

# Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times

BY M. H. MAATHUIS

*Seminar für Statistik, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland*  
maathuis@stat.math.ethz.ch

AND M. G. HUDGENS

*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.*  
mhudgens@bios.unc.edu

## SUMMARY

New methods and theory have recently been developed to nonparametrically estimate cumulative incidence functions for competing risks survival data subject to current status censoring. In particular, the limiting distribution of the nonparametric maximum likelihood estimator and a simplified naive estimator have been established under certain smoothness conditions. In this paper, we establish the large-sample behaviour of these estimators in two additional models, namely when the observation time distribution has discrete support and when the observation times are grouped. These asymptotic results are applied to the construction of confidence intervals in the three different models. The methods are illustrated on two datasets regarding the cumulative incidence of different types of menopause from a cross-sectional sample of women in the United States and subtype-specific HIV infection from a sero-prevalence study in injecting drug users in Thailand.

*Some key words:* Competing risk; Confidence interval; Current status data; Interval censoring; Nonparametric maximum likelihood estimator; Survival analysis.

## 1. INTRODUCTION

Current status data with competing risks arise in cross-sectional studies that assess the current status of individuals in the sample with respect to an event that can be caused by several mechanisms. An example is Cycle I of the Health Examination Survey in the United States ([MacMahon & Worcester, 1966](#)). This study recorded the age and menopausal status of the female participants, where menopausal status could be pre-menopausal, post-menopausal due to an operation or post-menopausal due to natural causes. Based on these data, the cumulative incidence of natural and operative menopause can be estimated as a function of age. A second example is the Bangkok Metropolitan Administration injecting drug users cohort study ([Kitayaporn et al., 1998](#); [Vanichseni et al., 2001](#)). This study recorded the age and HIV status of injecting drug users, where HIV status could be HIV negative, HIV positive with subtype B, HIV positive with subtype E or HIV positive with some other subtype. Based on these data, the subtype-specific cumulative incidence of HIV can be estimated as a function of age.

New methods and theory have recently been developed to nonparametrically estimate cumulative incidence functions based on current status data with competing risks. [Hudgens et al. \(2001\)](#) and [Jewell et al. \(2003\)](#) derived and studied the nonparametric maximum likelihood approach and also introduced several other estimators, including the so-called naive estimator of [Jewell et al. \(2003\)](#). [Groeneboom et al. \(2008b,c\)](#) and M.H. Maathuis, in an unpublished 2006 University of Washington PhD thesis, available at <http://stat.ethz.ch/~maathuis/papers>, derived the large-sample behaviour of the maximum likelihood estimator and the naive estimator in a smooth model that imposes certain smoothness conditions on the cumulative incidence functions and the observation time distribution. In this model, the local rate of convergence of the maximum likelihood estimator is  $n^{1/3}$  ([Groeneboom et al., 2008b](#), Theorem 4.17), slower than the usual  $n^{1/2}$  rate. Moreover, its limiting distribution is nonstandard and involves a self-induced system of slopes of convex minorants of Brownian motion processes plus parabolic drifts ([Groeneboom et al., 2008c](#), Theorems 1.7 and 1.8). The naive estimator has the same local rate of convergence as the maximum likelihood estimator, but its limiting distribution is simpler, since it does not involve a self-induced system ([Groeneboom et al., 2008c](#), Theorem 1.6).

In practice, recorded observation times are often discrete, making the smooth model unsuitable. We therefore study the large sample behaviour of the maximum likelihood estimator and the naive estimator in two additional models: a discrete model in which the observation time distribution has discrete support, and a grouped model in which the observation times are assumed to be rounded in the recording process, yielding grouped observation times.

We show that the large sample behaviour of the estimators in the discrete model is fundamentally different from that in the smooth model: the maximum likelihood estimator and the naive estimator converge locally at rate  $n^{1/2}$ , and their limiting distributions are identical and normal. These results are related to the work of [Yu et al. \(1998\)](#), who studied the asymptotic behaviour of the maximum likelihood estimator for current status data with discrete observation times in the absence of competing risks.

The grouped model is related to the work of [Woodroffe & Zhang \(1999\)](#) and [Zhang et al. \(2001\)](#), who considered the maximum likelihood estimator for a nonincreasing density when the observations are grouped. We are not aware, however, of any work on the maximum likelihood estimator for interval censored data with grouped observation times, even though such grouping frequently occurs in practice. For example, in the menopause data the ages of the women were grouped in the intervals 25–30, 30–35, 35–36, 36–37, . . . , 58–59 and recorded as the midpoints of these intervals. The menopausal status, on the other hand, was determined at the exact but unrecorded time of interview, yielding a mismatch between the recorded status and the recorded observation time. For example, if a 30.7-year old pre-menopausal woman is interviewed, she is recorded as pre-menopausal with rounded age 32.5. When ignoring the rounding, as done in previous analyses of these data, this is taken to mean that she was interviewed at age 32.5 and that she was pre-menopausal at that age. A correct interpretation of the data is, however, that she was pre-menopausal at some unknown age in the interval (30, 35]. In particular, the data do not reveal her menopausal status at age 32.5; in actuality, she might have been post-menopausal at that age, for example due to an operation.

The grouped model accounts for such grouping of observation times. We show that the likelihood in this model can be written in the same form as in the discrete model, but in terms of different parameters, representing weighted averages of the cumulative incidence functions over the grouping intervals, where the weights are determined by the observation time distribution. This similarity with the discrete model implies that the maximum likelihood estimator and the naive estimator in the grouped model can be computed with existing software, and that their limiting distributions can be derived as in the discrete model. However, since the likelihood is written

in terms of different parameters, the estimates under the grouped model must be interpreted differently. The ideas incorporated in the grouped model can be easily extended to other forms of interval censored data.

The asymptotic results in the three models are applied to the construction of confidence intervals, a problem that has received little attention until now. In the discrete and grouped models, confidence intervals can be constructed by standard methods, for example using the bootstrap or the limiting distributions derived in this paper. In the smooth model, the nonstandard limiting behaviour of the estimators makes the construction of confidence intervals less straightforward. In this case, we advocate using likelihood ratio confidence intervals (Banerjee & Wellner, 2001) based on the naive estimator.

## 2. MODELS

### 2.1. Exact observation times

Consider the usual competing risks setting where an event can be caused by  $K$  competing risks, with  $K \in \{1, 2, \dots\}$  fixed. The random variables of interest are  $(X, Y)$ , where  $X \in \mathbb{R}$  is the time of the event of interest, and  $Y \in \{1, \dots, K\}$  is the corresponding cause. The goal is to estimate the cumulative incidence functions  $F_0 = (F_{01}, \dots, F_{0K})$ , where  $F_{0k}(t) = \text{pr}(X \leq t, Y = k)$  for  $k = 1, \dots, K$ . The cumulative incidence functions are nonnegative, monotone nondecreasing, and satisfy  $\sum_{k=1}^K F_{0k}(t) = \text{pr}(X \leq t) \leq 1$ .

The difficulty in estimating the cumulative incidence functions is that we cannot observe  $(X, Y)$  directly. Rather, we observe the current status of a subject at a single random observation time  $C \in \mathbb{R}$ . Thus, at time  $C$  we observe whether or not the event of interest has occurred, and if and only if the event has occurred, we also observe the cause  $Y$ . We assume that  $C$  is independent of  $(X, Y)$ . Let  $G$  denote the distribution of  $C$ , and let  $(C, \Delta)$  denote the observed data, where  $\Delta = (\Delta_1, \dots, \Delta_{K+1})$  is an indicator vector for the status of the subject at time  $C$ :

$$\Delta_k = 1\{X \leq C, Y = k\} \quad (k = 1, \dots, K), \quad \Delta_{K+1} = 1\{X > C\}. \tag{1}$$

To make this concrete, consider the HIV data discussed in § 1, where  $X$  is the age at HIV infection,  $C$  is the age at screening and there are  $K = 3$  competing risks representing the HIV subtypes:  $Y = 1$  for subtype B,  $Y = 2$  for subtype E and  $Y = 3$  for other subtypes.

We consider the maximum likelihood estimator for  $F_0$  based on  $n$  independent and identically distributed observations of  $(C, \Delta)$ , denoted by  $(C_i, \Delta^i)$  ( $i = 1, \dots, n$ ), where  $\Delta^i = (\Delta_1^i, \dots, \Delta_{K+1}^i)$ . For any  $K$ -tuple  $(x_1, \dots, x_K)$  let  $x_+ = \sum_{k=1}^K x_k$  and, unless otherwise defined, let  $x_{K+1} = 1 - x_+$ . Moreover, define the set  $\mathcal{F}_K = \{F = (F_1, \dots, F_K) : F_1, \dots, F_K \text{ are cumulative incidence functions and } F_+(t) \leq 1 \text{ for all } t \in \mathbb{R}\}$ . A maximum likelihood estimator for  $F_0$  is defined as any  $\hat{F}_n = (\hat{F}_{n1}, \dots, \hat{F}_{nK}) \in \mathcal{F}_K$  satisfying  $l_n(\hat{F}_n) = \max_{F \in \mathcal{F}_K} l_n(F)$ , where  $l_n(F)$  is the loglikelihood

$$l_n(F) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{K+1} \Delta_k^i \log\{F_k(C_i)\}, \tag{2}$$

with the convention  $0 \log 0 = 0$ ; see also Jewell et al. (2003), equation (1).

We also consider the naive estimator  $\tilde{F}_n = (\tilde{F}_{n1}, \dots, \tilde{F}_{nK})$  of Jewell et al. (2003), whose  $k$ th component is defined as any  $\tilde{F}_{nk} \in \mathcal{F}_1$  satisfying  $l_{nk}(\tilde{F}_{nk}) = \max_{F_k \in \mathcal{F}_1} l_{nk}(F_k)$ , where

$$l_{nk}(F_k) = \frac{1}{n} \sum_{i=1}^n [\Delta_k^i \log\{F_k(C_i)\} + (1 - \Delta_k^i) \log\{1 - F_k(C_i)\}] \tag{3}$$

is the marginal loglikelihood for the reduced current status data  $(C_i, \Delta_k^i)$  ( $i = 1, \dots, n$ ), and  $\mathcal{F}_1$  is obtained from  $\mathcal{F}_K$  by taking  $K = 1$ . Since  $\tilde{F}_{nk}$  only uses the  $k$ th entry of the  $\Delta$ -vector, the naive estimator splits the estimation problem into  $K$  well-known univariate current status problems. Therefore, its computation and asymptotic theory follow straightforwardly from known results on current status data. But this simplification comes at a cost. For example,  $\tilde{F}_{n+}$  need not be bounded by unity, and the naive estimator has been empirically shown to be less efficient than the maximum likelihood estimator in the smooth model (Groeneboom et al., 2008c).

The R-package MLEcens provides an efficient and stable method to compute the maximum likelihood estimator. This algorithm first uses the Height Map Algorithm of Maathuis (2005) to compute the areas to which the maximum likelihood estimator can possibly assign probability mass, called maximal intersections. Next, it computes the amounts of mass that must be assigned to the maximal intersections. This involves solving a high-dimensional convex optimization problem, which is done using the support reduction algorithm of Groeneboom et al. (2008a). Jewell & Kalbfleisch (2004) describe an alternative algorithm for the computation of the maximum likelihood estimator, based on the pool adjacent violators algorithm of Ayer et al. (1955).

The maximum likelihood estimator and the naive estimator are not defined uniquely at all times. Gentleman & Vandal (2002) defined two types of nonuniqueness for estimators based on censored data: mixture nonuniqueness and representational nonuniqueness. Mixture nonuniqueness occurs when the probability masses assigned to the maximal intersections are nonunique. Representational nonuniqueness refers to the fact that the estimator is indifferent to the distribution of mass within the maximal intersections. The maximum likelihood estimator for current status data with competing risks is always mixture unique, see Theorem 2.20 in the thesis by M.H. Maathuis, and mixture uniqueness of the naive estimator follows as a special case of this. One can account for representational nonuniqueness of the estimators by providing a lower bound that assigns all mass to the right endpoints of the maximal intersections, and an upper bound that assigns all mass to the left endpoints of the maximal intersections.

## 2.2. Exact observation times with discrete support

Section 2.1 does not impose any assumptions on the observation time distribution  $G$ , and hence is valid for both continuous and discrete observation times. However, the formulas can be simplified when  $G$  is discrete. In this case, let  $G(\{s\})$  denote the point mass of  $G$  at  $s$ , and let  $\mathcal{S} = \{s \in \mathbb{R} : G(\{s\}) > 0\}$  denote the support of  $G$ , where  $\mathcal{S}$  is countable but possibly infinite. Defining

$$N_k(s) = \frac{1}{n} \sum_{i=1}^n \Delta_k^i 1\{C_i = s\} \quad (k = 1, \dots, K + 1), \quad s \in \mathcal{S},$$

and  $N(s) = \sum_{k=1}^{K+1} N_k(s)$ , the loglikelihood (2) reduces to

$$l_n(F) = \sum_{s \in \mathcal{S}} \sum_{k=1}^{K+1} N_k(s) \log\{F_k(s)\}, \quad (4)$$

and the marginal loglikelihood (3) for the naive estimator becomes

$$l_{nk}(F_k) = \sum_{s \in \mathcal{S}} [N_k(s) \log\{F_k(s)\} + \{N(s) - N_k(s)\} \log\{1 - F_k(s)\}].$$

The spaces  $\mathcal{F}_K$  and  $\mathcal{F}_1$  can also be simplified, as the nonnegativity, monotonicity and boundedness constraints only need to hold at points  $s \in \mathcal{S}$ .

2.3. Grouped observation times

In many applications, only rounded versions of the observation times are recorded, yielding grouped observation times. We introduce a new model for this type of data, called the grouped model. For any interval  $I$  on the real line, define  $G(I) = \int_{c \in I} dG(c)$ . Let  $\mathcal{I}$  be a countable but possibly infinite set of mutually exclusive intervals such that  $G(I) > 0$  for all  $I \in \mathcal{I}$ . For each  $I \in \mathcal{I}$ , let  $m(I)$  denote a unique point in the interval, for example its midpoint, and let  $\mathcal{M} = \{m(I) \in \mathbb{R} : I \in \mathcal{I}\}$ . For each  $m \in \mathcal{M}$ , let  $I(m)$  denote the corresponding interval in  $\mathcal{I}$ .

The observation scheme in the grouped model is as follows. As before, the current status of a subject is assessed at a single random time  $C \in \mathbb{R}$ , where  $C$  is independent of  $(X, Y)$ . The difference is that we no longer observe  $C$ . Instead, all observation times falling into interval  $I$  are grouped and rounded to  $m(I)$ . Thus, the observed data are  $(D, \Delta)$ , where  $D = \sum_{I \in \mathcal{I}} m(I) 1\{C \in I\}$  is the rounded version of  $C$ , and  $\Delta$  is the indicator vector corresponding to the status of the subject at the exact time  $C$ , as defined in (1). We study the maximum likelihood estimator and the naive estimator based on  $n$  independent and identically distributed observations of  $(D, \Delta)$ , which we denote by  $(D_i, \Delta^i)$  ( $i = 1, \dots, n$ ).

To derive the likelihood in the grouped model, we compute  $\text{pr}(D = d, \Delta = \delta)$  for  $d \in \mathcal{M}$  and  $\delta \in \{e_1, \dots, e_{K+1}\}$ , where  $e_k$  is the unit vector in  $\mathbb{R}^{K+1}$  with a 1 at the  $k$ th entry. Conditioning on the exact observation time  $C$  yields

$$\begin{aligned} \text{pr}(D = d, \Delta = \delta) &= \int \text{pr}(D = d, \Delta = \delta \mid C = c) dG(c) = \int_{c \in I(d)} \text{pr}(\Delta = \delta \mid C = c) dG(c) \\ &= \prod_{k=1}^{K+1} \left\{ \int_{c \in I(d)} F_{0k}(c) dG(c) \right\}^{\delta_k} = G\{I(d)\} \prod_{k=1}^{K+1} [H_{0k}\{I(d)\}]^{\delta_k}, \end{aligned} \tag{5}$$

where

$$H_{0k}\{I(d)\} = [G\{I(d)\}]^{-1} \int_{c \in I(d)} F_{0k}(c) dG(c) \quad (k = 1, \dots, K),$$

and  $H_{0,K+1}\{I(d)\} = 1 - H_{0+}\{I(d)\}$  are weighted averages of  $F_{01}, \dots, F_{0,K+1}$  over  $I(d)$  with weights determined by  $G$ . It is convenient to work with these weighted averages, as they must obey the same constraints as the cumulative incidence functions. More precisely, considering  $H_{0k}$  ( $k = 1, \dots, K$ ), as functions that maps  $m$  to  $H_{0k}\{I(m)\}$ , the constraints on  $F_{01}, \dots, F_{0K}$  imply that  $H_{01}, \dots, H_{0K}$  must be nonnegative and nondecreasing and satisfy  $H_{0+}\{I(m)\} \leq 1$  for all  $m \in \mathcal{M}$ . Let  $\mathcal{H}_K$  denote the space of such allowable  $K$ -tuples  $(H_1, \dots, H_K)$ .

The term  $G\{I(d)\}$  in the right-hand side of (5) can be dropped from the likelihood, as it does not depend on  $F$ . Hence, a maximum likelihood estimator for  $H_0 = (H_{01}, \dots, H_{0K})$  is defined as any  $\hat{H}_n \in \mathcal{H}_K$  satisfying  $l_n^{\text{group}}(\hat{H}_n) = \max_{H \in \mathcal{H}_K} l_n^{\text{group}}(H)$ , where

$$l_n^{\text{group}}(H) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{K+1} \Delta_k^i \log[H_k\{I(D_i)\}]. \tag{6}$$

Expression (6) has the same form as (2), but with  $F_k(C_i)$  replaced by the weighted average  $H_k\{I(D_i)\}$ . As in the discrete model, (6) can be simplified further:

$$l_n^{\text{group}}(H) = \sum_{I \in \mathcal{I}} \sum_{k=1}^{K+1} M_k(I) \log\{H_k(I)\}, \tag{7}$$

where

$$M_k(I) = \frac{1}{n} \sum_{i=1}^n \Delta_k^i 1\{D_i = m(I)\} \quad (k = 1, \dots, K + 1), \quad I \in \mathcal{I}.$$

Since the loglikelihood (7) has the same form as (4), and also the constraints on the maximization problems for the discrete and grouped models are equivalent, the maximum likelihood estimator in the grouped model can be computed with existing software. Moreover, its asymptotic theory follows straightforwardly from the theory for the discrete model. The important difference between the two models is, however, that the resulting estimates must be interpreted differently. In the discrete model, one estimates the cumulative incidence functions at points  $s \in \mathcal{S}$ . In the grouped model, the cumulative incidence functions are unidentifiable in general, and one estimates the weighted averages of the cumulative incidence functions over intervals  $I \in \mathcal{I}$ .

The naive estimator  $\tilde{H}_n$  in the grouped model can be derived analogously. Defining  $M(I) = \sum_{k=1}^{K+1} M_k(I)$ ,  $I \in \mathcal{I}$ , the marginal loglikelihood for the  $k$ th component is

$$l_{nk}^{\text{group}}(H_k) = \sum_{I \in \mathcal{I}} [M_k(I) \log\{H_k(I)\} + \{M(I) - M_k(I)\} \log\{1 - H_k(I)\}],$$

and  $\tilde{H}_{nk} \in \mathcal{H}_1$  is defined by  $l_{nk}^{\text{group}}(\tilde{H}_{nk}) = \max_{H_k \in \mathcal{H}_1} l_{nk}^{\text{group}}(H_k)$ .

*Remark 1.* In general,  $F_{0k}(m) \neq H_{0k}\{I(m)\}$ , but equality can occur in special situations. For example,  $F_{0k}(m) = H_{0k}\{I(m)\}$  if

- (i)  $F_{0k}$  is constant on  $I(m)$ , or
- (ii) both  $F_{0k}$  and  $G$  are linear on  $I(m)$  and  $m$  is the midpoint of  $I(m)$ , or
- (iii) the only mass of  $G$  on  $I(m)$  consists of a point mass at  $m$ .

Condition (iii) shows that the grouped model generalizes the discrete model.

### 3. LOCAL ASYMPTOTICS OF THE ESTIMATORS

#### 3.1. Strong consistency in the discrete and grouped models

The maximum likelihood estimator and the naive estimator are Hellinger consistent when the observation times are recorded exactly for any observation time distribution  $G$ , see Theorem 4.6 in the thesis by M. H. Maathuis. Using the equivalence between Hellinger distance and total variation distance, this implies consistency in total variation, see Corollary 4.7 in the thesis by M. H. Maathuis, which in turn implies strong pointwise consistency at all points  $s \in \mathcal{S}$  in the discrete model, as stated in Theorem 1.

**THEOREM 1.** *In the discrete model,  $\hat{F}_{nk}(s) \rightarrow F_{0k}(s)$  and  $\tilde{F}_{nk}(s) \rightarrow F_{0k}(s)$  almost surely as  $n \rightarrow \infty$  for all  $s \in \mathcal{S}$ .*

Since the form of the loglikelihood and the constraints on the allowable functions are identical in the discrete and grouped models, the proofs for the discrete model carry over directly to the grouped model. This leads to Theorem 2, which we give without proof.

**THEOREM 2.** *In the grouped model,  $\hat{H}_{nk}(I) \rightarrow H_{0k}(I)$  and  $\tilde{H}_{nk}(I) \rightarrow H_{0k}(I)$  almost surely as  $n \rightarrow \infty$  for all  $I \in \mathcal{I}$ .*

3.2. Limiting distributions in the discrete model

Denote the infimum and supremum of  $\mathcal{S}$  by  $s_{\text{inf}}$  and  $s_{\text{sup}}$ , and define

$$s_- = \sup\{x \in \mathcal{S} : x < s\} \quad \text{for } s \in \mathcal{S} \text{ with } s \neq s_{\text{inf}},$$

$$s_+ = \inf\{x \in \mathcal{S} : x < s\} \quad \text{for } s \in \mathcal{S} \text{ with } s \neq s_{\text{sup}}.$$

Define  $s \in \mathcal{S}$  to be a regular point if  $F_{0k}(s) = 0$  for all  $k = 1, \dots, K$  or the following two conditions hold.

*Condition 1.* If  $s \neq s_{\text{inf}}$  then  $s_- \in \mathcal{S}$  and for each  $k = 1, \dots, K$  either  $F_{0k}(s_-) < F_{0k}(s)$  or  $F_{0k}(s) = 0$ .

*Condition 2.* If  $s \neq s_{\text{sup}}$  then  $s_+ \in \mathcal{S}$  and for each  $k = 1, \dots, K$  either  $F_{0k}(s) < F_{0k}(s_+)$  or  $F_{0k}(s) = 0$ .

If  $\mathcal{S}$  is a finite set and  $s \in \mathcal{S} \setminus \{s_{\text{inf}}, s_{\text{sup}}\}$ , then  $s_-$  and  $s_+$  are simply the points directly to the left and right of  $s$ , and Conditions 1 and 2 are equivalent to requiring that for each  $k = 1, \dots, K$  either  $F_{0k}(s_-) < F_{0k}(s) < F_{0k}(s_+)$  or  $F_{0k}(s) = 0$ . As a second example, suppose that  $\mathcal{S}$  is the set of rational numbers. Then for any point  $s \in \mathcal{S}$  we have  $s \notin \{s_{\text{inf}}, s_{\text{sup}}\}$  and  $s_- = s = s_+$ . Hence, Conditions 1 and 2 are only satisfied if  $F_{0k}(s) = 0$  for all  $k = 1, \dots, K$ . [Yu et al. \(1998\)](#) introduced regular points in the current status model without competing risks. Our definition generalizes theirs by allowing for competing risks. Moreover, we allow the parameters to be on the boundary of the parameter space. For example,  $s \in \mathcal{S}$  can be a regular point when  $F_{0k}(s) = 0$  for some or all of the  $F_{0k}$ s, and  $s = s_{\text{sup}}$  can be a regular point when  $\sum_{k=1}^K F_{0k}(s) = 1$  or when  $F_{0k}(s) = \lim_{t \rightarrow \infty} F_{0k}(t)$  for some of the  $F_{0k}$ s.

We now introduce the following simple estimator for  $F_{0k}(s)$ :

$$\check{F}_{nk}(s) = N_k(s)/N(s) \quad (k = 1, \dots, K), \quad s \in \mathcal{S},$$

where we set  $0/0 = 0$ . This estimator is very simple, in the sense that  $\check{F}_{nk}$  does not obey monotonicity constraints and uses only the  $k$ th component of the  $\Delta$ -vector. Lemma 1 below states that  $\check{F}_n$  is the maximum likelihood estimator for  $F_0$  if the monotonicity constraints on the cumulative incidence functions are discarded. Next, Lemma 2 establishes that for any regular point  $s \in \mathcal{S}$ ,  $\hat{F}_n(s) = \tilde{F}_n(s) = \check{F}_n(s)$  with probability tending to one as  $n \rightarrow \infty$ . Hence, at such points the limiting distributions of  $\hat{F}_n(s)$  and  $\tilde{F}_n(s)$  equal the limiting distribution of  $\check{F}_n(s)$ . This yields asymptotic normality of  $\hat{F}_n(s)$  and  $\tilde{F}_n(s)$  at regular points, as stated in Theorem 3. All proofs are deferred to the Appendix.

LEMMA 1. Let  $\mathcal{F}_K^* = \{F = (F_1, \dots, F_K) : F_k(s) \geq 0 \text{ for } k = 1, \dots, K \text{ and } F_+(s) \leq 1 \text{ for all } s \in \mathcal{S}\}$ . Then  $l_n(\check{F}_n) \geq l_n(F)$  for all  $F \in \mathcal{F}_K^*$ , and  $l_n(\check{F}_n) > l_n(F)$  for all  $F \in \mathcal{F}_K^*$  such that  $F(s) \neq \check{F}_n(s)$  for some  $s \in \mathcal{S}$  with  $N(s) > 0$ .

LEMMA 2. For any regular point  $s \in \mathcal{S}$  in the discrete model,  $\text{pr}\{\hat{F}_n(s) = \tilde{F}_n(s) = \check{F}_n(s)\} \rightarrow 1$ , as  $n \rightarrow \infty$ .

THEOREM 3. For any regular point  $s \in \mathcal{S}$  in the discrete model,

$$n^{1/2}\{\hat{F}_n(s) - F_0(s)\} = n^{1/2} \begin{pmatrix} \hat{F}_{n1}(s) - F_{01}(s) \\ \vdots \\ \hat{F}_{nK}(s) - F_{0K}(s) \end{pmatrix}$$

is asymptotically normal with zero mean and covariance matrix  $V(s)$ , where  $V(s)$  is a  $K \times K$  matrix with entries

$$\{V(s)\}_{k,\ell} = [F_{0k}(s)1\{k = \ell\} - F_{0k}(s)F_{0\ell}(s)] / G(\{s\}), \quad k, \ell \in \{1, \dots, K\}.$$

For any finite collection of regular points  $s_1, \dots, s_p$  in  $\mathcal{S}$ , the stacked vector  $n^{1/2}\{\hat{F}_n(s_1) - F_0(s_1), \dots, \hat{F}_n(s_p) - F_0(s_p)\}$  is asymptotically normal with zero mean and block diagonal covariance matrix with blocks  $V(s_1), \dots, V(s_p)$ . Consistent estimators for the elements of  $V(s)$ ,  $s \in \mathcal{S}$ , are

$$\{\hat{V}_n(s)\}_{k,\ell} = [\hat{F}_{nk}(s)1\{k = \ell\} - \hat{F}_{nk}(s)\hat{F}_{n\ell}(s)] / N(s), \quad k, \ell \in \{1, \dots, K\}.$$

The same results hold for the naive estimator, that is, when  $\hat{F}_n$  is replaced by  $\tilde{F}_n$ .

*Remark 2.* If  $F_{0k}(s) > 0$  for all  $k = 1, \dots, K$  and  $\sum_{k=1}^K F_{0k}(s) = 1$ , then the matrix  $V(s)$  is positive-semidefinite with rank  $K - 1$ . If  $F_{0k}(s) = 0$  or  $F_{0k}(s) = 1$ , then the  $k$ th row and the  $k$ th column of  $V(s)$  are zero vectors, and the corresponding limiting distributions of  $\hat{F}_{nk}(s)$  and  $\tilde{F}_{nk}(s)$  should be interpreted as degenerate distributions consisting of a point mass at zero. More details can be found in the proof of Theorem 3.

### 3.3. Limiting distributions in the grouped model

Denote the infimum and supremum of  $\mathcal{M}$  by  $m_{\inf}$  and  $m_{\sup}$ , and define

$$\begin{aligned} \{m(I)\}_- &= \sup\{x \in \mathcal{M} : x < m(I)\} \quad \text{for } I \in \mathcal{I} \text{ with } m(I) \neq m_{\inf}, \\ \{m(I)\}_+ &= \inf\{x \in \mathcal{M} : x > m(I)\} \quad \text{for } I \in \mathcal{I} \text{ with } m(I) \neq m_{\sup}. \end{aligned}$$

If  $\{m(I)\}_- \in \mathcal{M}$  let  $I_- = I[\{m(I)\}_-]$ , and if  $\{m(I)\}_+ \in \mathcal{M}$  let  $I_+ = I[\{m(I)\}_+]$ . We say that  $I \in \mathcal{I}$  is a regular interval if  $H_{0k}(I) = 0$  for all  $k = 1, \dots, K$  or the following two conditions hold: (i) if  $m(I) \neq m_{\inf}$  then  $\{m(I)\}_- \in \mathcal{M}$  and for each  $k = 1, \dots, K$  either  $H_{0k}(I_-) < H_{0k}(I)$  or  $H_{0k}(I) = 0$ , and (ii) if  $m(I) \neq m_{\sup}$  then  $\{m(I)\}_+ \in \mathcal{M}$  and for each  $k = 1, \dots, K$  either  $H_{0k}(I) < H_{0k}(I_+)$  or  $H_{0k}(I) = 0$ .

Analogously to  $\tilde{F}_n$  in the discrete model, we define a simple estimator in the grouped model:

$$\check{H}_{nk}(I) = M_k(I) / M(I) \quad (k = 1, \dots, K), \quad I \in \mathcal{I}.$$

The proofs and results for the discrete model can now be translated directly to the grouped model, by replacing regular points  $s \in \mathcal{S}$  by regular intervals  $I \in \mathcal{I}$ ,  $\hat{F}_n(s)$  by  $\hat{H}_n(I)$ ,  $\tilde{F}_n(s)$  by  $\check{H}_n(I)$ ,  $\tilde{F}_n(s)$  by  $\check{H}_n(I)$ ,  $F_0(s)$  by  $H_0(I)$  and  $N_k(s)$  by  $M_k(I)$  for  $k = 1, \dots, K + 1$ . We therefore only give the main result in Theorem 4, without proof.

**THEOREM 4.** For any regular interval  $I \in \mathcal{I}$  in the grouped model,

$$n^{1/2}\{\hat{H}_n(I) - H_0(I)\} = n^{1/2} \begin{pmatrix} \hat{H}_{n1}(I) - H_{01}(I) \\ \vdots \\ \hat{H}_{nK}(I) - H_{0K}(I) \end{pmatrix}$$

is asymptotically normal with zero mean and covariance matrix  $U(I)$ , where  $U(I)$  is a  $K \times K$  matrix with entries

$$\{U(I)\}_{k,\ell} = [H_{0k}(I)1\{k = \ell\} - H_{0k}(I)H_{0\ell}(I)] / G(I), \quad k, \ell \in \{1, \dots, K\}.$$



Moreover, for any finite collection of regular intervals  $I_1, \dots, I_p$ , the stacked vector  $n^{1/2}\{\hat{H}_n(I_1) - H_0(I_1), \dots, \hat{H}_n(I_p) - H_0(I_p)\}$  is asymptotically normal with mean vector zero and block diagonal covariance matrix with blocks  $U(I_1), \dots, U(I_p)$ . Consistent estimators for the elements of  $U(I)$ ,  $I \in \mathcal{I}$ , are

$$\{\hat{U}_n(I)\}_{k,\ell} = [\hat{H}_{nk}(I)1\{k = \ell\} - \hat{H}_{nk}(I)\hat{H}_{n\ell}(I)]/M(I), \quad k, \ell \in \{1, \dots, K\}.$$

The same results hold for the naive estimator, that is, when  $\hat{H}_n$  is replaced by  $\tilde{H}_n$ .

As in Theorem 3, a degenerate limiting distribution should be interpreted as point mass at zero.

### 3.4. Theoretical motivation for the grouped model

The asymptotic results provide a theoretical motivation for the grouped model, since a contradiction arises with respect to rates of convergence when the grouping of observation times is ignored. To see this, consider the menopause data and the HIV data, and suppose that the grouping of observation times is ignored, meaning that the recorded observation times are interpreted as exact observation times. This assumption was made in previous analyses of the menopause data, see [Krailo & Pike \(1983\)](#); [Jewell et al. \(2003\)](#); [Jewell & Kalbfleisch \(2004\)](#). Under this assumption, the discrete model is most appropriate for the menopause data, since there are numerous ties in the recorded observation times, see § 5.2. On the other hand, the smooth model seems most appropriate for the HIV data, since this dataset contains very few ties in the recorded observation times, see § 5.3. This would imply that the local rate of convergence of the maximum likelihood estimator and the naive estimator at the recorded observation times is  $n^{1/2}$  for the menopause data, while it is  $n^{1/3}$  for the HIV data.

In reality, however, the observation times were continuous in both datasets, and they were rounded in the recording process. In the menopause data, this rounding was substantial, into one-year or five-year intervals, while in the HIV data it was minimal, into one-day intervals. Since rounding implies discarding information, it seems impossible that more rounding, as in the menopause data, leads to a faster local rate of convergence at the recorded observation times. This apparent contradiction can be resolved by modelling the grouping of the observation times. For the grouped model, rounding or grouping of the observation times indeed yields a faster rate of convergence, but not for the cumulative incidence functions at the recorded observation times, but for weighted averages of the cumulative incidence functions over the grid cells. These weighted averages are smooth functionals of the cumulative incidence functions and thus can be estimated at rate  $n^{1/2}$ , see [Jewell et al. \(2003\)](#) and Chapter 7 in the thesis by M. H. Maathuis.

## 4. CONSTRUCTION OF POINTWISE CONFIDENCE INTERVALS

### 4.1. Confidence intervals in the discrete and grouped models

In the discrete and grouped models, the large-sample behaviour of the maximum likelihood estimator and the naive estimator at regular points or intervals is standard, and hence confidence intervals can be constructed by any standard method, for example using the asymptotic normal distribution or the bootstrap. For instance, let  $s \in \mathcal{S}$  be a regular point in the discrete model. Then an asymptotic  $(1 - \alpha)100\%$  confidence interval for  $F_{0k}(s)$  is

$$\hat{F}_{nk}(s) \pm n^{-1/2}z_{1-\alpha/2}[\{\hat{V}_n(s)\}_{k,k}]^{1/2},$$

where  $z_{1-\alpha/2}$  is the  $(1 - \alpha/2)$ -quantile of the standard normal distribution. Similarly, considering a regular interval  $I \in \mathcal{I}$  in the grouped model, an asymptotic  $(1 - \alpha)100\%$  confidence interval

for  $H_{0k}(I)$  is

$$\hat{H}_{nk}(I) \pm n^{-1/2} z_{1-\alpha/2} [\{\hat{U}_n(I)\}_{k,k}]^{1/2}. \quad (8)$$

#### 4.2. Confidence intervals in the smooth model

In the smooth model, the large-sample behaviour of the maximum likelihood estimator and the naive estimator is nonstandard, making the construction of confidence intervals less straightforward. In principle, one can construct confidence intervals using the limiting distribution of the maximum likelihood estimator, but this approach entails several difficulties. First, the limiting distribution involves parameters from the underlying distributions that need to be estimated. Moreover, Theorems 1.7 and 1.8 of [Groeneboom et al. \(2008c\)](#) suggest that these parameters cannot be separated from the limiting distribution, in the sense that it seems impossible to write the limiting distribution as  $cZ$ , where  $c$  is a constant depending on the underlying distribution and  $Z$  is a universal limit. Hence, one would need to simulate the limiting distribution on a case by case basis. Conducting such simulations is nontrivial, see [Groeneboom & Wellner \(2001\)](#).

One might also consider the nonparametric bootstrap to construct confidence intervals based on the maximum likelihood estimator or the naive estimator. However, it is likely that the bootstrap is inconsistent in this setting, given recent results of [Kosorok \(2008\)](#) and [Sen et al. \(2010\)](#) on inconsistency of the bootstrap for the closely related Grenander estimator.

Subsampling ([Politis & Romano, 1994](#)), a variant of the bootstrap, can be applied to construct asymptotically valid confidence intervals for the cumulative incidence functions based on the maximum likelihood estimator or the naive estimator. A drawback of subsampling is that it requires a tuning parameter, the subsample size, which is difficult to choose in practice.

Finally, one can consider likelihood ratio confidence intervals based on the naive estimator. Although the naive estimator has been shown empirically to be less efficient than the maximum likelihood estimator ([Groeneboom et al., 2008c](#), Fig. 3), it has the advantage that its large sample behaviour is simpler. For a fixed failure cause, the limiting distribution of the naive estimator is identical to the limiting distribution of the maximum likelihood estimator for current status data without competing risks ([Groeneboom et al., 2008c](#), Theorem 1.6). Hence, the likelihood ratio theory of [Banerjee & Wellner \(2001\)](#) applies, and confidence intervals can be constructed by inverting likelihood ratio tests ([Banerjee & Wellner, 2005](#)). These confidence intervals have the appealing property that they do not require estimation of parameters from the underlying distribution, nor any tuning parameters. Simulation studies by [Banerjee & Wellner \(2005\)](#) showed that for current status data without competing risks, likelihood ratio based confidence intervals are typically preferable over confidence intervals based on the limiting distribution or subsampling. In the smooth model, we therefore recommend using likelihood ratio confidence intervals based on the naive estimator.

## 5. EXAMPLES

### 5.1. Simulation

It is not clear how well the asymptotic distributions of § 3.2 and § 3.3 approximate the finite sample behaviour of the estimators, especially for grids that are dense relative to  $n$ . We therefore conducted a simulation study, using the following discrete model:  $\text{pr}(Y = 1) = 0.6$ ,  $\text{pr}(Y = 2) = 0.4$ ,  $X | Y = 1 \sim \text{Ga}(5, 3)$ , and  $X | Y = 2 \sim \text{Ga}(9, 2)$  where  $\text{Ga}(a, b)$  denotes a Gamma distribution with shape parameter  $a$  and scale parameter  $b$ . The distribution of  $C$  was taken to be uniform on one of the following grids: (i)  $\{10, 20, 30\}$ , called gap 10, (ii)  $\{6, 8, \dots, 34\}$ , called gap 2, (iii)  $\{5.5, 6.0, \dots, 35.0\}$ , called gap 0.5 and (iv)  $\{5.1, 5.2, \dots, 35.0\}$ , called gap 0.1. For each of the four resulting models, 1000 datasets of sample size  $n = 1000$  were simulated. For

each dataset we computed symmetric 95% asymptotic confidence intervals for the cumulative incidence functions at the points  $t_0 = (10, 20, 30)$ , based on the normal distribution and the bootstrap, using both the maximum likelihood estimator and the naive estimator.

The results for  $F_{01}$  are shown in Fig. 1. The results for  $F_{02}$  are similar, and are therefore omitted. Confidence intervals based on the maximum likelihood estimator behave very similarly to confidence intervals based on the naive estimator, while there is a large difference between normal and bootstrap based confidence intervals for the denser grids. The increase in width of the normal based confidence intervals for the denser grids is caused by the decrease of  $nG(\{t_0\})$ , which can be viewed as the expected effective sample size for the simple estimator  $\tilde{F}_n$  at  $t_0$ . As a result, the variance of the asymptotic normal distribution increases by a factor of 5 or 6 between each pair of successive grids. The empirical variance of the estimators, on the other hand, increases somewhat for the denser grids, but not by much, due to the stabilizing effect of the monotonization that takes place in the maximum likelihood estimator and the naive estimator. As a result, the normal based confidence intervals give substantial overcoverage. This breakdown of the normal limit is already apparent for the larger time-points in the relatively coarse grid, gap 2, which has an average of 67 observation times per grid point. The bootstrap variance was found to be a better approximation of the empirical variance of the estimators, suggesting the use of bootstrap intervals over asymptotic normal intervals in practice. However, the undercoverage of the bootstrap intervals at  $t_0 = 10$  becomes more substantial as the grids become denser. This points to inconsistency of the bootstrap for very dense grids, which is in line with the theory discussed in § 4.2.

### 5.2. Menopause data

We consider data on 2423 women in the age range 25–59 years from Cycle I of the Health Examination Survey of the National Center for Health Statistics (MacMahon & Worcester, 1966). Among other things, these women were asked to report: their current age; whether they were pre- or postmenopausal; and if they were postmenopausal, the age and cause of menopause, where the cause could be natural or operative. Since MacMahon & Worcester (1966) found marked terminal digit clustering in the reported ages of menopause, Krailo & Pike (1983) excluded these from the analysis. The remaining information can be viewed as current status data with competing risks. Nonparametric estimates of the cumulative incidences of the two types of menopause were computed by Jewell et al. (2003) and Jewell & Kalbfleisch (2004) under the assumption that the recorded ages of the women at the time of the interview were exact. However, this was not the case. Instead, the ages were grouped into the intervals 25–30, 30–35, 35–36, 36–37 . . . , 58–59 and recorded as the midpoints of these intervals, yielding 26 age groups with a minimum of 45 and an average of 93 observations per age group. This is comparable to gap 2 in our simulation study, see § 5.1.

We add to the previous analyses of these data in two ways: we use the grouped model, which is clearly appropriate for these data, and we provide confidence intervals. Figure 2 shows the maximum likelihood estimator and the naive estimator for the weighted averages of the cumulative incidence functions, together with 95% normal and bootstrap confidence intervals based on the maximum likelihood estimator. As in our simulation study, the confidence intervals based on the normal distribution are wider than those based on the bootstrap.

### 5.3. HIV data

The Bangkok Metropolitan Administration injecting drug users cohort study (Kitayaporn et al., 1998; Vanichseni et al., 2001) was established in 1995 to better understand HIV transmission and to assess the feasibility of conducting a phase III HIV vaccine efficacy

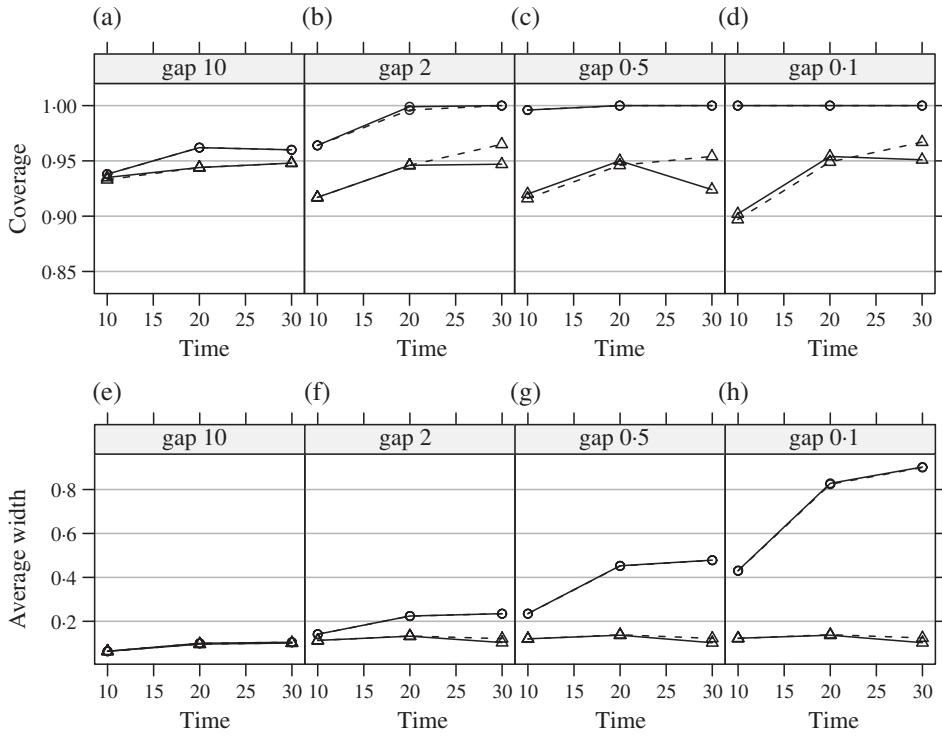


Fig. 1. Simulation: Coverage (a)–(d) and average width (e)–(h) of the four 95% confidence intervals for  $F_{01}(t_0)$  as a function of  $t_0$ , corresponding to grids gap 10 {(a) and (e)}, gap 2 {(b) and (f)}, gap 0.5 {(c) and (g)}, and gap 0.1 {(d) and (h)}. The confidence intervals were based on the normal distribution ( $\circ$ ) and the bootstrap ( $\Delta$ ), using the maximum likelihood estimator (solid line) and the naive estimator (dashed line). The bootstrap confidence intervals are based on 750 bootstrap samples.

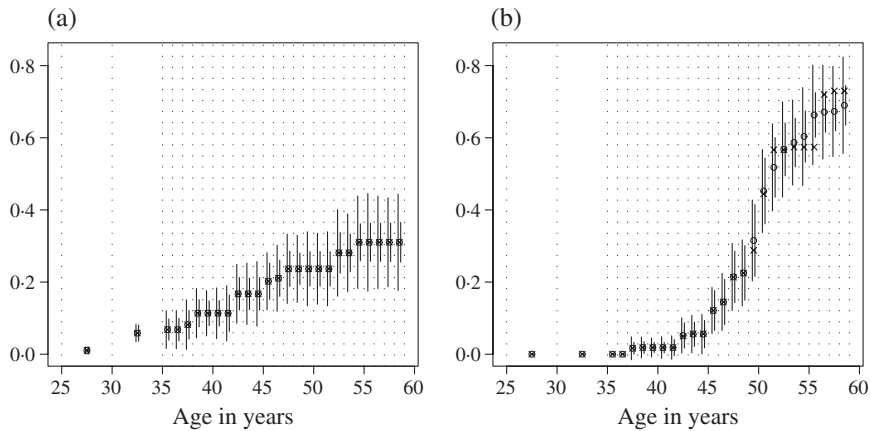


Fig. 2. Menopause data. The maximum likelihood estimator  $\hat{H}_n$  ( $\circ$ ) and the naive estimator  $\hat{H}_n$  ( $\times$ ) for the weighted averages of the cumulative incidence of (a) operative and (b) natural menopause over the age groups. The estimators are plotted at the midpoints of the age groups which are indicated by the dotted vertical lines. The two solid vertical line segments in each age group are 95% asymptotic confidence intervals based on the maximum likelihood estimator: the left line segment is based on the normal approximation (8) and the right line segment is a symmetric bootstrap confidence interval based on 1000 bootstrap samples.

trial in an injecting drug users population in Bangkok. We consider data on 1366 injecting drug users in this study who were screened from May to December 1996 and who were under 35 years of age. Among this group, 393 were HIV positive, with 114 infected with subtype B, 238 infected with subtype E, 5 infected by another or mixed subtype and 36 infected with missing subtype. The subjects with other, mixed or missing subtypes were grouped in a remainder category. All ages were recorded in days, leading to a small number of ties: among the 1366 subjects, there were 1212 distinct ages, and the mean number of observations per distinct age was 1.13. In light of this, we analyse these data using the smooth model. Figure 3 shows the maximum likelihood estimator and the naive estimator for the subtype-specific cumulative incidence of HIV, together with 95% likelihood ratio confidence intervals based on the naive estimator.

## 6. OBSERVATION TIME DISTRIBUTION OR GROUPING DEPENDENT ON $n$

There are interesting connections between our work and a 2011 unpublished technical report by Tang, Banerjee and Kosorok, available at <http://www.stat.lsa.umich.edu/~moulib/information.html>, who studied current status data without competing risks when the observation time distribution depends on the sample size  $n$ . More precisely, let  $X$  be a random event time with distribution  $F_0$  and let  $C^{(n)}$  be a random observation time with distribution  $G^{(n)}$ , where  $G^{(n)}$  is a discrete distribution on an equidistant grid with spacings  $n^{-\gamma}$  for some  $\gamma \in (0, 1)$ . Without loss of generality, assume this grid is on  $[0, 1]$ . Consider the nonparametric maximum likelihood estimator  $\hat{F}_n$  for  $F_0$  based on  $n$  independent and identically distributed observations of  $(C^{(n)}, \Lambda^{(n)})$ , where  $\Lambda^{(n)} = 1\{X \leq C^{(n)}\}$ . Let  $t_0 \in (0, 1)$  be a time-point of interest, and let  $t_n$  be the largest support point of  $G^{(n)}$  smaller than  $t_0$ . Assuming  $F_0$  satisfies certain smoothness conditions in a neighbourhood of  $t_0$ , Tang, Banerjee and Kosorok found that the limiting distribution of the maximum likelihood estimator depends crucially on  $\gamma$ . For  $\gamma < 1/3$  the limiting distribution of  $n^{(1-\gamma)/2}\{\hat{F}_n(t_n) - F_0(t_0)\}$  is normal with zero mean and variance  $F_0(t_0)\{1 - F_0(t_0)\}$ . Hence, for such sparse grids, the maximum likelihood estimator behaves as in the discrete model, up to a different rate of convergence. For  $\gamma > 1/3$ , on the other hand, the limiting distribution of  $n^{1/3}\{\hat{F}_n(t_0) - F_0(t_0)\}$  is determined by the slope of the convex minorant of a Brownian motion process plus parabolic drift, showing that the maximum likelihood estimator behaves as in the smooth model. The case  $\gamma = 1/3$  forms the boundary between these two scenarios and yields a new limiting distribution.

Combining our work with that of Tang, Banerjee and Kosorok yields two extensions. First, consider a grouped model for current status data without competing risks, where the grouping intervals depend on  $n$ . More precisely, let  $X$  be an event time with distribution  $F_0$ , let  $C$  be an observation time with distribution  $G$ , and let  $\Lambda = 1\{X \leq C\}$ . Assume the support of  $G$  is  $[0, 1]$ , and let  $\mathcal{I}_n$  be the set of intervals formed by the grid cells of an equidistant grid on  $[0, 1]$  with spacings  $n^{-\gamma}$  for some  $\gamma \in (0, 1)$ . Assume that the observation time  $C$  is rounded to the midpoint of the interval in which it falls, and denote this rounded observation time by  $D^{(n)}$ . One can now consider the nonparametric maximum likelihood estimator for  $F_0$  based on  $n$  independent and identically distributed copies of  $(D^{(n)}, \Lambda)$ . Since the likelihood in this grouped model can be written in exactly the same form as the likelihood in the discrete model, and since also the constraints on the two optimization problems are equivalent, the work of Tang, Banerjee and Kosorok should carry over to this model, with the only difference that everything should be written in terms of weighted averages of  $F_0$  over the grid cells. Second, consider the discrete model for current status data with competing risks, where the support of  $G$  depends on  $n$ . Then the results of Tang, Banerjee and Kosorok should carry over to the naive estimator  $\tilde{F}_{nk}$ , since this estimator can be viewed as a maximum likelihood estimator based on reduced current status

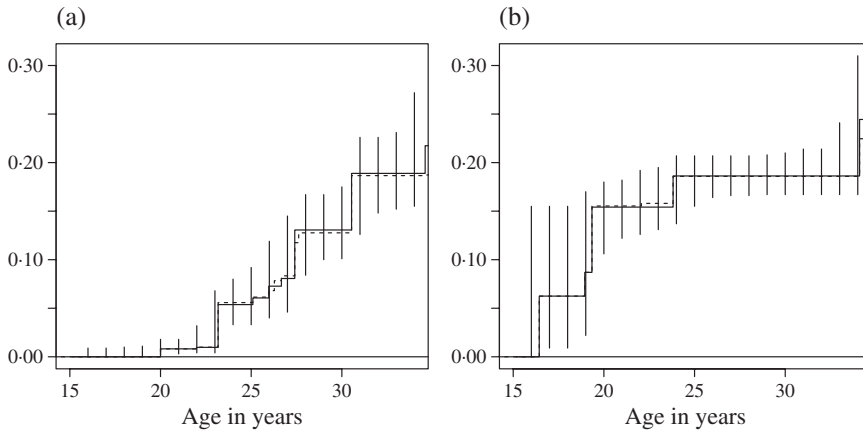


Fig. 3. HIV data. The maximum likelihood estimator  $\hat{F}_n$  (dashed) and the naive estimator  $\check{F}_n$  (solid) for the cumulative incidence of HIV subtypes (a) B and (b) E as a function of age, using the smooth model. The solid vertical lines represent 95% pointwise confidence intervals at times 16, . . . , 34, based on the likelihood ratio method for the naive estimator.

data without competing risks. The same holds for the naive estimator  $\check{H}_{nk}$  in the grouped model when the grouping intervals depend on  $n$ .

ACKNOWLEDGEMENT

We thank Kachit Choopanya, Dwip Kitayaporn, Timothy D. Mastro, Philip A. Mock and Suphak Vanichseni for allowing us to use the data from the Bangkok Metropolitan Administration Injecting Drug Users cohort study. We thank Moulinath Banerjee for discussing the connections between this paper and his work with Runlong Tang and Michael Kosorok. We thank the associate editor and two anonymous referees for helpful suggestions and comments. M.G. Hudgens was supported by a grant from the National Institutes of Health, U.S.A.

APPENDIX

*Proof of Lemma 1.* Due to the absence of monotonicity constraints on  $\mathcal{F}_K^*$ , the maximizer of  $l_n(F)$  over  $\mathcal{F}_K^*$  can be determined separately for each  $s \in \mathcal{S}$ . Thus, fix  $s \in \mathcal{S}$ , and define  $l_n(F, s) = \sum_{k=1}^{K+1} N_k(s) \log\{F_k(s)\}$ . Moreover, define  $\mathcal{K} = \{k \in \{1, \dots, K + 1\} : N_k(s) > 0\}$  and  $\mathcal{K}^C = \{1, \dots, K + 1\} \setminus \mathcal{K}$ . First, suppose  $\mathcal{K} = \emptyset$ . Then  $l_n(F, s) = 0$  for any choice of  $F_k(s)$  ( $k = 1, \dots, K$ ), and hence  $\check{F}_n(s)$  is a maximizer of  $l_n(F, s)$ . Next, suppose  $\mathcal{K} \neq \emptyset$ , or equivalently,  $N(s) > 0$ . Then any maximizer of  $l_n(F, s)$  subject to the constraint  $F_+(s) \leq 1$  must set  $F_k(s) = 0$  for  $k \in \mathcal{K}^C$ . Hence, for  $k \in \mathcal{K}^C$  the maximizer is unique and equals  $\check{F}_{nk}(s)$ . If  $|\mathcal{K}| = 1$ , then  $l_n(F, s)$  contains only one nonzero term, and it is clear that the corresponding  $F_k(s)$  should be set to 1, which equals  $\check{F}_{nk}(s)$ . If  $|\mathcal{K}| > 1$ , we define  $k^* = \max \mathcal{K}$ . Then  $N_{k^*}(s) = N(s) - \sum_{k \in \mathcal{K} \setminus \{k^*\}} N_k(s)$  and any maximizer of  $l_n(F, s)$  over  $\mathcal{F}_K^*$  must satisfy  $F_{k^*}(s) = 1 - \sum_{k \in \mathcal{K} \setminus \{k^*\}} F_k(s)$ . Hence, we can write  $l_n(F, s) = \sum_{k \in \mathcal{K} \setminus \{k^*\}} N_k(s) \log\{F_k(s)\} + \{N(s) - \sum_{k \in \mathcal{K} \setminus \{k^*\}} N_k(s)\} \log\{1 - \sum_{k \in \mathcal{K} \setminus \{k^*\}} F_k(s)\}$ . This function is strictly concave in  $F_k(s)$  for  $k \in \mathcal{K} \setminus \{k^*\}$ . The unique maximizer can be determined by solving  $\partial l_n(F, s) / \partial F_k(s) = 0$  for  $k \in \mathcal{K} \setminus \{k^*\}$ , which yields  $\check{F}_{nk}(s)$ ,  $k \in \mathcal{K}$ .  $\square$

*Proof of Lemma 2.* Let  $s \in \mathcal{S}$  be a regular point in the discrete model. We first consider the maximum likelihood estimator for the so-called basic case, where  $s \notin \{s_{\text{inf}}, s_{\text{sup}}\}$ . Let  $\mathcal{K}^+ = \{k \in \{1, \dots, K\} : F_{0k}(s) > 0\}$ . For  $k \in \{1, \dots, K\} \setminus \mathcal{K}^+$ , we have  $N_k(s) = 0$ . Hence, the corresponding  $F_k(s)$ s do not contribute to the likelihood and we directly obtain that the corresponding estimators satisfy  $\hat{F}_{nk}(s) = \check{F}_{nk}(s) = 0$ . So we are done if  $\mathcal{K}^+ = \emptyset$ . Otherwise, we are left to show

$\text{pr}[\bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s) = \check{F}_{nk}(s)\}] \rightarrow 1$  as  $n \rightarrow \infty$ . Define the events

$$A_n(s) = \bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s_-) < \hat{F}_{nk}(s) < \hat{F}_{nk}(s_+)\}, \quad B_n(s) = \bigcap_{k \in \mathcal{K}^+ \cup \{K+1\}} \{N_k(s) > 0\}.$$

The assumptions on  $s$  imply  $F_{0k}(s_-) < F_{0k}(s) < F_{0k}(s_+)$  for  $k \in \mathcal{K}^+$ . By combining this with the consistency of  $\hat{F}_n$  in Theorem 1, it follows that  $\text{pr}\{A_n(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ . Moreover, the law of large numbers,  $G(\{s\}) > 0$ ,  $F_{0k}(s) > 0$  for  $k \in \mathcal{K}^+$  and  $F_{0+}(s) < 1$  imply  $\text{pr}\{B_n(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ . Hence,  $\text{pr}\{A_n(s) \cap B_n(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ , and the proof for the basic case can be completed by showing that the event  $\{A_n(s) \cap B_n(s)\}$  implies  $\bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s) = \check{F}_{nk}(s)\}$ . We do this using contraposition. Thus, suppose  $\{A_n(s) \cap B_n(s)\}$  holds. This implies  $k^* = K + 1$  in the proof of Lemma 1, and it follows that  $\check{F}_{nk}(s)$ ,  $k \in \mathcal{K}^+$ , is the unique solution of  $\partial l_n(F)/\partial F_k(s) = 0$ ,  $k \in \mathcal{K}^+$ . Now assume there is a  $j \in \mathcal{K}^+$  such that  $\hat{F}_{nj}(s) \neq \check{F}_{nj}(s)$ . Then there must be a  $\bar{k} \in \mathcal{K}^+$  such that  $\partial l_n(F)/\partial F_{\bar{k}}(s)|_{\hat{F}_n(s)} \neq 0$ . Let  $\sigma \in \{-1, +1\}$  be the sign of  $\partial l_n(F)/\partial F_{\bar{k}}(s)|_{\hat{F}_n(s)}$ , and define  $\hat{F}_n^{\text{new}}(s) = \hat{F}_n(s) + \gamma \sigma e_{\bar{k}}$ , where  $e_{\bar{k}}$  is the unit vector in  $\mathbb{R}^K$  with a 1 at the  $k$ th entry. Then for  $\gamma > 0$  sufficiently small, replacing  $\hat{F}_n(s)$  by  $\hat{F}_n^{\text{new}}(s)$  increases the loglikelihood. Moreover, this replacement does not violate the constraints of  $\mathcal{F}_K$ , as for  $\gamma > 0$  sufficiently small we have  $\hat{F}_{n\bar{k}}(s_-) < \hat{F}_{n\bar{k}}^{\text{new}}(s) < \hat{F}_{n\bar{k}}(s_+)$  and  $\hat{F}_{n+}^{\text{new}}(s) < \hat{F}_{n+}(s) \leq 1$ . This shows that  $\hat{F}_n$  cannot be the maximum likelihood estimator, which is a contradiction.

If  $s \neq s_{\text{inf}}$  and  $s = s_{\text{sup}}$ , we distinguish two cases. If  $F_{0+}(s) < 1$ , the proof of the basic case goes through with the only change that  $A_n(s) = \bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s_-) < \hat{F}_{nk}(s)\} \cap \{\hat{F}_{n+}(s) < 1\}$ . If  $F_{0+}(s) = 1$ , then  $N_{K+1}(s) = 0$  and  $1 - F_+(s)$  does not contribute to the loglikelihood. Hence, the maximum likelihood estimator must satisfy  $\hat{F}_{n+}(s) = 1$  and this equals  $\check{F}_{n+}(s)$  if  $N(s) > 0$ . If  $K = 1$ , this implies  $\text{pr}\{\hat{F}_n(s) = \check{F}_n(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ , so that we are done. If  $K > 1$ , we use the proof for the basic case with the following changes. We define  $A_n(s) = \bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s_-) < \hat{F}_{nk}(s)\}$  and  $B_n(s) = \bigcap_{k \in \mathcal{K}^+} \{N_k(s) > 0\}$ . As before, we have  $\text{pr}\{A_n(s) \cap B_n(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ . We will therefore show that  $\{A_n(s) \cap B_n(s)\}$  implies  $\bigcap_{k \in \mathcal{K}^+} \{\hat{F}_{nk}(s) = \check{F}_{nk}(s)\}$ , using contraposition. Thus, assume  $\{A_n(s) \cap B_n(s)\}$  holds. This implies  $k^* = \max \mathcal{K}^+$  in the proof of Lemma 1, meaning that  $\check{F}_{nk}$ ,  $k \in \mathcal{K}^+$ , are found by solving  $\partial l_n(F)/\partial F_k(s) = 0$  for  $k \in \mathcal{K}^+ \setminus \{k^*\}$  and setting  $\check{F}_{nk^*}(s) = 1 - \sum_{k \in \mathcal{K}^+ \setminus \{k^*\}} \check{F}_{nk}(s)$ . Assume  $\hat{F}_{nk}(s) \neq \check{F}_{nk}(s)$  for some  $k \in \mathcal{K}^+$ . Then there must be a  $\bar{k} \in \mathcal{K}^+ \setminus \{k^*\}$  such that  $\partial l_n(F)/\partial F_{\bar{k}}(s)|_{\hat{F}_n(s)} \neq 0$ . Define  $\sigma$  as the sign of  $\partial l_n(F)/\partial F_{\bar{k}}(s)|_{\hat{F}_n(s)}$ , and define  $\hat{F}_n^{\text{new}}(s) = \hat{F}_n(s) + \gamma \sigma e_{\bar{k}} - \gamma \sigma e_{k^*}$ . Then for  $\gamma > 0$  sufficiently small, replacing  $\hat{F}_n(s)$  by  $\hat{F}_n^{\text{new}}(s)$  increases the loglikelihood. Moreover, this replacement does not violate the constraints of  $\mathcal{F}_K$ , as for  $\gamma > 0$  sufficiently small we have  $\hat{F}_{n\bar{k}}(s_-) < \hat{F}_{n\bar{k}}^{\text{new}}(s)$ ,  $\hat{F}_{nk^*}(s_-) < \hat{F}_{nk^*}^{\text{new}}(s)$  and  $\hat{F}_{n+}^{\text{new}}(s) = \hat{F}_{n+}(s) = 1$ . Hence,  $\hat{F}_n$  cannot be the maximum likelihood estimator, again a contradiction.

The proof for the maximum likelihood estimator is completed by considering two remaining special cases. If  $s = s_{\text{inf}}$  and  $s \neq s_{\text{sup}}$ , then the proof for the basic case goes through with the only change that  $A_n(s) = \bigcap_{k \in \mathcal{K}^+} \{0 < \hat{F}_{nk}(s) < \hat{F}_{nk}(s_+)\}$ . If  $s = s_{\text{inf}} = s_{\text{sup}}$ , then  $|\mathcal{S}| = 1$  and monotonicity constraints do not play any role in the maximum likelihood estimator, so that  $\hat{F}_n = \check{F}_n$  follows immediately.

The proof for the naive estimator follows directly from the proof for the maximum likelihood estimator by taking  $K = 1$ . To see this, let  $k \in \{1, \dots, K\}$  and recall that the naive estimator is the maximum likelihood estimator for the reduced current status data  $(\Delta_k^i, C_i)$  ( $i = 1, \dots, n$ ). Hence, the proof for the maximum likelihood estimator implies  $\text{pr}\{\hat{F}_{nk}(s) = \check{F}_{nk}^{\text{red}}(s)\} \rightarrow 1$  as  $n \rightarrow \infty$ , where  $\check{F}_{nk}^{\text{red}}$  is the simple estimator based on the reduced data. The proof is completed by observing that  $\check{F}_{nk}^{\text{red}} = \check{F}_{nk}$ .  $\square$

*Proof of Theorem 3.* Because of Lemma 2, it is sufficient to derive the limiting distribution of  $\check{F}_n$ . Let  $k \in \{1, \dots, K\}$  and  $s \in \mathcal{S}$ . Since  $\text{pr}\{N(s) > 0\} \rightarrow 1$  as  $n \rightarrow \infty$ , we can assume  $N(s) > 0$ . We first consider the case  $0 < F_{0k}(s) < 1$ . Then

$$\begin{aligned} n^{1/2} \{\check{F}_{nk}(s) - F_{0k}(s)\} &= N(s)^{-1} n^{1/2} \{N_k(s) - F_{0k}(s)N(s)\} \\ &= N(s)^{-1} n^{-1/2} \sum_{i=1}^n \{\Delta_k^i - F_{0k}(s)\} \mathbf{1}\{C_i = s\}, \end{aligned}$$

and the result follows from  $N(s) \rightarrow_p G(\{s\})$ , the multivariate central limit theorem, and Slutsky's lemma, e.g., Van der Vaart (1998, Lemma 2.8(iii)).

If  $F_{0k}(s) = 0$ , then  $N_k(s) = 0$  and hence  $\check{F}_{nk}(s) = 0 = F_{0k}(s)$  always. Similarly, if  $F_{0k}(s) = 1$ , we have  $N_k(s) = N(s)$  and hence  $\check{F}_{nk}(s) = 1 = F_{0k}(s)$  whenever  $N(s) > 0$ . These results are in agreement with the theorem, since in these cases  $\{V(s)\}_{k,k} = 0$ , leading to a degenerate limiting distribution that should be interpreted as a point mass at zero. It can be easily verified that the off-diagonal elements  $\{V(s)\}_{k,j} = 0$  for  $j \in \{1, \dots, K\}$ ,  $j \neq k$ , are also correct in these cases.  $\square$

#### REFERENCES

- AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. & SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**, 641–47.
- BANERJEE, M. & WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699–731.
- BANERJEE, M. & WELLNER, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–24.
- GENTLEMAN, R. & VANDAL, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Can. J. Statist.* **30**, 557–71.
- GROENEBOOM, P., JONGBLOED, G. & WELLNER, J. A. (2008a). The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scand. J. Statist.* **35**, 385–99.
- GROENEBOOM, P., MAATHUIS, M. H. & WELLNER, J. A. (2008b). Current status data with competing risks: consistency and rates of convergence of the MLE. *Ann. Statist.* **36**, 1031–63.
- GROENEBOOM, P., MAATHUIS, M. H. & WELLNER, J. A. (2008c). Current status data with competing risks: limiting distribution of the MLE. *Ann. Statist.* **36**, 1064–89.
- GROENEBOOM, P. & WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comp. Graph. Statist.* **10**, 388–400.
- HUDGENS, M. G., SATTEN, G. A. & LONGINI, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* **57**, 74–80.
- JEWELL, N. P. & KALBFLEISCH, J. D. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics* **5**, 291–306.
- JEWELL, N. P., VAN DER LAAN, M. J. & HENNEMAN, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika* **90**, 183–97.
- KITAYAPORN, D., VANICHSENI, S., MASTRO, T. D., RAKTHAM, S., VANİYAPONGS, T., DES JARLAIS, D. C., WASI, C., YOUNG, N. L., SUJARITA, S., HEYWARD, W. L. & ESPARZA, J. (1998). Infection with HIV 1 subtypes B and E in injecting drug users screened for enrollment into a prospective cohort in Bangkok, Thailand. *J. Acq. Immun. Defic. Hum. Retrovir.* **19**, 289–95.
- KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 282–292. Institute of Mathematics and Statistics Collection 1. Ed. N. Balakrishnan, E. A. Peña and M. J. Silvapulle. IMS Collections. Beachwood, OH: Institute of Mathematics and Statistics.
- KRAILO, M. D. & PIKE, M. C. (1983). Estimation of the distribution of age at natural menopause from prevalence data. *Am. J. Epidemiol.* **117**, 356–61.
- MAATHUIS, M. H. (2005). Reduction algorithm for the MLE for the distribution function of bivariate interval censored data. *J. Comp. Graph. Statist.* **14**, 352–62.
- MACMAHON, B. & WORCESTOR, J. (1966). Age at menopause, United States 1960–1962. *National Center for Health Statist. Vital Health Statist.* **11**, 161–73.
- POLITIS, D. N. & ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031–50.
- SEN, B., BANERJEE, M. & WOODROOFE, M. (2010). Inconsistency of the bootstrap: the Grenander estimator. *Ann. Statist.* **38**, 1953–77.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
- VANICHSENI, S., KITAYAPORN, D., MASTRO, T. D., MOCK, P. A., RAKTHAM, S., DES JARLAIS, D. C., SUJARITA, S., SRISUWANVILAI, L., YOUNG, N. L., WASI, C., SUBBARAO, S., HEYWARD, W. L., ESPARZA, J. & CHOOPANYA, K. (2001). Continued high HIV-1 incidence in a vaccine trial preparatory cohort of injection drug users in Bangkok, Thailand. *AIDS* **15**, 397–405.
- WOODROOFE, M. & ZHANG, R. (1999). Isotonic estimation for grouped data. *Statist. Prob. Lett.* **45**, 41–7.
- YU, Q., SCHICK, A., LI, L. & WONG, G. Y. C. (1998). Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *Can. J. Statist.* **26**, 619–27.
- ZHANG, R., KIM, J. & WOODROOFE, M. (2001). Asymptotic analysis of isotonic estimation for grouped data. *J. Statist. Plan. Infer.* **98**, 107–17.

[Received October 2009. Revised November 2010]