i) \right]^{1-\delta_i} \left[ (f(T_i)\pi^{(0)}(S_i - T_i))^{I(\varepsilon_i=0)} \left( \int_{M_{i1}}^{M_{i2}} f(u)\pi^{(1)}(S_i - u)du \right)^{I(\varepsilon_i=1)} \times \left( \int_{Y_{i1}}^{Y_{i2}} f(u)\pi^{(2)}(S_i - u)du \right)^{I(\varepsilon_i=2)} \left( \int_{0}^{S_i} f(u)\pi^{(3)}(S_i - u)du \right)^{I(\varepsilon_i=3)} \right]^{\delta_i}.
\]

(2.5)
Note that when \( \pi^{(1)} = \pi^{(2)} = 0 \), the likelihood (2.5) reduces to (2.2). When \( \pi^{(1)} = \pi^{(2)} = 0 \) and \( \pi^{(0)} \) is a constant, it becomes a constant multiple of the likelihood corresponding to non-informatively interval censored data. If \( \pi^{(0)} = \pi^{(1)} = \pi^{(2)} = 0 \), it reduces to the current status likelihood (2.1).

While the proposed model is specific to the data at hand, it can easily be adjusted for arbitrary types of recall, which need not even be ordered.

The factors in the product likelihood (2.5) have different forms in different cases. We now show that they can be expressed as the common density of some random vector with respect to a suitable dominating measure.

The main challenge to obtaining a common format of the data lies in the fact that \( M_{i1}, M_{i2}, Y_{i1} \) and \( Y_{i2} \) are the ages of the \( i \)th individual at specified calendar times. We make use of the fact that these observables are functions of \( T_i \) and the date of birth of the \( i \)th individual. Specifically, for the \( i \)th subject, let \( m_i \) be the serial number of the month of birth within the year of birth and \( d_i \) be the time (measured in years) from the beginning of the month of birth till the event of birth. For the sake of simplicity, we assume that every year has duration 1 and every month has duration 1/12.

When \( \epsilon_i = 1 \), i.e., the month of the event is recalled, we write

\[
M_{i1} = \left\lfloor 12(d_i + T_i) \right\rfloor / 12 - d_i, \\
M_{i2} = M_{i1} + 1/12,
\]  

(2.6)
where $\lfloor \cdot \rfloor$ is the integer part of its argument. Thus, the variables $\lfloor 12(d_i + T_i) \rfloor$, $m_i$ and $d_i$ can be obtained from $M_{i1}$, $M_{i2}$, $m_i$ and $d_i$ and vice versa.

Likewise, when $\epsilon_i = 2$, i.e., the year of the event is recalled, we write

$$
Y_{i1} = \lfloor (T_i + d_i + (m_i - 1)/12) \rfloor - (d_i + (m_i - 1)/12),
$$

$$
Y_{i2} = Y_{i1} + 1.
$$

(2.7)

It is clear that the variables $\lfloor (T_i + d_i + (m_i - 1)/12) \rfloor$, $m_i$ and $d_i$ are equivalent to $Y_{i1}$, $Y_{i2}$, $m_i$ and $d_i$. Therefore, we define the variable

$$
V_i = \begin{cases} 
T_i & \text{if } \epsilon_i = 0, \delta_i = 1, \\
\lfloor 12(d_i + T_i) \rfloor / 12 & \text{if } \epsilon_i = 1, \delta_i = 1, \\
\lfloor (T_i + d_i + (m_i - 1)/12) \rfloor & \text{if } \epsilon_i = 2, \delta_i = 1, \\
0 & \text{if } \epsilon_i = 3, \delta_i = 1, \text{ or if } \delta_i = 0,
\end{cases}
$$

(2.8)

which captures the essential part of the occasionally observable variables $T_i$, $M_{i1}$, $M_{i2}$, $Y_{i1}$ and $Y_{i2}$, and work with the observable vector

$$
Y_i = (S_i, V_i, \epsilon_i, \delta_i, m_i, d_i).
$$

(2.9)

We have already assumed that the $T_i$s (time-to-event) are samples from the distribution $F$ and the $S_i$s (ages on interview date) are samples from another distribution. We now denote by $G_1$, $G_2$ and $G_3$ the distributions of $S_i$, $m_i$ and $d_i$, respectively, for every $i$. The distribution $G_2$ is defined over the set $\{1, 2, \ldots, 12\}$, and $G_3$ is defined over the interval $[0, 1/12]$. The latter
assumption disregards the fact that $d_i$ is known only up to days (measured as fixed fractions of a year), to keep the description simple.

Theorem 1 presented below gives the density of $Y_i$, after the subscript $i$ is dropped for simplicity. The dominating probability measure used for defining this density is $\mu = \vartheta_1 \times \vartheta_2 \times \vartheta_3 \times \vartheta_4 \times \vartheta_5 \times \vartheta_6$ where $\vartheta_1$ is the measure with respect to which $G_1$ has a density (e.g., the counting or the Lebesgue measure, depending on whether $G_1$ is discrete or continuous), $\vartheta_2$ is the sum of the counting and the Lebesgue measures, each of $\vartheta_3$, $\vartheta_4$ and $\vartheta_5$ is the counting measure and $\vartheta_6$ is the Lebesgue measure Ash (2000).

**Theorem 1.** The density of $Y = (S, V, \varepsilon, \delta, m, d)$ with respect to the measure $\mu$ is

$$h(s, v, \varepsilon, \delta, m, d) = \begin{cases} 
g_1(s)g_2(m)g_3(d)\bar{F}(s) & \text{if } \delta = 0, 
g_1(s)g_2(m)g_3(d)f(v)\pi^{(0)}(s-v)I_{v<s} & \text{if } \varepsilon = 0 \text{ and } \delta = 1, 
g_1(s)g_2(m)g_3(d)\int_{v-d}^{v+1-d-m-1} f(u)\pi^{(1)}(s-u)du & \text{if } \varepsilon = 1 \text{ and } \delta = 1, 
g_1(s)g_2(m)g_3(d)\int_{v-d}^{v+1-d-m-1} f(u)\pi^{(2)}(s-u)du & \text{if } \varepsilon = 2 \text{ and } \delta = 1, 
g_1(s)g_2(m)g_3(d)\int_{0}^{s} f(u)\pi^{(3)}(s-u)du & \text{if } \varepsilon = 3 \text{ and } \delta = 1, 
\end{cases}$$

(2.10)

where $g_1$, $g_2$ and $g_3$ are the densities of $G_1$, $G_2$ and $G_3$ with respect to the measures $\vartheta_1$, $\vartheta_5$ and $\vartheta_6$, respectively.
Theorem 1 implies that the likelihood (2.5) can be written as

\[ \prod_{i=1}^{n} [\bar{F}(S_i)]^{1-\delta_i} \left[ (f(V_i)\pi^{(0)}(S_i-V_i))^{I(\varepsilon_i=0)} \left( \int_{V_i-d_i}^{V_i-d_i+m_i-\frac{1}{12}} f(u)\pi^{(1)}(S_i-u)du \right)^{I(\varepsilon_i=1)} \times \left( \int_{V_i-d_i}^{V_i-d_i+m_i-\frac{1}{12}+1} f(u)\pi^{(2)}(S_i-u)du \right)^{I(\varepsilon_i=2)} \left( \int_{0}^{S_i} f(u)\pi^{(3)}(S_i-u)du \right)^{I(\varepsilon_i=3)} \right]^{\delta_i}, \]

\[ = \prod_{i=1}^{n} h(S_i, V_i, \varepsilon_i, \delta_i, m_i, d_i) \prod_{i=1}^{n} g_1(S_i) g_2(m_i) g_3(d_i), \] (2.11)

where the \( i \)th factor is the conditional density of \((V_i, \varepsilon_i, \delta_i)\) given \((S_i, m_i, d_i)\).

3. Parametric estimation

Suppose the forms of the functions \( \bar{F}, f, \pi^{(0)}, \pi^{(1)}, \pi^{(2)} \) and \( \pi^{(3)} \) in the likelihood (2.5) are known up to a few parameters, and accordingly they are written as \( \bar{F}_\theta, f_\theta, \pi^{(0)}_\eta, \pi^{(1)}_\eta, \pi^{(2)}_\eta \) and \( \pi^{(3)}_\eta \), respectively. The MLE of the (possibly vector) parameters \( \theta \) and \( \eta \) are obtained by maximizing (2.5).

Since the equivalent likelihood (2.11) is identified as a product of conditional densities, standard results for consistency and asymptotic normality of the MLE become applicable. The regularity conditions for these results would then be specified in terms of the density of \( Y_i \). In the first section of the Supplementary material, we provide easily verifiable sufficient conditions that involve the density \( f_\theta \) (the density of \( T_i \)) and the functions \( \pi^{(0)}_\eta, \pi^{(1)}_\eta, \pi^{(2)}_\eta \) and \( \pi^{(3)}_\eta \), which define the conditional probability distribution of the random variable \( \varepsilon_i \) given \( T_i \) and \( S_i \).
4. Non-parametric estimation

4.1 Non-parametric MLE

Before embarking on the task of estimation, we need to visit the issue of identifiability.

**Theorem 2.** The distribution functions $G_1$ and $F$, and recall probabilities $\pi^{(k)}$, $k = 0, 1, 2, 3$ are identifiable from $h$ in (2.10).

The likelihood (2.5) is difficult to maximize because of the integrals contained in the expression. In order to simplify it, we assume that the function $\pi^{(k)}$ in (2.5) is piecewise constant, having the form

\[
\begin{align*}
\pi^{(0)}(x) &= b_{01}I(x_1 < x \leq x_2) + b_{02}I(x_2 < x \leq x_3) + \ldots + b_{0L}I(x_L < x < \infty), \\
\pi^{(1)}(x) &= b_{11}I(x_1 < x \leq x_2) + b_{12}I(x_2 < x \leq x_3) + \ldots + b_{1L}I(x_L < x < \infty), \\
\pi^{(2)}(x) &= b_{21}I(x_1 < x \leq x_2) + b_{22}I(x_2 < x \leq x_3) + \ldots + b_{2L}I(x_L < x < \infty), \\
\pi^{(3)}(x) &= b_{31}I(x_1 < x \leq x_2) + b_{32}I(x_2 < x \leq x_3) + \ldots + b_{3L}I(x_L < x < \infty). 
\end{align*}
\]

where $0 = x_1 < x_2 < \ldots < x_L$ are a chosen set of time-points and $b_{l1}, b_{l2}, \ldots, b_{lL}$, $l = 0, 1, 2, 3$ are unspecified parameters taking values in the range $[0, 1]$ such that $\sum_{l=0}^{3} b_{lj} = 1$ for $j = 1, 2, \ldots, L$. Then the likelihood
4.1 Non-parametric MLE

(4.13) reduces to

\[
L = \prod_{i=1}^{n} [F(S_i)]^{1 - \delta_i} \left[ f(T_i) \left( \sum_{l=1}^{L} b_{0l} I(W_{l+1}(S_i) < T_i \leq W_l(S_i)) \right) \right]^{I(\epsilon_i = 0)} \\
\times \left\{ \sum_{l=1}^{L} b_{1l} \left( F(\min(W_l(S_i), M_{i2})) - F(\max(W_{l+1}(S_i), M_{i1})) \right) \right\}^{I(\epsilon_i = 1)} \\
\times \left\{ \sum_{l=1}^{L} b_{2l} \left( F(\min(W_l(S_i), Y_{i2})) - F(\max(W_{l+1}(S_i), Y_{i1})) \right) \right\}^{I(\epsilon_i = 2)} \\
\times \left\{ \sum_{l=1}^{L} b_{3l} \left( F(W_l(S_i)) - F(W_{l+1}(S_i)) \right) \right\}^{I(\epsilon_i = 3)} \delta_i, \tag{4.13}
\]

where \( W_l(S_i) = (S_i - x_l) \vee t_{\text{min}} \) for \( l = 1, \ldots, L \) and \( W_{L+1}(S_i) = t_{\text{min}}, i = 1, 2, \ldots, n \). The likelihood (4.13) involves probabilities assigned to intervals of the type \([t, t_{\text{max}}]\) or \((t, t_{\text{max}}]\), as per the baseline probability distribution.

Since these intervals have overlap, we try to write them as unions of some disjoint intervals. Let \( \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5 \) be sets of indices \( i \) (between 1 and \( n \)) that satisfy the conditions \( \delta_i = 0, \delta_i \epsilon_i = 1, \delta_i (1 - \epsilon_i) = 1, \delta_i \epsilon_i = 2 \) and \( \delta_i \epsilon_i = 3 \), respectively. Consider the intervals

\[
A_i = (S_i, t_{\text{max}}) \quad \text{for } i \in \mathcal{I}_1; \\
A_i' = [T_i, t_{\text{max}}) \quad \text{for } i \in \mathcal{I}_2; \\
A_i'' = (T_i, t_{\text{max}}] \quad \text{for } i \in \mathcal{I}_2; \\
A_{il} = \begin{cases} 
(W_l(S_i), t_{\text{max}}], & l = 1, \ldots, L, \\
[W_l(S_i), t_{\text{max}}], & l = k + 1,
\end{cases} \quad \text{for } i \in \mathcal{I}_2 \cup \mathcal{I}_3;
\]

and
4.1 Non-parametric MLE

\[ B_{il} = [W_{l+1}(S_i) \lor M_{i1}, W_{l+1}(S_i) \land M_{i1}] \quad \text{for } i \in \mathcal{I}_4 \& l = 1, \ldots, L; \]

\[ C_{il} = [W_{l+1}(S_i) \lor Y_{i1}, W_{l+1}(S_i) \land Y_{i1}] \quad \text{for } i \in \mathcal{I}_5 \& l = 1, \ldots, L. \]

(4.14)

and the sets

\[ A_1 = \{ A_i : i \in \mathcal{I}_1 \}; \]

\[ A_2 = \{ A_i \setminus A'_i : i \in \mathcal{I}_2 \}; \]

\[ A_3 = \{ A'_i : i \in \mathcal{I}_2 \}; \]

\[ A_4 = \{ A_{i(l+1)} \setminus A_{il} : 1 \leq l \leq L \text{ and } i \in \mathcal{I}_3 \}; \]

\[ A_5 = \{ B_{il} : 1 \leq l \leq L \text{ and } i \in \mathcal{I}_4 \}; \]

\[ A_6 = \{ C_{il} : 1 \leq l \leq L \text{ and } i \in \mathcal{I}_5 \}. \]

(4.15)

As \( F \) is absolutely continuous, the elements of \( A_2 \) and \( A_3 \) are distinct with probability 1. Let \( n_i \) be the cardinality of \( \mathcal{I}_i \), \( i = 1, 2, 3, 4, 5 \). We arrange the singleton elements of \( A_2 \) in increasing order, and denote them as \( B_1, B_2, \ldots, B_{n_2} \). We also arrange the elements of \( A_3 \) in the corresponding order and denote them as \( B_{n_2+1}, B_{n_2+2}, \ldots, B_{2n_2} \). We then collect the unique elements of \( A_1 \cup A_4 \cup A_5 \cup A_6 \) that are distinct from \( B_1, B_2, \ldots, B_{2n_2} \), and denote them as \( B_{2n_2+1}, B_{2n_2+2}, \ldots, B_M \). Observe that the collection \( B_1, B_2, \ldots, B_M \) consists of the distinct elements of \( \bigcup_{i=1}^{6} A_i \), arranged in a particular order. Denote the non-empty subsets of the index set \( \{1, 2, \ldots, M\} \) by \( s_1, s_2, \ldots, s_{2^M - 1} \). Define

\[ I_r = \left\{ \bigcap_{i \in s_r} B_i \right\} \cap \left\{ \bigcap_{i \notin s_r} B^c_i \right\} \quad \text{for } r = 1, 2, \ldots, 2^M - 1. \]

(4.16)
Some of the $I_r$s may be empty sets, denoted here by $\phi$. Let

$$C = \{s_r : I_r \neq \phi, 1 \leq r \leq 2^M - 1\}; \quad (4.17)$$

$$A = \{I_r : I_r \neq \phi, 1 \leq r \leq 2^M - 1\}. \quad (4.18)$$

It can be verified that the elements of $A$ are distinct and disjoint.

Note that each of the intervals $B_1, \ldots, B_M$ is a union of disjoint sets that are members of $A$. For any Borel set $A$, suppose $P(A)$ is the probability assigned to $A$ as per the probability distribution $F$. Let $p_r = P(I_r)$, for $I_r \in A$. Then the likelihood (4.13) reduces to

$$L = \prod_{i \in I_1} \left( \sum_{r : I_r \subseteq A_i \atop s_r \in C} p_r \right) \times \prod_{i \in I_2} \left( 1 - \sum_{l=1}^L (b_{1l} + b_{2l} + b_{3l}) I_{(T_l \in A_{i(1+1)} \setminus A_{il})} \right) \cdot \left( \sum_{r : I_r \subseteq A_i \setminus A_{il}} p_r \right)$$

$$\times \prod_{i \in I_3} \left[ \sum_{l=1}^L b_{3l} \left( \sum_{r : I_r \subseteq A_{i(l+1)} \setminus A_{il}} \atop s_r \in C \right) p_r \right] \times \prod_{i \in I_4} \left[ \sum_{l=1}^L b_{1l} \left( \sum_{r : I_r \subseteq A_i \setminus A_{il}} \atop s_r \in C \right) p_r \right]$$

$$\times \prod_{i \in I_5} \left[ \sum_{l=1}^L b_{2l} \left( \sum_{r : I_r \subseteq A_i \setminus A_{il}} \atop s_r \in C \right) p_r \right]. \quad (4.19)$$

Thus, maximizing the likelihood (4.13) amounts to maximizing (4.19) with respect to $p_r$ for $s_r \in C$.

There is a partial order among the members of $C$ in the sense that some
sets are contained in others. Consider the following subsets of $\mathcal{C}$.

$$C_1 = \{ s : s \in \mathcal{C}; \text{ there is another element } s' \in \mathcal{C}, \text{ such that } s \subset s' \},$$

$$C_2 = \{ s : s \in \mathcal{C}; \text{ there is another element } s' \in \mathcal{C}, \text{ such that } s' \setminus (s \cap s') \text{ consists of a singleton } j \text{ and } s \setminus (s \cap s') = \{ j + n_2 \} \},$$

$$C_0 = \mathcal{C} \setminus (C_1 \cup C_2). \quad (4.20)$$

We now show that maximization of the likelihood can be restricted to $C_0$.

**Theorem 3.** Maximizing the likelihood (4.19) with respect to $p_r$ for $s_r \in \mathcal{C}$ is equivalent to maximizing it with respect to $p_r$ for $s_r \in C_0$, i.e.,

$$\max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}} p_r = 1} L = \max_{p_r : p_r \in [0,1], \sum_{s_r \in C_0} p_r = 1} L.$$

It transpires from the above theorem that the likelihood has the same maximum value, irrespective of whether $s_r$ is chosen from the class $\mathcal{C}$ or $C_0$. Therefore, we can replace $\mathcal{C}$ by $C_0$ in (4.19).

Let us relabel the intervals $I_j$, $s_j \in C_0$, by $J_1, J_2, \ldots, J_\nu$. Further, let $\mathcal{A}_0 = \{ J_1, J_2, \ldots, J_\nu \}$ and $q_j = P(J_j)$ for $j = 1, 2, \ldots, \nu$. If the likelihood (4.19) is rewritten with the condition $s_r \in \mathcal{C}$ replaced by the equivalent condition $I_r \in \mathcal{A}$, then Theorem 3 shows that the latter condition can be replaced by $I_r \in \mathcal{A}_0$. In other words, maximizing the likelihood (4.19) is
equivalent to maximizing

\[
L(p, \eta) = \prod_{i \in \mathcal{I}_1} \left( \sum_{j: J_{ij} \subseteq A_i} q_j \right) \times \prod_{i \in \mathcal{I}_2} \left( 1 - \sum_{l=1}^{L} (b_{1l} + b_{2l} + b_{3l}) I_{(T_l \in A_{(l+1)} \setminus A_d)} \right) \times \left( \sum_{j: J_{ij} \subseteq A_i \setminus A'_i} q_j \right) \\
\times \prod_{i \in \mathcal{I}_3} \left[ \sum_{l=1}^{L} b_{2l} \left( \sum_{j: J_{ij} \subseteq C_{il}} q_j \right) \right] \times \prod_{i \in \mathcal{I}_4} \left[ \sum_{l=1}^{L} b_{3l} \left( \sum_{j: J_{ij} \subseteq B_{il}} q_j \right) \right] \times \prod_{i \in \mathcal{I}_5} \left[ \sum_{l=1}^{L} b_{2l} \left( \sum_{j: J_{ij} \subseteq C_{il}} q_j \right) \right] = \prod_{i=1}^{n} \left( \sum_{j=1}^{\nu} \alpha_{ij} q_j \right),
\]

with respect to the vector parameters \( p = (q_1, q_2, \ldots, q_\nu)^T \) and \( \eta = (b_{11}, \ldots, b_{1L}, b_{21}, \ldots, b_{2L}, b_{31}, \ldots, b_{3L})^T \), subject to the restrictions \( \sum_{j=1}^{\nu} q_j = 1, 0 \leq q_1, \ldots, q_\nu \leq 1 \), where

\[
\alpha_{ij} = \begin{cases} 
I_{(J_{ij} \subseteq A_i)} & \text{if } i \in \mathcal{I}_1, \\
\left( 1 - \sum_{l=1}^{L} (b_{1 l} + b_{2 l} + b_{3 l}) I_{(T_l \in A_{(l+1)} \setminus A_d)} \right) I_{(J_{ij} \subseteq A_i \setminus A'_i)} & \text{if } i \in \mathcal{I}_2, \\
\sum_{l=1}^{L} b_{1 l} I_{(J_{ij} \subseteq A_{(l+1)} \setminus A_d)} & \text{if } i \in \mathcal{I}_3, \\
\sum_{l=1}^{L} b_{2 l} I_{(J_{ij} \subseteq B_{il})} & \text{if } i \in \mathcal{I}_4, \\
\sum_{l=1}^{L} b_{3 l} I_{(J_{ij} \subseteq C_{il})} & \text{if } i \in \mathcal{I}_5,
\end{cases}
\]

for \( i = 1, \ldots, n \), and \( j = 1, \ldots, \nu \).

Now consider the set \( \mathcal{A}_2 = \{ \{ T_i \}, \ i \in \mathcal{I}_2 \} \) defined in (4.15), with cardinality set \( n_2 \) (defined after (4.15)). The task of maximization is simplified
4.2 Self-consistency approach for estimation

further through the following result, which is interesting by its own right.

**Theorem 4.** The set $\mathcal{A}_2$ is contained in the set $\mathcal{A}_0$ almost surely. Further, if $G$ is a discrete distribution with finite support, then the probability of $\mathcal{A}_0$ being equal to $\mathcal{A}_2$ goes to one as $n \to \infty$.

We are now ready for the next result regarding the existence and uniqueness of the NPMLE. The uniqueness is established probabilistically under the condition that $n_2$, the number of cases with exact recall, goes to infinity.

**Theorem 5.** The likelihood (4.21) has a maximum. Further, if $G$ is a discrete distribution with finite support, then the probability that it has a unique maximum goes to one, as $n_2 \to \infty$.

### 4.2 Self-consistency approach for estimation

We follow the footsteps of Efron (1967) and Turnbull (1976) to obtain the NPMLE through the self consistency approach. For $i = 1, 2, \ldots, n$, let

$$L_{ij} = \begin{cases} 
1 & \text{if } T_i \in J_j, \\
0 & \text{otherwise},
\end{cases}$$

When $i \in \mathcal{I}_2$, the value of $L_{ij}$ is known. If $i \notin \mathcal{I}_2$, its expectation with respect to the probability vector $p$ is given by

$$E(L_{ij}) = \frac{\alpha_{ij}q_j}{\sum_{j=1}^{\nu} \alpha_{ij}q_j} = \mu_{ij}(p), \text{ say.} \quad (4.23)$$
4.2 Self-consistency approach for estimation

Thus, \( \mu_{ij}(p) \) represents the probability that the \( i \)-th observation lies in \( J_j \).

The average of these probabilities across the \( n \) individuals,

\[
\frac{1}{n} \sum_{i=1}^{n} \mu_{ij}(p) = \pi_j(p), \quad \text{say,} \tag{4.24}
\]

should indicate the probability of the interval \( J_j \). Thus, it is reasonable to expect that the vector \( p \) would satisfy the equation

\[
q_j = \pi_j(p) \quad \text{for} \quad 1 \leq j \leq \nu. \tag{4.25}
\]

An estimator of \( p \) may be called self consistent if it satisfies (4.25). The form of these equations suggests the following iterative procedure.

**Step I.** Obtain a set of initial estimates \( q_j^{(0)} \) (1 \( \leq j \leq m \)).

**Step II.** At the \( n \)th stage of iteration, use current estimate, \( p^{(n)} \), to evaluate \( \mu_{ij}(p^{(n)}) \) for \( i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, \nu \) and \( \pi_j(p^{(n)}) \) for \( j = 1, 2, \ldots, \nu \) from (4.23) and (4.24), respectively.

**Step III.** Obtain updated estimates \( p^{(n+1)} \) by setting \( q_j^{(n+1)} = \pi_j(p^{(n)}) \).

**Step IV.** Return to Step II with \( p^{(n+1)} \) replacing \( p^{(n)} \).

**Step V.** Iterate; stop when the required accuracy has been achieved.

The following theorem shows that equation (4.25) defining a self consistent estimator must be satisfied by an NPML estimator of \( p \).

**Theorem 6.** An NPML estimator of \( p \) must be self consistent.
4.3 A computationally simpler estimator

The computational complexity of the NPMLE depends on the number of segments \( k \) used in the piecewise constant formulation of the function \( \pi_\eta \). One can conceive of a computational simplification on the basis of Theorem 3. According to this theorem, the NPMLE has mass only at points of exact recall of the event, when \( n \) is large. In such a case, the likelihood (4.21) involves \( J_j \)'s that are singletons only.

Formally, let \( t_1, \ldots, t_{n_2} \) be the ordered set of distinct ages at event that have been perfectly recalled, and \( q_1^*, \ldots, q_{n_2}^* \) be the probability masses allocated to them. The likelihood (4.21), subject to the constraint that \( q_j = 0 \) whenever \( J_j \notin A_2 \), is equivalent to the unconstrained maximization of

\[
L(p^*, \eta) = \prod_{i=1}^{n} \left[ \sum_{j=1}^{n_2} \alpha_{ij} q_j^* \right],
\]

(4.26)

with respect to the parameters \( p^* = (q_1^*, \ldots, q_{n_2}^*)^T \) and \( \eta \), over the set

\[
\mathcal{R}^* = \left\{ (p^*, \eta) \mid \sum_{j=1}^{n_2} q_j^* = 1, \ 0 \leq q_1^*, \ldots, q_{n_2}^* \leq 1, \ 0 \leq b_1 \leq \cdots \leq b_k \leq 1 \right\}.
\]

Let the likelihood (4.26) be maximized at \( (\hat{p}^*, \hat{\eta}^*) \), where \( \hat{p}^* = (\hat{q}_1^*, \ldots, \hat{q}_{n_2}^*)^T \).

We define an approximate NPMLE (AMLE) of \( F \) as

\[
\hat{F}_n(t) = \sum_{j: t_j \leq t} \hat{q}_j^*.
\]

(4.27)
4.4 Consistency of the estimators

Let $\Theta$ be the set of all distribution functions over $[t_{\text{min}}, t_{\text{max}}]$, i.e.,

$$\Theta = \{ F : [t_{\text{min}}, t_{\text{max}}] \to [0, 1]; F \text{ right continuous, nondecreasing}; \hspace{1cm} (4.28)$$

$$F(t_{\text{min}}) = 0; F(t_{\text{max}}) = 1 \}.$$

and $\overline{\Theta}$ be the set of all sub-distribution functions, i.e.,

$$\overline{\Theta} = \{ F : [t_{\text{min}}, t_{\text{max}}] \to [0, 1]; F \text{ right continuous, nondecreasing}; \hspace{1cm} (4.29)$$

$$F(t_{\text{min}}) = 0; F(t_{\text{max}}) \leq 1 \}.$$

Note that, with respect to the topology of vague convergence, $\overline{\Theta}$ is compact by Helley’s selection theorem. Further, let $F_0$ denote the true distribution of the time of occurrence of landmark events with density $f_0$, and $F_0(t_{\text{min}}) = 0$.

For any given distribution $F \in \Theta$ having masses restricted to the set $\{t_1, \ldots, t_{n_2}\}$, the log of the likelihood (4.26) can be rewritten as a function of $F$ (instead $q_{1}^{*}, \ldots, q_{n_2}^{*}$) as

$$\ell(F) = \sum_{i=1}^{n} \log \left[ \sum_{j=1}^{n_{2}} \alpha_{ij} \{ F(t_{j}) - F(t_{j-}) \} \right]. \hspace{1cm} (4.30)$$

Define the set

$$E = \{ F : F \in \Theta, E[\ell(F) - \ell(F_0)] = 0 \}, \hspace{1cm} (4.31)$$

which is an equivalence class of the true distribution $F_0$. 

Strong consistency of the AMLE and weak consistency of the NPMLE are established by the following theorems.

**Theorem 7.** In the above set-up, the AMLE \( \{ \hat{F}_n \} \) converges almost surely to the equivalence class \( \mathcal{E} \) of the true distribution \( F_0 \), in the topology of vague convergence.

**Theorem 8.** In the set-up described before Theorem 7, the NPMLE \( \{ \hat{F}_n \} \) converges in probability to the equivalence class \( \mathcal{E} \) of the true distribution \( F_0 \), in terms of the Lévy distance.

5. Simulation of performance

5.1 Parametric estimation

We consider the MLEs based on the current status likelihood \( (2.1) \) (described here as Current Status MLE), the likelihood \( (2.2) \) based on binary recall (described here as Binary Recall MLE) and the likelihood \( (2.5) \) based on partial recall (described here as Partial Recall MLE). Computation of the three MLEs is done through numerical optimization of likelihood using the Quasi-Newton method Nocedal and Wright (2006).

For the purpose of simulation, we generate samples of time-to-event from the Weibull distribution with shape and scale parameters \( \theta_1 \) and \( \theta_2 \), respectively. Thus, \( \theta = (\theta_1, \theta_2) \). We generate the recall probabilities through
5.1 Parametric estimation

the multinomial logistic model, \( \log \left( \frac{\pi_k(u)}{\pi_0(u)} \right) = \alpha_k + \beta_k u, \ k = 1, 2, 3. \)

Since \( \sum_{k=0}^{3} \pi_k(u) = 1 \), the probabilities can be written as

\[
\begin{align*}
\pi_0(u) &= 1 / \left( 1 + \sum_{k=1}^{3} e^{\alpha_k + \beta_k u} \right), \\
\pi_k(u) &= e^{\alpha_k + \beta_k u} / \left( 1 + \sum_{k=1}^{3} e^{\alpha_k + \beta_k u} \right), \quad k = 1, 2, 3, \\
\end{align*}
\]

(5.32)

where \( \eta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3) \). Further, we generate the ‘age at interview’ from the discrete uniform distribution over [8,21].

We use the following sets of values of the parameters.

(i) \( \theta = (10, 12) \) and \( \eta = (-0.05, -0.05, -0.05, 0.01, 0.01, 0.01) \),

(ii) \( \theta = (10, 12) \) and \( \eta = (-2, -1, -0.4, 0.05, 0.3, 0.02) \),

(iii) \( \theta = (10, 12) \) and \( \eta = (-2, -0.7, -1, 0.5, 0.06, 0.2) \),

(iv) \( \theta = (10, 12) \) and \( \eta = (-2, -2, -2, 0.3, 0.08, 0.08) \).

Note that for the chosen value of \( \theta \), the median of the Weibull distribution turns out to be 11.6, which is in line with the median estimated from the data described in Section 4 under a simpler model [Mirzai et al. (2015)].

Also, the chosen values of \( \eta \) correspond to the following probabilities of different types of recall, five years after the event.

(i) \( \pi_0(5) = \pi_1(5) = \pi_2(5) = \pi_3(5) = 0.25 \),

(ii) \( \pi_0(5) = 0.28, \pi_1(5) = 0.46, \pi_2(5) = 0.21, \pi_3(5) = 0.05 \),
5.1 Parametric estimation

(iii) $\pi_\eta(0)(5) = 0.23$, $\pi_\eta(1)(5) = 0.15$, $\pi_\eta(2)(5) = 0.23$, $\pi_\eta(3)(5) = 0.38$,

(iv) $\pi_\eta(0)(5) = 0.5$, $\pi_\eta(1)(5) = 0.1$, $\pi_\eta(2)(5) = 0.1$, $\pi_\eta(3)(5) = 0.3$.

Choice (iv) is meant to favour the Binary Recall MLE, as the chances of partial recall are slim. Choice (ii) should favour the Partial Recall MLE. Choice (iii), with a high probability attached to ‘no recall’, gives Current Status MLE its best chance. Choice (i) does not favour any single method.

While computing the Binary Recall MLE, we assume the following form of the non-recall probability function $\pi_\eta$:

$$\log \left( \frac{\pi_\eta(u)}{1 - \pi_\eta(u)} \right) = \alpha + \beta u.$$

We run 1000 simulations for each of the above combinations of parameters, for sample size $n = 100$, to compute the empirical bias, the standard deviation (Stdev) and the mean squared error (MSE) for the MLEs of the parameter $\theta = (\theta_1, \theta_2)$, the median time-to-event, and $\pi_\eta(0)(5)$ (the exact recall probability 5 years after the event), based on the three likelihoods. These indicators of performance, for the combinations of parameter values given in case (i) to case (iv), are reported in Table I.

In cases (i)–(iii), it is found that the bias and the standard deviation (and consequently the MSE) of the Partial Recall MLE is generally less than (and sometimes comparable to) those of the other two estimators and
### 5.1 Parametric estimation

Table 1: Bias, standard deviation (Stdev) and MSE of estimated parameters for \( n = 100 \)

| Case | Param | Current Status MLE | Binary Recall MLE | Partial Recall MLE |
|------|-------|--------------------|-------------------|--------------------|
|      |       | Bias   | Stdev | MSE | Bias   | Stdev | MSE | Bias   | Stdev | MSE |
| (i)  | \( \theta_1 \) | 1.698  | 5.368 | 31.67 | 0.487  | 1.701 | 3.127 | 0.247  | 1.07 | 1.207 |
|      | \( \theta_2 \) | -0.071 | 0.329 | 0.113 | -0.023 | 0.233 | 0.055 | -0.01  | 0.165 | 0.027 |
|      | Median | -0.047 | 0.338 | 0.116 | -0.012 | 0.241 | 0.058 | -0.003 | 0.172 | 0.029 |
|      | \( \pi_\eta^{(0)}(5) \) | - | - | - | -0.004 | 0.054 | 0.003 | 0.0001 | 0.054 | 0.002 |
| (ii) | \( \theta_1 \) | 1.845 | 5.27 | 31.15 | 0.52 | 1.745 | 3.314 | 0.214 | 0.952 | 0.952 |
|      | \( \theta_2 \) | -0.058 | 0.341 | 0.119 | -0.01 | 0.341 | 0.051 | -0.011 | 0.145 | 0.021 |
|      | Median | -0.031 | 0.347 | 0.121 | -0.002 | 0.24 | 0.058 | -0.005 | 0.152 | 0.023 |
|      | \( \pi_\eta^{(0)}(5) \) | - | - | - | -0.018 | 0.057 | 0.004 | 0.0007 | 0.053 | 0.003 |
| (iii) | \( \theta_1 \) | 1.93 | 5.091 | 29.63 | 0.573 | 1.828 | 3.669 | 0.381 | 1.322 | 1.893 |
|      | \( \theta_2 \) | -0.07 | 0.331 | 0.114 | -0.024 | 0.243 | 0.059 | -0.007 | 0.182 | 0.033 |
|      | Median | -0.037 | 0.337 | 0.115 | -0.011 | 0.255 | 0.065 | -0.002 | 0.193 | 0.037 |
|      | \( \pi_\eta^{(0)}(5) \) | - | - | - | -0.026 | 0.056 | 0.004 | -0.003 | 0.06 | 0.004 |
| (iv)  | \( \theta_1 \) | 1.803 | 5.333 | 31.66 | 0.262 | 1.291 | 1.735 | 0.253 | 1.146 | 1.377 |
|      | \( \theta_2 \) | -0.062 | 0.332 | 0.114 | -0.018 | 0.191 | 0.04 | -0.014 | 0.174 | 0.031 |
|      | Median | -0.036 | 0.34 | 0.117 | -0.012 | 0.202 | 0.041 | -0.008 | 0.185 | 0.034 |
|      | \( \pi_\eta^{(0)}(5) \) | - | - | - | -0.012 | 0.064 | 0.004 | 0.001 | 0.067 | 0.004 |
its performance improves with increasing sample size. The Current Status MLE, which uses the least amount of information from the data, has the poorest performance even in case (iii), where a substantial proportion of the subjects are designed to have no recollection of the event date. The substantial gap between the performance of the Binary Recall MLE and the Partial Recall MLE shows that the later estimator is able to utilize the additional information available from partial recall data. A similar table for \( n = 1000 \) is given in supplementary material, to save space. The conclusions are similar, though all the methods perform better.

5.2 Non-parametric estimation

We generate sample times-to-event \((T)\) from the Weibull distribution with shape and scale parameters \( \theta = (10, 12) \) as before, but truncate the generated samples to the interval \([8, 16]\). This truncated distribution has median of 11.6. The corresponding ‘time of interview’ \((S)\) is generated from the discrete uniform distribution over \([8, \ldots, 21]\). These choices are in line with the data set described in Section 1 and lead to about 29% cases of no-occurrence of the event till the time of interview \((S < T)\). As for the recall probabilities, we use (4.12) with \( L = 4, x_1 = 0, x_2 = 3, x_3 = 6, x_4 = 9 \) and three sets of values of the parameters, described below.
5.2 Non-parametric estimation

Case (a) $b_0 = (0.15, 0.10, 0.08, 0.05), b_1 = (0.28, 0.2, 0.15, 0.1), b_2 = (0.22, 0.25, 0.17, 0.1), b_3 = (0.35, 0.45, 0.6, 0.75)$, which correspond to overall probabilities of exact recall $E[\pi^{(0)}(S-T)|S>T] = 0.10$, recall up to calendar month $E[\pi^{(1)}(S-T)|S>T] = 0.20$, recall up to calendar year $E[\pi^{(2)}(S-T)|S>T] = 0.20$ and no recall $E[\pi^{(3)}(S-T)|S>T] = 0.50$.

Case (b) $b_0 = (0.69, 0.55, 0.49, 0.31), b_1 = b_2 = (0.08, 0.05, 0.03, 0.02), b_3 = (0.15, 0.35, 0.45, 0.65)$ which correspond to overall exact recall probability 0.55, calendar month recall probability 0.05, calendar year recall probability 0.05 and no-recall probability 0.35.

Case (c) $b_0 = 1 - (b_1 + b_2 + b_3), b_1 = b_2 = b_3 = (0.25, 0.25, 0.25, 0.25)$, which correspond to equal probability (0.25) of each type of recall.

It has been observed by Mirzaei and Sengupta (2016) that in the special case of binary recall, the performances of AMLE and NPMLE are comparable. Therefore, we choose not to run simulation for NPMLE, which involves heavier computation. Instead, we compare the performance of the AMLE estimated from (4.27) (described here as Partial Recall AMLE) with those of the AMLE based on (2.2), proposed by Mirzaei and Sengupta (2016) (described here as Binary Recall AMLE), and the empirical estimate of $F$ (described here as EDF). The EDF is used only as a hypothetical benchmark of performance that could have been achieved with complete data.
The Partial Recall AMLE is implemented by assuming that $L, x_1, x_2, \ldots, x_L$ in (4.12) are known, while the likelihood (4.26) is recursively maximized alternately with respect to the probability parameter $p^*$ and the nuisance parameter $\eta = (b_0, b_1, b_2, b_3)^T$.

Figure 2 shows plots of the bias, the variance and the mean square error (MSE) of the three estimators for different ages, when $n = 100$ and parameters of the recall functions (4.12) are chosen as in Cases (a), (b) and (c). The Partial Recall AMLE is found to have smaller bias, variance and MSE than the Binary Recall AMLE, although its performance is expectedly poorer than that of the EDF.

Plots similar to Figure 2 for $n = 1000$ are given in the supplementary material. At that sample size, the performance parameters of partial AMLE are found to be closer to those of EDF than those of binary AMLE.

6. Adequacy of the Model

One can use the chi-square goodness of fit test to check how well the assumed parametric model actually fits the data. For this purpose, the data may be transformed to the vector $Y = (S, V, \varepsilon, \delta, m, d)$, and the support of the distribution of this vector may be appropriately partitioned, depending on the availability of data. An example is given in the next section.
Figure 2: Comparison of bias, variance and MSE of the estimator for $n = 100$ in cases (a) (top panel), (b) (middle panel) and (c) (bottom panel)
Modeling of the recall probability functions is a critical issue. One has to choose suitable functional forms, and also strike a balance between a flexible model and a parsimonious one with fewer parameters. We provide below an exploratory technique for selecting the functional forms.

As we have seen in Section 4, use of the piecewise constant form (4.12) of the recall probabilities reduces the likelihood (2.5) to the likelihood (4.13). If the distribution of $T$ is known, one can obtain the MLE of the parameters $b_{l1}, b_{l2}, \ldots, b_{lk}$, $l = 0, 1, 2, 3$. The conditional MLE of the piecewise constant functions $\pi^{(1)}, \pi^{(2)}, \pi^{(3)}$ and $\pi^{(0)}$, for any given $F_\theta$ can be obtained iteratively. By using a candidate parametric form $\pi^{(1)}_{\hat{\eta}}, \pi^{(2)}_{\hat{\eta}}, \pi^{(3)}_{\hat{\eta}}$ and $\pi^{(0)}_{\hat{\eta}}$, one can first estimate the MLEs $\hat{\theta}$ and $\hat{\eta}$ and then compare the plots of $\hat{\pi}^{(1)}_{\hat{\eta}}, \hat{\pi}^{(2)}_{\hat{\eta}}, \hat{\pi}^{(3)}_{\hat{\eta}}$ and $\hat{\pi}^{(0)}_{\hat{\eta}}$ with the plots of the conditional MLE of the piecewise constant versions of $\pi^{(1)}, \pi^{(2)}, \pi^{(3)}$ and $\pi^{(0)}$, with $F_\theta$ held fixed at $F_{\hat{\theta}}$. An example of this graphical check is given in the next section.

7. Data Analysis

For the data set described in Section 4, the landmark event is the onset of menarche in young and adolescent females. In a parametric analysis, we used the Weibull model for menarcheal age and the multinomial logistic model for the recall probabilities $\pi^{(0)}_{\eta}, \pi^{(1)}_{\eta}, \pi^{(2)}_{\eta}$ and $\pi^{(3)}_{\eta}$, as in Section 5.1.
Table 2: Different estimates of parameters for the menarcheal data

| Estimator               | $\theta_1$ (Stdev) | $\theta_2$ (Stdev) | Median (Stdev) |
|-------------------------|--------------------|--------------------|---------------|
| Current Status MLE      | 10.76 (0.51)       | 12.18 (0.07)       | 11.77 (0.031) |
| Binary Recall MLE       | 10.86 (0.33)       | 12.33 (0.05)       | 11.92 (0.018) |
| Partial Recall MLE      | 10.37 (0.24)       | 12.39 (0.04)       | 11.96 (0.0096) |

We used the three different methods mentioned in Section 5.1 for estimating the parameters $\theta_1$ and $\theta_2$ as well as the median of the age at menarche. Table 2 gives a summary of the findings. The Partial Recall MLEs have smaller standard errors than those of the other two estimators.

Figure 3 shows the survival functions estimated from the three parametric methods and the Partial Recall AMLE presented in Section 4.3 (with knot points of the recall probability functions chosen as in the first paragraph of Section 5.2). The parametric MLEs are not very far from the Partial Recall AMLE. Though there appears to be little difference among the parametric MLEs, their standard errors are different. Figure 4 shows the plots of the widths of the asymptotic pointwise 95% confidence intervals of the estimated survival function based on the three likelihoods. It is clear that the confidence intervals for the Partial Recall MLE are narrower.

In order to formally check how well the assumed parametric model fits
Figure 3: Survival functions for the menarcheal data based on four methods

Figure 4: Width of asymptotic pointwise 95% confidence intervals of survival function for the menarcheal data based on three parametric methods
the data, we use the chi-square goodness of fit test, by discretizing the range of the hexatuple \((S, V, \varepsilon, \delta, m, d)\). The range of \(S\) is split into the intervals \([7, 14]\) and \((14, 21]\), while the range of \(V\) is split into the sets \(\{0\}\), \((0, 11.84]\) and \((11.84, 21]\) (11.84 being the median of the observed non-zero values of \(V\)). The ranges of \(\varepsilon\), \(\delta\) and \(m\) have four points (0, 1, 2 and 3), two points \(\{0\text{ and }1\}\) and twelve points (integers 1 to 12), respectively, none of which are clubbed. The range of \(d\) is the interval \([0, 1/12]\), which is not split.

When \(\delta = 0\), the value of \(\varepsilon\) is irrelevant and \(V = 0\), i.e., there are only two bins for the two groups of values of \(S\). When \(\delta = 1\) and \(\varepsilon = 3\), \(V\) can only be zero and again there are only two bins. When \(\delta = 1\) and \(\varepsilon = 0, 1\) or 2, in each case there are four bins arising from two groups of values of \(S\) and two groups of non-zero values of \(V\). Thus, we have a total of 16 bins.

In order to avoid small expected frequency in some cells, the groups \(\{\delta = 0\}\) and \(\{\delta = 1\} \cap \{\varepsilon = 2\} \cap \{V > 11.84\}\) were not split by the value of \(S\). After this pruning, we have a reduced total of 14 bins. There are 8 parameters to estimate. Thus, the null distribution should be \(\chi^2\) with 5 degrees of freedom. The p-value of the test statistic for the given data happens to be 0.188. Therefore, violation of the chosen model is not indicated.

We now check the adequacy of the functional form of the \(\pi_{\eta}^{(l)}\)s by comparing the \(\pi_{\eta}^{(b)}\)s with the conditional MLE of the corresponding piecewise
Figure 5: Comparison of estimated logistic recall probabilities with estimated piecewise constant recall probabilities with (a) 4 pieces, (b) 8 pieces constant function in (4.12), as indicated in the last section. For the given data, the largest value of $S_i - T_i$ in a perfectly recalled case happens to be 10.88 years. Therefore, we consider recall functions over the interval 0 to 12 years. With $F$ chosen as Weibull and $\theta_1$ and $\theta_2$ fixed at the values reported in the last row of Table 2, we obtained the conditional MLE of the values of $\pi_{\eta}^{(0)}, \pi_{\eta}^{(1)}, \pi_{\eta}^{(2)}, \pi_{\eta}^{(3)}$ in different segments of equal length. Figure 5(a) shows the plots of the estimated recall probabilities under the logistic and the piecewise constant models, with number of segments $L = 4$. The estimated functions are found to be close to each other for $l = 0, 1, 2, 3$. Figure 5(b) shows the same plots for $L = 8$. The finer partition seems unnecessary.
We have seen the cumulative proportions of decreasing degrees of recall for different age ranges in the case of the menarcheal data in Figure 6. As an additional check for the assumed model, we consider the model based estimates of these cumulative proportions for ages \( s = 11, 14, 17 \) and 20 (i.e., at the middle of the respective age intervals). We used the Partial Recall MLE of parameters \( \hat{\theta} \) and \( \hat{\eta} \) to calculate \( f_{\hat{\theta}} \) and \( \pi_{\hat{\eta}}^{(j)} \) for \( j = 0, 1, 2, 3 \) and then computed the requisite probabilities through numerical integration.

Figure 6 shows the cumulative proportions in different age groups (solid lines) along with the corresponding model based estimates (dashed lines). The estimated probabilities are quite close to the empirical proportions.
8. Concluding Remarks

The aim of this paper has been to offer a realistic model for time-to-event based on partial recall information through an informative censoring model, where the range of relevant dates may depend on calendar time (rather than time elapsed since the event). The simulations and the data analysis of the menarcheal data set show that there is much to be gained from partial recall information in the form of the event falling in a calendar month or a calendar year. Many other forms of partial recall information may be handled in a similar way. As the simulations reported in Section 5 show, a particular category of partial recall (e.g., recall up to a calendar month or year) is justified if that category is not very rare in the data.

The recalled time-to-event can sometimes be erroneous. Grouping of the uncertainly recalled event date by the calendar month or year may reduce the error to some extent. If one adopts this solution, the method presented in this paper provides a viable method of analysis. Skinner and Humphreys (1999), while working with data without instances of non-recall, has modeled erroneously recalled time-to-event as $t_i' = t_i k_i$, where $t_i$ is the correct time-to-event and $k_i$ is a multiplicative error of recall that is independent of $t_i$. Since $k_i$s are unobservable, they have used a mixed-effects regression model to account for erroneous recalls. One may investigate whether a
similar adjustment in the term $f_\theta(T_i)$ of the likelihood (2.5), improves the analysis.

It would also be of interest to see the effect of covariates on the time-to-event, under a suitable regression model. This problem will be taken up in future.

**Supplementary Materials**

Contain the proofs and part of simulation results.

**Acknowledgements**

This research is partially sponsored by the project “Physical growth, body composition and nutritional status of the Bengal school aged children, adolescents, and young adults of Calcutta, India: Effects of socioeconomic factors on secular trends”, funded by the Neys Van Hoogstraten Foundation of the Netherlands. The authors thank Professor Parasman Dasgupta of the Biological Anthropology Unit of ISI, for making the data available for this research. The work of the corresponding author has been partially supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (contracts #HHSN275201200005C, #HHSN267200700019C).
References

Ash, R. B. (2000). *Probability and Measure Theory*. Burlington, MA: Harcourt/Academic Press.

Dasgupta, P. (2015), *Physical Growth, Body Composition and Nutritional Status of Bengali School aged Children, Adolescents and Young adults of Calcutta, India: Effects of Socioeconomic Factors on Secular Trends*. (in collaboration with M. Nubé, D. Sengupta and M. de Onis), The Hague: Neys-van Hoogstraten Foundation.

URL: http://www.neys-vanhoogstraten.nl/wp-content/uploads/2015/06/Academic-Report-ID-158.pdf

Efron, B. (1967), “The two sample problem with censored data,” *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 831–853.

Gillespie, B., dÁrcy, H., Schwartz, K., Bobo, J. K., and Foxma, B. (2006), “Recall of age of weaning and other breastfeeding variables,” *International Breastfeeding Journal*, 1, 4.

Harel, Y., Overpeck, M. D., Jones, D. H., Scheidt, P. C., Bijur, P. E., Trumble, A. C., and Anderson, J. (1994), “The effects of recall on estimating annual nonfatal injury rates for children and adolescents,” *American Journal of Public Health*, 84(4), 599–605.

Joffe, M., Villard, L., Li, Z., Plowman, R., and Vessey, M. (1995), “A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England,” *Journal of Epidemiology and Community Health*, 49, 314–319.

Koo, M. M., and Rohana, T. E. (1997), “Accuracy of short-term recall of age at menarche,” *Annals of Human Biology*, 24, 61–64.
Lee, E. T., and Wang, J. W. (2003), *Statistical Methods for Survival Data Analysis*, New York: John Wiley.

Mathiowetz, N. A., and Ouncanb, G. J. (1988), “Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment,” *Journal of Business & Economic Statistics*, 6(2), 221–229.

Mirzaei, S. S., and Sengupta, D. (2016), “Nonparametric estimation of time-to-event distribution based on recall data in observational studies,” *Lifetime Data Analysis*, 22, 473–503.

Mirzaei, S. S., Sengupta, D., and Das, R. (2015), “Parametric estimation of menarcheal age distribution based on recall data,” *Scandinavian Journal of Statistics*, 42, 290–305.

Nocedal, J., and Wright, S. J. (2006), *Numerical Optimization*, New York: Springer.

Skinner, C. J., and Humphreys, K. (1999), “Weibull regression for lifetimes measured with error,” *Lifetime Data Analysis*, 5, 23–37.

Turnbull, B. W. (1976), “The empirical distribution function with arbitrarily grouped, censored and truncated data,” *Journal of the Royal Statistical Society, Series B*, 38, 290–295.

*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, USA

E-mail: (mirzaeisalehas@mail.nih.gov)

Indian Statistical Institute, Kolkata, India, E-mail: (sdebasis@isical.ac.in)

North Carolina State University, E-mail: (rghosal@ncsu.edu)
S1. Parametric estimation

Regularity conditions

Sufficient conditions for consistency given in Theorem 7.1.1 of Lehman (1999).

(C1) The parameters $\theta$ and $\eta$ are identifiable with respect to the family of densities $f_\theta$ of the time-to-event and the family of functions $\pi^{(k)}_\eta$, $k = 1, 2, 3$. In other words, $f_{\theta_1} = f_{\theta_2}$ implies $\theta_1 = \theta_2$ and congruence of $\pi^{(k)}_{\eta_1}$ and $\pi^{(k)}_{\eta_2}$ for $k = 1, 2, 3$ implies $\eta_1 = \eta_2$.

(C2) The parameter spaces for $\theta$ and $\eta$ are open.

(C3) The set $A_1 = \{ t : f_{\theta}(t) > 0 \}$ is independent of $\theta$ and the set $A_2 = \{ t : \pi^{(k)}_\eta(t) \in (0, 1), k = 0, 1, 2, 3 \}$ is independent of $\eta$.

(C4) The functions $f_{\theta}(t)$, $\pi^{(1)}_\eta(t)$, $\pi^{(2)}_\eta(t)$ and $\pi^{(3)}_\eta(t)$ are differentiable with respect to $\theta$ and $\eta$ for all $t$ such that the derivative is absolutely bounded.
by a $\mu$-integrable function.

The additional conditions for asymptotic normality are conditions 1-5 of Theorem 18 (Ferguson, 1996, Chapter 18), where the log-likelihood is

$$
\ell(\theta, \eta) = \sum_{i=1}^{n} \left[ \delta_i I(\epsilon_i=3) \log \left( \int_0^{S_i} f_{\theta}(u) \pi_0^{(3)}(S_i - u) \, du \right) 
+ \delta_i I(\epsilon_i=2) \log \left( \int_{Y_{1i}} f_{\theta}(u) \pi_0^{(2)}(S_i - u) \, du \right) 
+ \delta_i I(\epsilon_i=1) \log \left( \int_{M_{1i}} f_{\theta}(u) \pi_0^{(1)}(S_i - u) \, du \right) 
+ \delta_i I(\epsilon_i=0) \log \left( f_{\theta}(T_i) \pi_0^{(0)}(S_i - T_i) \right) + (1 - \delta_i) \log \left( F_{\theta}(S_i) \right) \right].
$$

(S1.1)

**Proof of Theorem 1**

In the second case, the density can be derived as,

$$
h(s, w, 1, 1, m, d)
= g_1(s)g_2(m)g_3(d) \frac{\partial P(W < w, \delta = 1, \epsilon = 1|s, m, d)}{\partial w}
= g_1(s)g_2(m)g_3(d) \lim_{h \to 0} \frac{P(w < W \leq w + h, \delta = 1, \epsilon = 1|s, m, d)}{h}
= g_1(s)g_2(m)g_3(d) \lim_{h \to 0} \frac{P(w < T \leq w + h, T \leq s, \epsilon = 1)}{h}
= g_1(s)g_2(m)g_3(d) \lim_{h \to 0} \frac{P(w < T \leq w + h, \epsilon = 1)}{h}
= g_1(s)g_2(m)g_3(d) \lim_{h \to 0} \frac{E_T[P(w < T \leq w + h|T)\pi_0(s - T)I(w \leq s)]}{h}
= g_1(s)g_2(m)g_3(d) \lim_{h \to 0} \frac{\int_{w}^{w+h} f_\theta(u) \pi_0(s - u) \, du I_{(w \leq s)}}{h}
= g_1(s)g_2(m)g_3(d) f_\theta(w) \pi_0(s - w) I_{(w \leq s)}.
$$
The density in the other cases can be obtained by considering the corresponding probability masses:

\[
\begin{align*}
  h(s, w, \varepsilon, 0, m, d) &= P(W = 0, \delta = 0|s, m, d)g_1(s)g_2(m)g_3(d) \\
  &= P(T > S|S = s)g_1(s)g_2(m)g_3(d) = \bar{F}_\theta(s)g_1(s)g_2(m)g_3(d); \\
  h(s, w, 0, 1, m, d) &= E_T[g_1(s)g_2(m)g_3(d)P(T \leq s|T, m, d, s)\pi_1(s - T)] \\
  &= \int_0^s g_1(s)g_2(m)g_3(d)f_\theta(u)\pi_1(s - u)du \\
  &= g_1(s)g_2(m)g_3(d)\int_0^s f_\theta(u)\pi_1(s - u)du \\
  h(s, w, 2, 1, m, d) &= g_1(s)g_2(m)g_3(d)P(W = w, \varepsilon = 2, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)P([12(d + T)]/12 = w, \varepsilon = 2, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)P(12w \leq 12(d + T) < 12w + 1, \varepsilon = 2, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)\int_{w - d}^{w + d} f_\theta(u)\pi_2(s - u)du; \\
  h(s, w, 3, 1, m, d) &= g_1(s)g_2(m)g_3(d)P(W = w, \varepsilon = 3, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)P([T + d + (m - 1)/12] = w, \varepsilon = 3, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)P(w - d - (m - 1)/12 \leq T < w + 1 - d - (m - 1)/12, \varepsilon = 3, \delta = 1|s, m, d) \\
  &= g_1(s)g_2(m)g_3(d)\int_{w - d - \frac{m - 1}{12}}^{w + 1 - d - \frac{m - 1}{12}} f_\theta(u)\pi_3(s - u)du;
\end{align*}
\]

**S2. Non-parametric estimation**

Proof of Theorem 2

In last theorem it is shown that, the density of \(Y = (S, V, \varepsilon, \delta, m, d)\) with respect to
the measure $\mu$ is

$$
\begin{aligned}
h(s, v, \varepsilon, \delta, m, d) &=
\begin{cases}
g_1(s)g_2(m)g_3(d)F(s) & \text{if } \delta = 0,
g_1(s)g_2(m)g_3(d)f(v)\pi^{(0)}(s - v)I_{v<s} & \text{if } \varepsilon = 0 \text{ and } \delta = 1,
g_1(s)g_2(m)g_3(d)\int_{v-d}^{v+1-d} f(u)\pi^{(1)}(s - u)du & \text{if } \varepsilon = 1 \text{ and } \delta = 1,
g_1(s)g_2(m)g_3(d)\int_{v-d}^{v+1-d} \frac{e^{-\frac{u-s}{\mu}}}{\mu^{\frac{1+\mu}{2}}} f(u)\pi^{(2)}(s - u)du & \text{if } \varepsilon = 2 \text{ and } \delta = 1,
g_1(s)g_2(m)g_3(d)\int_{0}^{s} f(u)\pi^{(3)}(s - u)du & \text{if } \varepsilon = 3 \text{ and } \delta = 1,
\end{cases}
\end{aligned}
$$

\tag{S2.2}

where $g_1$, $g_2$ and $g_3$ are the densities of $G_1$, $G_2$ and $G_3$ with respect to the measures $\vartheta_1$, $\vartheta_5$ and $\vartheta_6$, respectively.

(a) We have, from (S2.2), $g_1(s) = \int_{v,\varepsilon,\delta,m,d} h(s, v, \varepsilon, \delta, m, d)$ and hence $G_1$ are identifiable from $h$. It is the same for $g_2$, $G_2$ and $g_3$, $G_3$.

Also we have

$$
\tilde{F}(s) = \frac{h(s, 0, \varepsilon, 0, m, d)}{g_1(s)g_2(m)g_3(d)},
$$

and

$$
\pi_1(s - v) = \frac{h(s, v, 1, 1, m, d)}{g_1(s)g_2(m)g_3(d)},
$$

that show $F$ and $\pi_1$ are identifiable form $h$.

(b) For the sake of contradiction, let us assume there are two $\pi_2$s, say $\pi^1_2$ and $\pi^2_2$, such that their substitution in the right hand side of (S2.2) produces the same function
\( h_1 = h_2 = h \). By differentiating \( h \) w.r.t. \( v \), we get

\[
\frac{dh(s, v, 2, 1, m, d)}{dv} = f((v + 1/12) - d) \pi_1^2(s - (v + 1/12) + d) - f(v - d) \pi_2^1(s - v + d)
\]

\[
= f((v + 1/12) - d) \pi_2^1(s - (v + 1/12) + d) - f(v - d) \pi_2^1(s - v + d).
\]

Hence,

\[
f((v + 1/12) - d) \pi_1^1(s - (v + 1/12) + d) - \pi_2^2(s - (v + 1/12) + d)
\]

\[
= f(v - d) \left[ \pi_1^1(s - v + d) - \pi_2^2(s - v + d) \right],
\]

which implies,

\[
\frac{f((v + 1/12) - d)}{f(v - d)} = \frac{\pi_1^1(s - v + d) - \pi_2^2(s - v + d)}{\pi_1^1(s - (v + 1/12) + d) - \pi_2^2(s - (v + 1/12) + d)} > 0 \quad \forall s.
\]

Since the numerator and the denominator are the same function evaluated at two different points, we have either

\[
\pi_1^1(s - v + d) - \pi_2^2(s - v + d) > 0 \quad \forall s,
\]

or,

\[
\pi_1^1(s - v + d) - \pi_2^2(s - v + d) < 0 \quad \forall s.
\]

Without loss of generality, let \( \pi_1^1(s - v + d) - \pi_2^2(s - v + d) > 0 \quad \forall s \), i.e., \( \pi_1^1 > \pi_2^2 \).

Since

\[
h_1 = \int_{v-d}^{(v+1/12)-d} f(u) \pi_1^1(s - u) du
\]

\[
= \int_{v-d}^{(v+1/12)-d} f(u) \left[ \pi_1^1(s - u) - \pi_2^2(s - u) \right] du
\]

\[
+ \int_{v-d}^{(v+1/12)-d} f(u) \pi_1^1(s - u) du,
\]
we have

\[ h_1 - h_2 = \int_{v-d}^{(v+1/12) - d} f(u) \left[ \pi_2^1(s - u) - \pi_2^2(s - u) \right] du > 0, \]

which contradicts the assumption. Therefore, \( \pi_2 \) is uniquely defined for any given \( h \). A similar argument can be used to show that \( \pi_3 \) is identifiable from \( h \). From the identity \( \sum_{k=0}^{3} \pi_k = 1 \), we conclude that all the \( \pi \)s are identifiable from \( h \).

**Proof of Theorem 3**

From the definitions of \( C \) and \( C_0 \), we can rewrite the likelihood (4.20) as follows.

\[
L = \prod_{i \in I_1} \left( \sum_{r: i \subseteq A_{i}} p_r + \sum_{r: i \subseteq A_{i}, s_r \in C \setminus C_0} p_r \right) \times \prod_{i \in I_2} \left( 1 - \sum_{l=1}^{L} (b_{1l} + b_{2l} + b_{3l}) I_{(T_i \in A_{(i+1)} \setminus A_{i})} \right) \cot \left( \sum_{r: i \subseteq A_{i}, s_r \in C \setminus C_0} p_r + \sum_{r: i \subseteq A_{i}, s_r \in C_0} p_r \right) \\
\times \prod_{i \in I_3} \left[ \sum_{l=1}^{L} b_{3l} \left( \sum_{r: i \subseteq A_{(i+1)} \setminus A_{i}, s_r \in C \setminus C_0} p_r + \sum_{r: i \subseteq A_{i}, s_r \in C_0} p_r \right) \right] \\
\times \prod_{i \in I_4} \left[ \sum_{l=1}^{L} b_{2l} \left( \sum_{r: i \subseteq C_{i}, s_r \in C_0} p_r + \sum_{r: i \subseteq C_{i}, s_r \in C_0} p_r \right) \right]. \quad (S2.3) \]

For any \( s_r \in C \setminus C_0 \), let \( A_r = \{I_r' : s_r \in C_0, s_r \subset s_r'\} \). By definition of \( C_0 \), \( A_r \) is a non-empty set. The elements of \( A_r \) are disjoint sets consisting of unions of intervals, which are subsets of \( [t_{\min}, t_{\max}] \). Let \( I_{r'} \) be that member of \( A_r \) which satisfies the condition ‘there is \( \alpha \in I_{r'} \) such that \( \alpha < \beta \) whenever \( \beta \in I_{r'} \) for any \( I_{r'} \in A_r \)’ (in some
sense, it is the minimal element in $A_r$). We are going to show that by moving mass from any $I_r$ to $I_{r^*}$, there won’t be reduction in the contribution of any individual to the likelihood (20). The change in the likelihood would be through the sets $B_j$ such that $j \in (s_{r^*} \setminus s_r)$.

We shall check the effect of shift of mass on the contribution of each individual $(i = 1, \ldots, n)$ to the likelihood.

Case (a). For any $j \in (s_{r^*} \setminus s_r)$, let $i_j$ be such that $B_j = A_{i_j}$ for $i_j \in I_1$. Since $I_{r^*} \subseteq A_{i_j}$, but $I_r \not\subseteq A_{i_j}$, contribution of individual $i_j$ will increase by shifting mass from $I_r$ to $I_{r^*}$.

Case (b). For any $j \in (s_{r^*} \setminus s_r)$, let $i_j$ be such that $B_j = A_{i_j} \setminus A'_{i_j}$ for $i_j \in I_2$. Since $I_{r^*} \subseteq A_{i_j}$, but $I_r \not\subseteq A_{i_j}$, by construction $B_{n_2+j} = A'_{i_j}$ which is disjoint from $B_j$, and we have $n_2 + j \not\in s_{r^*}$, i.e., $n_2 + j \not\in s_r$. This implies $I_r \not\subseteq B_j$ and $I_r \not\subseteq B_{n_2+j}$, i.e., $I_r \not\subseteq A_{i_j}$ and $I_r \not\subseteq A'_{i_j}$. Therefore, contribution of individual $i_j$ will increase by shifting mass from $I_r$ to $I_{r^*}$.

Case (c). For any $j \in (s_{r^*} \setminus s_r)$, let $i_j$ be such that $B_j = A_{i_j(t+1)} \setminus A'_{i_j}$ for $i_j \in I_3$, $l = 1, 2, \ldots, k$. Contribution of individual $i_j$ will increase by shifting mass from $I_r$ to $I_{r^*}$ because $I_{r^*} \in A_{i_j(t+1)}$ and $I_{r^*} \in A_{i_jl}$, but $I_r$ is not in either of them.

Case (d). For any $j \in (s_{r^*} \setminus s_r)$, let $i_j$ be such that $B_j = B_{i_jl}$ for $i_j \in I_4$, $l = 1, 2, \ldots, k$. Contribution of individual $i_j$ will increase by shifting mass from $I_r$ to $I_{r^*}$ because $I_{r^*} \in B_{i_jl}$, but $I_r \not\in B_{i_jl}$.

Case (e). For any $j \in (s_{r^*} \setminus s_r)$, let $i_j$ be such that $B_j = C_{i_jl}$ for $i_j \in I_5$, $l = 1, 2, \ldots, k$. Contribution of individual $i_j$ will increase by shifting mass from $I_r$ to $I_{r^*}$ because $I_{r^*} \in C_{i_jl}$, but $I_r \not\in C_{i_jl}$.
It follows that maximizing $L$ can be restricted to $\{p_r : s_r \in C_0\}$.

**Proof of Theorem 4**

Let $i \in \mathcal{I}_2$ and index $j_i$ be such that $S_{j_i} = \{j : T_i \in B_j\}$. Since time-to-event has absolutely continuous distribution, the recalled times $T_i$ for $i \in \mathcal{I}_2$ are distinct with probability 1. Therefore $T_i \in \{B_1, B_2, \ldots, B_{n_2}\}$ almost surely and $I_{j_i} = T_i$ with probability 1. Also, $S_{j_i} = \{j : T_i \in B_j\} \in C_0$. Therefore, $A_2 \subseteq A_0$.

The interview times are discrete valued with finite domain; $x_1, x_2, \ldots, x_k$ are also finite. So, there are finite number of sets in the form of $A_{i\overline{1}}, B_{i\overline{1}}, C_{i\overline{1}}$. Therefore, even when $n$ is large, there is at most a finite number ($\text{say } N$) of distinct sets of the form

$$A_s = \left\{ \bigcap_{i \in s} B_i \right\} \bigcap \left\{ \left( \bigcap_{i \in \mathcal{I}_1 \cup \mathcal{I}_3 \cup \mathcal{I}_4 \cup \mathcal{I}_5 \setminus s} B_i^r \right) \right\},$$

where $s \subseteq \mathcal{I}_1 \cup \mathcal{I}_3 \cup \mathcal{I}_4 \cup \mathcal{I}_5$.

Let $s^{(1)}, s^{(2)}, \ldots, s^{(N)}$ be the index sets corresponding to the $N$ distinct sets described above. Consider a member of $A_0$, say $I_s$, where $s$ is a subset of $\{1, 2, \ldots, n\}$. If $s \subseteq \mathcal{I}_2$, then it is already a singleton. If not, it can be written as $s^{(j)} \cup (s \setminus s^{(j)})$, with $s^{(j)} \subseteq \mathcal{I}_1 \cup \mathcal{I}_3 \cup \mathcal{I}_4 \cup \mathcal{I}_5$ and $s \setminus s^{(j)} \subseteq \mathcal{I}_2$ for some $j \in \{1, 2, \ldots, N\}$. Let us consider further cases.

Case (a). Let $s = s^{(j)} \cup \{r\}$ for $r \in \mathcal{I}_2$. In this case, $I_s$ is either a singleton or a null set. If it is a null set, then it cannot be a member of $A$, and hence of $A_0$. Thus, Case (a) contributes only singletons to $A_0$.

Case (b). Let $s = s^{(j)} \cup \{r_1, r_2, \ldots, r_p\}$, for $r_1, r_2, \ldots, r_p \in \mathcal{I}_2$ when $p > 1$. In this case, $I_s$ is either a singleton or a null set. Since the absolute continuity of the time-to-event distribution almost surely precludes coincidence of two sample values (say, $T_{r_1}$ and $T_{r_2}$),
I_s is a null set with probability 1. In summary, Case (b) cannot contribute anything other than a singleton to $A_0$.

Case (c). Let $s = s^{(j)}$. The probability that a specific individual (say, the $i$-th one) has the landmark event at an age contained in $A_{s^{(j)}}$ is $P(T_i \in A_{s^{(j)}}, i \in \mathcal{I}_2)$. Since this quantity is strictly positive, the probability that none of the $n$ individuals have had the landmark event in $A_{s^{(j)}}$ and recalled the date is $(1 - P(T_i \in A_{s^{(j)}}, i \in \mathcal{I}_2))^n$, which goes to zero as $n \to \infty$. Thus, the probability that there is $i \in \mathcal{I}_2$ such that $T_i \in A_{s^{(j)}}$ goes to one as $n \to \infty$. Therefore, $I_s \cup \{i\} = I_s \cap \{T_i\}$ is non-null. Consequently $I_s \in A_2$, which means $A_0 \subset A_2$ almost surely. It follows that $P[A_2 = A_0]$ goes to one as $n \to \infty$.

The statement of the theorem follows by combining these cases.

Proof of Theorem 5

From (4.22), the log-likelihood is

$$\ell(p, \eta) = \sum_{i=1}^{n} \ln \left( \sum_{j=1}^{v} \alpha_{ij} q_{j} \right).$$  \hfill (S2.4)

We maximize $\ell(p, \eta)$ periodically with respect to $p$ and $\eta$. If $(p^{(n)}, \eta^{(n)})$, be the iterate at the $n$th stage, the next iterate $(p^{(n+1)}, \eta^{(n+1)})$ is defined by

$$\eta^{(n+1)} = \begin{cases} 
\eta^{(n)} & \text{if } n \text{ is even}, \\
\text{argmax } \ell(p^{(n)}, \eta^{(n)}) & \text{if } n \text{ is odd}, \\
\eta \in M_2 
\end{cases}$$  \hfill (S2.5)

$$p^{(n+1)} = \begin{cases} 
P^{(n)} & \text{if } n \text{ is odd}, \\
\text{argmax } \ell(p, \eta^{(n)}) & \text{if } n \text{ is even}, \\
p \in M_1 
\end{cases}$$  \hfill (S2.6)

where $M_1 = \{p : \sum_{j=1}^{v} q_j = 1, 0 \leq q_i \leq 1, i = 1, 2, \ldots, v\}$ and $M_2 = \{\eta : \eta_i = (b_{i1}, b_{i2}, \ldots, b_{ik}), 0 \leq b_{il} \leq 1, \ \forall i, \forall l\}$. We shall show that the functions $\ell(p, \cdot)$ and $\ell(\cdot, \eta)$
are concave over the convex sets $M_1$ and $M_2$, respectively, so that there exists a maximum at each iteration. Thus, in each stage there is an increase in the likelihood (4.22), which is bounded by $(k\epsilon)^n$, and the sequence of partially maximized likelihoods converges. Under the conditions of the theorem, we shall also show that the objective function is strictly concave, which implies the uniqueness of the maximum at each stage, with probability tending one when $n_2$ goes to infinity. Eventually, as $M_1 \times M_2$ is a closed set, the sequence of maxima obtained at successive stages converges to a unique limit, with probability going to one.

Let $B$ be an $n \times v$ matrix such that the $ij$th be $\alpha_{ij}$. For fixed $b$, the partial derivative of (S2.4) with respect to $p$ is

$$\frac{\partial \ell}{\partial p} = \sum_{i=1}^{n} \frac{B_i}{B_i^T p}$$

where $B_i$ is the $i$th row of the matrix $B$. The second derivative or the Hessian is

$$\frac{\partial \ell}{\partial p \partial p^T} = -\sum_{i=1}^{n} \frac{B_iB_i^T}{(B_i^T p)^2}$$

which is a non-positive definite matrix. Therefore, $\ell$ is a concave function over a convex and bounded domain, which ensures existence of maxima; see Simon and Blume (1994).

Now, we need to show that the Hessian matrix is negative definite in the long run. It is enough to show that for any non-zero vector $u$,

$$P \left( \sum_{i=1}^{n} \frac{(B_i^T u)^2}{(B_i^T p)^2} = 0 \right) \rightarrow 0.$$ 

In other words, we need to show that for any arbitrary non-zero vector $u$,

$$P \left( B_i^T u = 0 \quad \forall i \right) = P \left( Bu = 0 \right) \rightarrow 0.$$ 

For $i \in I_2$, $B_i$ has only one non-zero element. In this situation, the equation $B_i^T u = 0$ implies that the corresponding element of $u$ is zero. Moreover, Theorem 4 implies that
the intervals $J_j \in A_0$ associated with columns of $B$ correspond only to singleton members $(A_2)$ with probability tending to 1. Therefore, with probability tending to one, the event $Bu = 0$ coincides with the event $u = 0$, which has probability zero.

For fixed $p$, the first derivative of (S2.4) with respect to $\eta$ is

$$\frac{\partial \ell}{\partial \eta} = \sum_{i=1}^{n} \frac{A_ip}{B_i^T p}$$

where $A_i$ is the $3k \times m$ matrix with the $(l, j)$th element given by $\frac{\partial \alpha_{ij}}{\partial b_1^l}$, for $l = 1, 2, \ldots, k$, $\frac{\partial \alpha_{ij}}{\partial b_{2l}}$, for $l = k + 1, \ldots, 2k$ and $\frac{\partial \alpha_{ij}}{\partial b_{3l}}$, for $l = 2k + 1, \ldots, 3k$.

The Hessian with respect to $\eta$ is

$$\frac{\partial^2 \ell}{\partial \eta \partial \eta^T} = -\sum_{i=1}^{n} \left(B_i^T p\right)^{-2} A_ipp^T A_i^T$$

which is non-positive definite matrix. Therefore $\ell$ is a concave function over a convex domain, which guarantees the existence of a maximum; see Simon and Blume (1994).

Now, to show the Hessian matrix is negative definite with probability tending to one, we need to show that for any arbitrary non-zero vector $v$,

$$P \left(v^T A_i p = 0 \quad \forall i \right) \to 0.$$

(S2.10)

For $i \in I_2$,

$$A_ip = - \left( \sum_{j=1}^{u} q_j \cdot I(J_j \subseteq A_i) \right) \left( I(T_i \in A_{i1}), \ldots, I(T_i \in A_{ik}) \right)^T,$$

(S2.11)

which is a vector with a non-zero element exactly at one place. The condition $v^T A_i p = 0$ is equivalent to the requirement that the element of $v$ corresponding to the non-zero element of $A_i p$ is zero. On the other hand, as $n_2 \to \infty$,

$$P \left( \sum_{l \in I_2} I((S_i - T_i) \in [x_l, x_{l+1}]) = 0 \right) = \left[ P\left((S_i - T_i) \in [x_l, x_{l+1}]|\delta_i \varepsilon_i = 1\right) \right]^{n_2} \to 0 \quad \forall l.$$
Hence, for all $l = 1, \ldots, k$, there is at least one $i \in \mathcal{I}_2$ such that $T_i \in A_{il}$, with probability tending to one. Therefore, the condition $v^T A_i p = 0 \ \forall i \in \mathcal{I}_2$ reduces, with probability tending to one, to the requirement that all the elements of $v$ are zero. Therefore, for $v \neq 0$, we have $P (v^T A_i p = 0, \ \forall i) \leq P (v^T A_i p = 0, \ \forall i \in \mathcal{I}_2) \to 0$. Thus, the probability that the Hessian matrix defined in (S2.9) is negative definite goes to one. This completes the proof.

**Proof of Theorem 6**

We can incorporate the constraint $\sum_{j=1}^v q_j = 1$, by using the Lagrange multiplier, to maximize

$$\ell = \sum_{i=1}^n \left( \ln \left( \sum_{j=1}^v \alpha_{ij} q_j \right) \right) + \lambda \left( \sum_{j=1}^v q_j - 1 \right).$$

(S2.12)

By setting the derivative of $\ell$ with respect to $\lambda$ equal to 0, we have

$$\frac{\partial \ell}{\partial \lambda} = \sum_{j=1}^v q_j - 1 = 0.$$  

(S2.13)

On the other hand, by setting the derivative of $\ell$ with respect to $q_j$s equal to 0, we obtain

$$\frac{\partial \ell}{\partial q_j} = \sum_{i=1}^n \frac{\alpha_{ij}}{\sum_{r=1}^v \alpha_{ir} q_r} - \lambda = 0 \ \ \forall j = 1, 2, \ldots, m.$$  

(S2.14)

By multiplying both sides of (S2.14) by $q_j$ and adding them over all values of $j$, we get

$$\sum_{j=1}^v \sum_{i=1}^n \frac{\alpha_{ij} q_j}{\sum_{r=1}^v \alpha_{ir} q_r} = \lambda \sum_{j=1}^v q_j,$$

(S2.15)

which simplifies, after interchange of the summations and utilization of (S2.13), to

$$\lambda = n.$$  

(S2.16)

By substituting into (S2.14) the optimum value of $\lambda$ obtained above, we have

$$\sum_{i=1}^n \frac{\alpha_{ij}}{\sum_{r=1}^v \alpha_{ir} q_r} = n \ \ \text{for} \ j = 1, \ldots, v,$$

for $j = 1, \ldots, v$. 


that is

$$\frac{1}{n} \sum_{i=1}^{n} \alpha_{ij} q_j = q_j \quad \text{for } j = 1, \ldots, v,$$

Thus, $\pi_j(p) = q_j \ \forall j$, and the statement is proved.

**Proof of Theorem 7**

We use Theorem 3.1 of Wang (1985), which was used by Gentleman and Geyer (1994). There are five assumptions that we need to check in order to establish consistency of the AMLE.

The first assumption requires a separable compactification of the parameter space $\Theta$. In our case, the set $\Theta$ serves this purpose. For metric we can use the Lévy distance, and compactness follows from the Helley selection theorem. In order to establish separability (Billingsley, 1968, p. 239), we use the Homeomorphic mapping of $[t_{\min}, t_{\max}]$ to $[0, 1]$.

To take care of non-identifiability as in Redner (1981), the equivalence class $E$ defined by

$$E = \{ F : F \in \Theta, E[ \ell(F) - \ell(F_0) ] = 0 \}, \quad (S2.17)$$

is regarded as a single point in $\Theta$.

Let, for $r = 1, 2, \ldots, V_r(F)$ be the Lévy neighborhood of $F \in \Theta$ with radius $1/r$. For such a sequence of decreasing open neighborhoods, Wang (1985)'s second assumption requires that, for any $F_0$ in $\Theta$, there is a function $F_r : \Theta \rightarrow V_r(F_0)$ such that (a) $\ell(F) - \ell(F_r(F))$ is locally dominated on $\Theta$ and (b) $F_r(F)$ is in $\Theta$ if $F \in \Theta$. We define $F_r(F) = \frac{1}{r+1} F + \frac{r}{r+1} F_0$. Since $\| F_r(F) - F_0 \| = \frac{1}{r+1} \| F - F_0 \|$, and the Lévy distance is dominated by the Kolmogorov-Smirnov distance, it is clear that $F_r(F) \in V_r(F_0)$. 
Condition (b) is obviously satisfied. As for condition (a), note that

\[
\sup_{F \in \Theta} \left[ \ell(F) - \ell(F_{r}) \right] \\
= \sup_{F \in \Theta} \ln \left( \frac{\sum_{j=1}^{n} \alpha_{ij} (F(t_{j}) - F(t_{j-1}))}{r+1} \right) \\
\leq \ln(r + 1),
\]

which has finite expectation. Thus, \( \ell(F) - \ell(F_{r}(F)) \) is globally dominated on \( \Theta \).

The third assumption requires that \( E[\ell(F) - \ell(F_{r}(F))] < 0 \) for \( F_{0} \in \Theta, F \in \Theta, F \neq F_{0} \). Here, \( F_{0} \) needs to be interpreted as \( \mathcal{E} \), and the result follows along the lines of the proof of Lemma 4.4 of Wang (1985).

The fourth and fifth assumptions require that \( \ell(F) - \ell(F_{r}(F)) \) is lower and upper semicontinuous for \( F \in \Theta \) except for a null set of points (which may depend on \( F \) only in the case of upper semicontinuity). Both the conditions follow from the portmanteau theorem (Billingsley, 1968, p. 11), as argued by Gentleman and Geyer (1994). No null set needs to be invoked.

The result follows from Theorem 3.1 of Wang (1985) as all the assumptions hold.
Proof of Theorem 8

Theorem 7 implies that

\[ \inf_{F \in \mathcal{E}} d_L(\tilde{F}_n, F) \to 0 \quad \text{as } n \to \infty \quad \text{with probability 1.} \]

Therefore \( P(\inf_{F \in \mathcal{E}} d_L(\tilde{F}_n, F) > \epsilon) \to 0. \)

By using Theorem 4, we have \( P(\omega: \tilde{F}_n(\omega) = \hat{F}_n(\omega)) \to 1, \) and we conclude

\[ P \left( \inf_{F \in \mathcal{E}} d_L(\tilde{F}_n, F) > \epsilon \right) \to 0. \]

S3. Simulation of performance

Parametric estimation:

Table 1 shows the bias, the standard deviation (Stdev) and the mean squared error (MSE) for the MLEs of the parameter \( \theta = (\theta_1, \theta_2) \), the median of time-to-event, and the estimated exact recall probability 5 years after the event, based on the three likelihoods, for the combination of parameter values in case (i) to case (iv) of Section 5.1 of the main paper, for sample size \( n = 1000 \). As with simulations with \( n = 100 \) reported in Table 1 of the paper, in cases (i)-(iii), it is found that the bias and the standard deviation (and consequently the MSE) of the Partial Recall MLE is less than those of the other two estimators and its performance improves with increasing sample size. In case (iv) also (the case where the parameters are chosen to produce lesser proportion of partial recalls), it is seen that the overall performance of the proposed Partial Recall MLE is better than that of the Binary Recall MLE, even though for sample size 1000, the bias of the Binary recall MLE of some parameters is smaller.
| Case | Param | Current Status MLE | Binary Recall MLE | Partial Recall MLE |
|------|-------|---------------------|--------------------|--------------------|
|      |       | Bias    | Stdev | MSE     | Bias    | Stdev | MSE     | Bias    | Stdev | MSE     |
| (i)  | $\theta_1$ | 0.119   | 0.811 | 0.672   | 0.037   | 0.497 | 0.248   | 0.09    | 0.329 | 0.116   |
|      | $\theta_2$ | -0.007  | 0.103 | 0.011   | -0.004  | 0.075 | 0.006   | -0.003  | 0.053 | 0.002   |
|      | Median  | -0.005  | 0.106 | 0.011   | -0.003  | 0.08  | 0.006   | -0.0001 | 0.056 | 0.003   |
|      | $\pi_{\eta}^{(0)}(5)$ | -   | -   | -      | -0.0002 | 0.0161 | 0.0002 | 0.0003 | 0.016 | 0.0002 |
| (ii) | $\theta_1$ | 0.123   | 0.803 | 0.66    | 0.046   | 0.483 | 0.235   | 0.038   | 0.288 | 0.09    |
|      | $\theta_2$ | -0.007  | 0.103 | 0.011   | -0.002  | 0.071 | 0.005   | -0.002  | 0.046 | 0.002   |
|      | Median  | -0.004  | 0.105 | 0.011   | -0.001  | 0.077 | 0.006   | -0.0008 | 0.048 | 0.002   |
|      | $\pi_{\eta}^{(0)}(5)$ | -   | -   | -      | -0.014  | 0.017 | 0.0005  | 0.0004  | 0.017 | 0.0003  |
| (iii)| $\theta_1$ | 0.123   | 0.842 | 0.723   | 0.043   | 0.528 | 0.281   | 0.107   | 0.395 | 0.167   |
|      | $\theta_2$ | -0.007  | 0.103 | 0.011   | -0.002  | 0.077 | 0.006   | -0.002  | 0.057 | 0.003   |
|      | Median  | -0.005  | 0.105 | 0.011   | -0.001  | 0.083 | 0.007   | 0.001   | 0.062 | 0.003   |
|      | $\pi_{\eta}^{(0)}(5)$ | -   | -   | -      | -0.019  | 0.017 | 0.0006  | -0.0002 | 0.017 | 0.0003  |
| (iv) | $\theta_1$ | 0.108   | 0.798 | 0.649   | 0.036   | 0.406 | 0.166   | 0.069   | 0.358 | 0.133   |
|      | $\theta_2$ | -0.009  | 0.101 | 0.01    | -0.0001 | 0.06  | 0.003   | -0.0007 | 0.05  | 0.002   |
|      | Median  | -0.006  | 0.103 | 0.011   | 0.001   | 0.064 | 0.004   | 0.001   | 0.057 | 0.003   |
|      | $\pi_{\eta}^{(0)}(5)$ | -   | -   | -      | -0.012  | 0.02  | 0.0005  | 0.002   | 0.02  | 0.0004  |
Non-Parametric estimation:

Figure 1 shows the plots of the bias, the variance and the mean square error (MSE) of the three estimators for different ages, when \( n = 1000 \) and the parameters of the recall probability functions (4.12) are chosen as in Cases (a), (b) and (c). The Partial Recall AMLE is found to have smaller bias, variance and MSE than the Binary Recall AMLE estimator, although its performance is expectedly poorer than that of EDF. In contrast with the case of \( n = 100 \) reported in the paper, the performance parameters of the Partial Recall AMLE is found to be closer to those of the EDF (the benchmark usable only for complete data) than with those of the Binary Recall AMLE.
Figure 1: Comparison of bias, variance and MSE of the estimator for $n = 1000$ in cases (a) (top panel), (b) (middle panel) and (c) (bottom panel)
Bibliography

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Ferguson, T. S. (1996), *A Course in Large Sample Theory*, Texts in Statistical Science Series, London: Chapman & Hall, London.

Gentleman, R., and Geyer, C. J. (1994), “Maximum likelihood for interval censored data: consistency and computation,” *Biometrika*, 81(3), 618–623.

Lehman, E. L. (1999), *Elements of Large-Sample Theory*, New York: Springer-Verlag.

Redner, R. (1981), “Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions,” *Ann. Statist.*, 9(1), 225–228.

Simon, C. P., and Blume, L. (1994), *Mathematics for Economists*, New York: W. W. Norton.

Wang, J.-L. (1985), “Strong consistency of approximate maximum likelihood estimators with applications in nonparametrics,” *Annals of Statistics*, 13(3), 932–946.