

LOESS smoothed density estimates for multivariate survival data subject to censoring and masking

Peter Adamic*

Laurentian University, Sudbury, Ontario P3E 2C6, Canada

Jenna Guse

University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Abstract

Actuaries often encounter censored and masked survival data when constructing multiple-decrement tables. In this paper, we propose estimators for the cause-specific failure time density using LOESS smoothing techniques that are employed in the presence of left-censored data, while still allowing for right-censored and exact observations, as well as masked causes of failure. The smoothing mechanism is incorporated as part of an expectation-maximisation algorithm. The proposed models are applied to a bivariate African sleeping sickness data set.

Keywords

LOESS smoothing; Multiple decrement; Left censoring; Right censoring

1. Introduction

Actuaries often encounter censored and masked survival data when constructing single or multiple-decrement tables. Censoring is present in lifetime data when the time of death (or failure, in the more general survival context) is not known precisely, and masking is present when the cause of death is not known precisely. The survival literature is rather copious with respect to many parametric and semi-parametric models that can be employed when censoring and/or masking is present.

A non-parametric maximum likelihood estimator (NPMLE) of the cumulative incidence for competing risks data subject to right censoring was given by Aalen (1976) and Kalbfleisch & Prentice (1980). Dinse (1982) proposed an NPMLE for right-censored and masked competing risks data to be computed with the explicit use of an expectation-maximisation (EM) algorithm. Hudgens *et al.* (2001) first presented an NPMLE estimated using an EM algorithm for competing risks data, subject to both interval censoring and truncation. Jewell *et al.* (2003) and Groeneboom *et al.* (2008) continued similar studies of an NPMLE for current status data, an extreme form of interval-censored data, without masking. Subsequently, Adamic (2010, 2012) developed generalisations of Turnbull's (1974, 1976) classical univariate algorithms for modelling competing risks, but in a more general setting, where each failure was allowed to be associated with any subset of possible failure modes (i.e. masking). Overall, distribution-free models that can be employed in a multiple-decrement

*Correspondence to: Peter Adamic, Laurentian University, 935 Ramsey Lake Rd, Sudbury, Ontario P3E 2C6, Canada. Tel: (705)675-1151 x2325; E-mail: padamic@laurentian.ca

context have received relatively little attention to date, in part due to the complexity that censoring and masking introduces.

In other recent research, models have been developed to analyse competing risks data subject to various censoring schemes. Craiu & Reiser (2006) developed a dependent competing risks model using an EM-based approach for masked data. Wang *et al.* (2012) proposed a new model for the NPMLE of interval-censored and masked competing risks data, the random partition masking model, which does not rely on the assumption that masked failure causes are independent of failure time. Yu & Li (2012, 2014) examined the NPMLE proposed by Dinse (1982); they concluded that the NPMLE was inconsistent and not unique and they introduced a consistent NPMLE of the joint distribution function with right-censored and masked competing risks data under another new model (the dependent masking and right-censoring model).

In this paper, we develop a smoothed non-parametric density estimator for modelling multiple-decrement¹ failure times that are subject to the possibility of both left censoring as well as right censoring, while still allowing for exact observations. Despite the novelty of modelling masked competing risks data using an EM algorithm, there is a significant drawback associated with the various approaches that have been developed to date. As expounded in Hudgens *et al.* (2001), estimators of this type will have an unexpected and undesirable property, namely, that the resulting estimator of the survival distribution will be undefined over a potentially larger set of regions than the NPMLE of $S(t)$, ignoring failure type. Indeed, the problem is, quite intuitively, even more prevalent in the multiple-decrement environment: the self-consistent competing risks algorithms or *SC-CR algorithms* of Adamic (2010, 2012) can be seen to converge only over a class of intervals called *cause-specific innermost intervals*.

To remedy this problem, we have chosen to generalise a univariate kernel density estimate found in Braun *et al.* (2005) that was used to fill in the gaps between the *innermost intervals* that were created by invoking the self-consistent EM algorithm of Turnbull (1976). The converged estimator of the failure rate distribution is often difficult to smooth, due to the large gaps between innermost intervals as well as the multimodal shape of the distribution that will naturally arise when there are many gaps in the probability masses. As mentioned in Duchesne & Stafford (2001), adopting a kernel smoothed estimator at each iteration avoids the bias created by arbitrarily assigning probability mass at the right end points of the innermost intervals, as recommended by Pan (2000), and is better at borrowing more information from neighbouring data points than would otherwise be the case. Duchesne & Stafford (2001) go on to state that since the innermost intervals are no longer in the picture, the kernel modification moves the algorithm away from problem-causing areas – areas where Turnbull's algorithm can sometimes get stuck at local solutions (also see Li *et al.*, 1997). Our generalisation of the approach proposed by Braun *et al.* (2005) will be actualised in two ways: first, in order to smooth the data we will use local regression and the LOcal regrESSion (LOESS) fitting procedure developed by Cleveland (1979) and Cleveland & Devlin (1988) as opposed to kernel density estimation – indeed, the latter is just a special case of the former; second, our models will be developed in the more general multiple-decrement setting as opposed to the simpler univariate setting.

2. Definitions, Notation, and Terminology

This article will focus exclusively on the so-called doubly-censored data, that is, data that is subject to both left and right censoring (in addition to exact observations), as defined by Zhou (2004) below. This section is mostly from Adamic (2010).

¹ The terms *multiple decrement* and *competing risk* will be used synonymously.

Definition 2.1 Suppose lifetimes X_1, \dots, X_n are non-negative, independent, and identically distributed random variables subject to censoring. In the case of double censoring, we observe both left and right censoring (in addition to exact observations). That is, we observe

$$\{Z_i, \Delta_i\} = \begin{cases} (X_i, 1), & \text{if } L_i \leq X_i \leq R_i \\ (R_i, 0), & \text{if } X_i > R_i \\ (L_i, 2), & \text{if } X_i < L_i \end{cases}$$

where L_i and R_i are the left and right-censoring times, respectively, $\forall i = 1, \dots, n, R_i \geq L_i$. Also, Δ_i is an indicator of the type of censoring, if any.

In a multiple risk forum, it may be the case that some of the failure modes that could have been responsible for the actual failure are ruled out as possible candidates. In this case, the remaining subset of possible failure modes are considered to be *masked*. A formal definition is given below.

Definition 2.2 Suppose a failure mode $j, j \in \{1, \dots, k\}$, is associated with a failure time t . If the mode is not known exactly, but is known only to belong to a certain subset of all possible modes, then the failure is caused by a subset of modes that are said to be masked. Furthermore, following Park & Padgett (2006), if the failure mode is completely unknown, then the set of failure modes is said to be completely masked; if one or more of the failure modes can be ruled out, then the remaining subset of possible failure modes is considered to be partially masked.

In this paper, the possibility of determining some of the failure modes a posteriori, and then incorporating the new information into the model will not be directly addressed. However, the algorithms will be sufficiently versatile that once a mode of failure can be determined exactly (or reduced to a smaller set of possible modes), the algorithms can simply be executed again utilising the updated information. See Flehinger *et al.* (1998) for more details on carrying out analyses using updated masking information.

The following preliminary notation, based on the International Actuarial Notation (IAN) guidelines as found in Bowers *et al.* (1997), is applicable to any survival or life table context. First, define ${}_tq_x = \Pr\{\text{failure prior or equal to time } (t+x) \mid \text{subject is currently aged } x\}$. For example, in human mortality studies, ${}_{10}q_{45}$ would represent the probability a 45-year-old dies in the next 10 years. When working with a life table, it is often convenient to replace q with d to represent the number, instead of the probability, of subjects that fail in a given time interval. Thus, ${}_{10}d_x$ would represent the number of failures in the interval $(x, x+t]$. Also, let ${}_t p_x = \Pr\{\text{survival to time } (t+x) \mid \text{subject is currently aged } x\}$. For example, ${}_{10}p_{45}$ would represent the probability a 45-year old would survive to age 55 years. Clearly, ${}_t p_x + {}_t q_x = 1$. When $t = 1$, the subscript is often omitted. Thus, ${}_1 p_{45} = p_{45}$, could represent the probability a 45-year old survives 1 year.

To motivate the competing risks theory, some additional notation is required. A superscript in brackets will be used to identify which specific risk is being considered. For example, $b_x^{(j)}(t)$ would represent the hazard rate for risk j at time $x+t$, given a current age of x . Note the implicit assumption here that only one distinct cause (or failure mode) can be responsible for any particular failure. This assumption is germane to all of the theory presented in this paper. The superscript τ will be used to represent the set of all the risks taken together, in aggregate. For example, ${}_t p_x^{(\tau)}$ would represent the

probability of surviving all risks up to time t . The following relationships are given without proof:

$$h_x^{(j)}(t) = \frac{f(t, j)}{t p_x^{(j)}}, \quad b_x^{(j)}(t) = \sum_j h_x^{(j)}(t) \cdot {}_t q_x^{(j)} = \sum_j {}_t q_x^{(j)}$$

where $f(t, j)dt \approx \Pr\{(t < T \leq t + dt) \cap (J = j)\}$, and h denotes the hazard function. For example, ${}_t q_x^{(j)}$ would represent the probability of failure for some subject aged x prior or equal to time $x + t$ by cause j in the multiple risk forum.

A very important function to model in any competing risks analysis is the cumulative incidence function (CIF), and this is usually accomplished for all of the possible failure modes. The standard definition in the literature for the CIF for cause j is

$$F_n^{(j)}(t) = \Pr(T \leq t, J = j) = \int_0^t h^{(j)}(u) {}_u p_0^{(j)} du \tag{1}$$

where the joint pair of random variables (T, J) capture the time of failure and the corresponding mode of failure, respectively. It should also be noted that $F_n^{(j)}(t) = {}_t q_0^{(j)}$ under IAN. The most common estimator for the CIF in counting process notation is

$$\hat{F}_n^{(j)}(t) = \int_0^t \hat{S}(u) d\hat{\Lambda}_j(u) \tag{2}$$

which, as opined in Lawless (2003), is typically estimated non-parametrically using

$$\hat{F}_n^{(j)}(t) = \sum_{i: t_i \leq t} \hat{S}(t_i) \cdot \frac{\delta_{ij}}{n_i}, \quad j = 1, \dots, k \tag{3}$$

where δ_{ij} equals 0 or 1 depending upon whether or not subject i failed due to cause j , and n_i captures the number at risk at time t_i . Unfortunately, equation (3) cannot be used here for two main reasons. First, equation (3) can only be used for complete or right-censored data. Thus, this closed-form estimator for the CIF cannot be applied to doubly-censored data. Second, equation (3) assumes that all of the failure modes are known exactly. Consequently, if any of the failure modes are masked, this estimator cannot be applied. That is, all of the δ_{ij} terms cannot be extracted from the data.

It is also noteworthy to mention that there is a non-identifiability phenomenon at work in the multiple-decrement model, as described by Tsiatis (1975). An observational study consisting of data that only furnishes the joint pair (T_i, J_i) for each subject i (and possibly some additional indicator variables for the censored range or ranges) will not allow for discrimination between independent and dependent modes of failure. This fact was also noted by Cox (1959). In many cases however, this is the only available data. Thus, for the sake of model simplification, an independence assumption has traditionally been invoked between the various competing risks. Some recent research has focussed on developing dependent competing risks models (see Escarela & Carriere, 2003; Craiu & Reiser, 2006; Wang *et al.*, 2012; Li & Yu, 2014). The complication of mutually dependent failure modes will not be considered in the ensuing theory.

The assumption of independence is also found in another significant way, namely, in the manner in which censored observations are reported. Censoring can sometimes be accomplished in such a way that the number of exposures to the risk (or risks) depends on the censoring mechanism itself, a dynamic which is often termed *selection bias*, as noted by Gichangi & Vach (2005). The absence of selection bias implies that *independent censoring* is taking place (Gichangi & Vach, 2005), and the assumption of independent censoring means that the cause-specific hazards do not depend on

whether or not a particular subject was censored (Gichangi & Vach, 2005). Symbolically, this means that

$$\lim_{b \rightarrow 0} \frac{\Pr\{T_i < t + b, J = j \mid T_i \geq t\}}{b} = \lim_{b \rightarrow 0} \frac{\Pr\{T_i < t + b, J = j \mid T_i \geq t, T_i \in \xi_i\}}{b}$$

where subject i is censored by the observed set ξ_i , and will hold true for all of the k competing risks (Gichangi & Vach, 2005).

One of the primary characteristics of the estimators of the cause-specific CIFs that will be developed in this paper will be that they are members of the class of estimators that are called *self-consistent estimators*. A formal definition of the concept of self-consistency for the univariate case is given next, as found in Stafford (2005), and is followed by conceptualisations for use in the arena of competing risks.

Definition 2.3 An estimator $\hat{F}_n(t)$ for a cumulative distribution function (CDF) is said to be a self-consistent estimator of the CDF if $\hat{F}_n(t) = E_{\hat{F}_n}[F_n(t) \mid \Psi]$, where $F_n(t)$ is the empirical CDF under complete data and Ψ denotes the set of all of the observed data (complete or censored).

Essentially, Definition 2.3 is stating that at each time t , the expected value of the CDF is the same with entirely complete data or with censored and/or masked data. Since any censored failure can have a failure time at any t , we condition on the entire set of failures, Ψ . Note that in the multiple-decrement setting, the CDF is simply replaced by the cause-specific CIF. Following Duchesne & Stafford (2001), a generalisation for self-consistency pertaining to CIFs is offered as follows.

Definition 2.4 An estimator $\hat{F}^{(j)}(t)$ for a CIF is said to be a self-consistent estimator if $\hat{F}^{(j)}(t) = E_{\hat{F}}[F^{(j)}(t) \mid \Psi]$, where $F^{(j)}(t)$ is the empirical CIF under complete data, Ψ denotes the set of all of the observed data (complete or censored, with the possibility of masking), and \hat{F} is understood to include the current information in all of the CIFs for modes $1 \dots k$.

Consider the following as motivation for developing the SC-CR algorithm for doubly-censored data. Huang (2000) noted that self-consistent equations are essentially score equations of indicator variables appropriately defined, depending on the type of censoring. When the data are potentially censored, Huang (2000) also states that \hat{F}_n can be obtained by taking the conditional expectation of F_n , conditional on the observed data under the probability measure induced by \hat{F}_n itself. However, evaluation of the conditional expectations in the SC equations will require the use of an iterative scheme, as no closed-form expression for these expectations exist in these non-parametric cases. Also, as noted by Turnbull (1974), convergence of a self-consistent iterative method implies that the application of the SC equations at some iteration will essentially have no effect on the realised value of the estimator (up to some small tolerance of discrepancy, ϵ say). Indeed, the SC procedure can also be shown to be a special case of the more general EM algorithm of Dempster *et al.* (1977).

3. The SC-CR Algorithm for Doubly-Censored Data

3.1. A self-consistent algorithm

Section 3, which summarises the SC-CR algorithm for doubly-censored data, is from Adamic (2010) and/or Adamic *et al.* (2010). The idea of a self-consistent algorithm, as well as the concept of

self-consistency itself, were propounded by Efron (1967). Efron developed an iterative algorithm to estimate the CDF for right-censored data using

$$\tilde{F}_K(x) = \frac{1}{n} \left[N(x) - \sum_{L_i \leq x, \Delta_i = 0} \frac{1 - \tilde{F}_{K-1}(x)}{1 - \tilde{F}_{K-1}(L_i)} \right] \tag{4}$$

where $N(x) = \sum I[X_i \leq x]$, $\Delta_i = 0$ if X_i is right censored, and K indexes the iteration number (see Braun *et al.*, 2005). Turnbull (1974) extended Efron’s algorithm to included doubly-censored data, again, for a single risk. Also noteworthy is the fact that Turnbull (1974, 1976) proved that the resulting survival estimators, upon convergence of the algorithm, were NPMLs. The algorithm is quite well known, and as revealed by many researchers who employ survival techniques in practice (see Finkelstein, 1986; Yu *et al.*, 2000), it is one of the most common estimators that has been used in medical applications of censored survival analysis over the last several decades. For the sake of brevity, Turnbull’s univariate algorithm will not be summarised here. The reader is referred to Klein & Moeschberger (1997) for this purpose.

3.2. Developing the model

The SC-CR algorithm is carried out in the presence of doubly-censored failures, assuming the standard independence assumptions previously mentioned. However, it is further assumed that the masking probabilities are also independent of the failure cause itself. Based on these assumptions, the likelihood function under a double censoring scenario with the possibility of exact observations (where the cause of failure is known) and completely masked modes of failure can be expressed as

$$L \propto \prod_{i=1}^n \left[(1 - {}_i p_0)^{c_i} ({}_i p_0)^{r_i} \prod_{j=1}^k ({}_{t_i - t_{i-1}} q_{t_{i-1}}^{(j)})^{d_i^{(j)}} \right] \tag{5}$$

where c_i represents the number of left-censored observations, r_i the number of right-censored observations, and $d_i^{(j)}$ the number of failures for risk j , $\forall i = 1, \dots, n$. Based on the notation in the likelihood, “exact” should be taken here to mean, “since the last time point, t_{i-1} ”. Equivalently, it will be assumed that $[t_i, t_i] \equiv (t_{i-1}, t_i]$, which is consistent with the step-function nature of the CIF estimators.

The algorithm may be implemented as follows. For ease of appropriation, the steps, statements, and logic of the algorithm will be directly generalised from those of the single variable algorithm of Klein & Moeschberger (1997).

3.2.1. The SC-CR algorithm

Step 0: Provide initial estimates of the overall survival probabilities at each $t_r, {}_t p_0^{(\tau)}$, by equally allotting all of the probability mass for each time point. Also, find initial estimates for the probability of failing during each time interval by cause $j, {}_{t_r - t_{r-1}} q_{t_{r-1}}^{(j)}$, for $r = 1, \dots, m$, again by uniformly distributing the probability mass. Ignore all of the left-censored observations when calculating these estimates.

Step 1: Using the current estimates of ${}_t p_0^{(\tau)}$ and ${}_{t_r - t_{r-1}} q_{t_{r-1}}^{(j)}$, calculate estimates of the conditional probabilities $e_r^{(j)} = P[t_{r-1} < X^{(j)} \leq t_r \mid X^{(\tau)} \leq t_i]$, where $X^{(j)}$ represents the event of failure due to cause j , using

$$\hat{e}_r^{(j),K} = \frac{{}_{t_{r-1}} \hat{p}_0^{(\tau),K} \cdot {}_{t_r - t_{r-1}} \hat{q}_{t_{r-1}}^{(j),K}}{1 - {}_{t_i} \hat{p}_0^{(\tau),K}}, \quad \forall r \leq i \tag{6}$$

Step 2: Using the results of the previous step, estimate the number of cause-specific failures at time t_i by using

$$\hat{d}_i^{(j),K} = d_i^{(j)} + \sum_{i=r}^m c_i \hat{e}_{ir}^{(j),K} \tag{7}$$

Step 3: Compute ${}_{t_r-t_{r-1}}\hat{q}_{t_{r-1}}^{(j)} \forall r=1, \dots, m$, using a generalised cause-specific product-limit estimator based on the current estimates of $\hat{d}_i^{(j)}$. If

$$\sup_{t_{r,j}} \left| {}_{t_r-t_{r-1}}\hat{q}_{t_{r-1}}^{(j),K} - {}_{t_r-t_{r-1}}\hat{q}_{t_{r-1}}^{(j),K-1} \right| < \varepsilon$$

(jointly for all of the time points t_i and causes j , given some small predetermined $\varepsilon > 0$), stop the algorithm; otherwise return to Step 1.

So far, the issue of partial masking has been ignored. The possibility of partial masking, however, can be easily introduced into the algorithm. In essence, there will need to be a proper allocation of the left-censored observations into each of the partially masked sets, so that the property of self-consistency will be maintained. To this end, Step 1 would need to be modified as follows. Define

$$e_{irz}^{(j|S_{iz})} = P \left[t_{r-1} < X^{(j)} \leq t_r \mid X^{(S_{iz})} \leq t_i \right] \tag{8}$$

$\forall r \leq i, j \subseteq S_{iz}$, where $z = 1, \dots, y_i \leq 2^{i-1}$ indexes each of the distinct masked sets, S_{iz} , at time t_i . Then for each time point t_i , and for each competing risk j , compute

$$\hat{e}_{irz}^{(j|S_{iz}),K} = \frac{{}_{t_{r-1}}\hat{p}_0^{(\tau),K} \cdot {}_{t_r-t_{r-1}}\hat{q}_{t_{r-1}}^{(j),K}}{t_i \hat{q}_0^{(S_{iz}),K}}, \forall r \leq i, j \subseteq S_{iz} \tag{9}$$

where ${}_{t_i}\hat{q}_0^{(S_{iz})}$ represents the current probability of failing due to any of the causes that comprise S_{iz} up until time t_i . Step 2 would also need to be modified slightly. The term $\hat{d}_i^{(j),K}$ would now be calculated using

$$\hat{d}_i^{(j),K} = d_i^{(j)} + \sum_{i=r}^m \sum_z c_i^{(S_{iz})} \hat{e}_{irz}^{(j|S_{iz}),K} \tag{10}$$

where $c_i^{(S_{iz})}$ represents the number of left-censored observations at time t_i masked by the failure set S_{iz} .

A proof that the SC-CR algorithms (for both the partially masked and completely masked cases) produce self-consistent estimators of the CIFs for each failure mode can be found in Appendix. That the CIFs derived from the SC-CR algorithms are also NPMLEs is strongly suggested by the fact that the estimators are self-consistent (due to the fact that the expected values of the CIFs for all time points equals an NPMLE of each CIF in the absence of any left-censored observations or masking – as, for example, found in equation (3) of section 2). For the purposes of the present paper, we will be content to rely on the statistical merits of self-consistency.

3.3. Illustrative numerical examples

In this section, the SC-CR algorithms will be illustrated with some numerical examples. First, a simple example of Turnbull’s traditional univariate algorithm is furnished, and the results are depicted in Table 1.

In Table 2, an SC-CR algorithm for doubly-censored competing risks is shown, where the total failures for each risk (at each time point) equal those as found in the univariate case. This is useful for comparative purposes, particularly because the resulting CIFs at each time point sum up to the CDF

Table 1. Iterations 1, 2, 5 (one risk, doubly censored).

K	t_i	Left	Cause 1	Right	$\hat{F}(t)$
1	1	5	110	15	0.1880
	2	5	255	20	0.6382
	3	10	180	5	0.9902
2	1	5	118.3722	15	0.1957
	2	5	263.0723	20	0.6443
	3	10	183.5555	5	0.9906
5	1	5	118.4970	15	0.1959
	2	5	263.0086	20	0.6444
	3	10	183.4945	5	0.9906

Table 2. Iterations 1, 2, 5 (two competing risks, doubly censored, complete masking).

K	t_i	Left	Cause 1	Cause 2	Right	$\hat{F}^{(1)}(t)$	$\hat{F}^{(2)}(t)$
1	1	5	40	70	15	0.0684	0.1197
	2	5	145	110	20	0.3243	0.3138
	3	10	130	50	5	0.5786	0.4116
2	1	5	43.0444	75.3278	15	0.0712	0.1245
	2	5	149.5901	113.4822	20	0.3263	0.3181
	3	10	132.5679	50.9876	5	0.5763	0.4142
5	1	5	43.0898	75.4072	15	0.0712	0.1246
	2	5	149.5539	113.4547	20	0.3263	0.3181
	3	10	132.5238	50.9707	5	0.5763	0.4143

Table 3. Iterations 1, 2, 5 (two competing risks, doubly censored, partial masking).

K	t_i	Left	Cause 1	Cause 2	Right	$\hat{F}^{(1)}(t)$	$\hat{F}^{(2)}(t)$
1	1	5*	40	70	15	0.0684	0.1197
	2	5	145	110	20	0.3243	0.3138
	3	10	130	50	5	0.5786	0.4116
2	1	5*	44.9535	73.4187	15	0.0743	0.1214
	2	5	149.5901	113.4822	20	0.3294	0.3149
	3	10	132.5679	50.9876	5	0.5795	0.4111
5	1	5*	45.0918	73.4052	15	0.0745	0.1213
	2	5	149.5539	113.4547	20	0.3296	0.3148
	3	10	132.5238	50.9707	5	0.5796	0.4110

Note: *Partial masking: three failures are due to Cause 1, and two are due to either Cause 1 or Cause 2.

values as found in the univariate case. This fact was used in the proof of the self-consistency of the SC-CR algorithms and is shown empirically here. Finally, in Table 3, we have an example of the SC-CR algorithm under partial masking. In this case, the five left-censored observations at time 1 are

assumed to be masked in the following way: three of the failures were known to be caused by Cause 1; the other two failure causes were not known. Thus, in Table 3, we see that the Cause 1 CIF is larger than previously, as would be expected. Note also that the sum of the CIFs at each time point again equals those as found in the univariate case, as it must. Finally, it should be noted that convergence was achieved in all three cases after five iterations, assuming an ϵ of 0.0001.

4. LOESS Modification to the SC-CR Algorithm

4.1. Developing the model

As motivation, consider the case where data are interval censored. In such a situation, Stafford (2005), based on the work of Goutis (1997), argues that a natural extension of the standard kernel smoothing weight is to define w_i as a conditional expectation, conditional on the observed interval $I_i = (L_i, R_i]$. That is, if the weight is defined as

$$w_i = E \left[\frac{1}{b} \tilde{K} \left(\frac{X_i - x}{b} \right) \middle| I_i \right] \tag{11}$$

the kernel density estimate of f is

$$\check{f}(x) = \frac{1}{n} \sum_{i=1}^n E \left[\frac{1}{b} \tilde{K} \left(\frac{X_i - x}{b} \right) \middle| I_i \right] \tag{12}$$

for a fixed kernel function \tilde{K} , as related in Stafford (2005). Now, Braun *et al.* (2005) espouse an intriguing way to evaluate equation (12): use an iterative algorithm such as Turnbull’s algorithm, and the kernel density estimate itself to evaluate the conditional expectation at each iteration. For our development here, we will translate this insightful approach for the case of doubly-censored data, use the more general framework of LOESS smoothing instead of kernel smoothing (again, kernel smoothing is simply a special case of local regression), and further generalise to the multivariate level comprising of multiple decrements.

We will first introduce the basic framework for the LOESS procedure proposed by Cleveland (1979) and expanded upon by Cleveland & Devlin (1988). Let y_i ($i = 1, \dots, n$) be measurements of the dependent variable, and let $x_i = (x_{i1}, \dots, x_{ip})$, $i = 1, \dots, n$, be n measurements of p independent variables. Suppose that the data are generated by $y_i = f(x_i) + \epsilon_i$, where f is a smooth function and the ϵ_i are errors assumed to be independent and identically distributed with mean 0 and variance σ^2 . Locally weighted regression provides an estimate of the smooth function, $\check{f}(x)$. The estimate of f at x , where x is any value in the p -dimensional space of independent variables, is found by defining a neighbourhood in the space of independent variables where each point in the neighbourhood is weighted according to its distance from x , with points closer to x being weighted more heavily than points further from x . Therefore, a weight function and neighbourhood size must be specified in order to complete the LOESS smoothing technique. Cleveland (1979) defined four properties that any weight function, W , should possess when performing locally weighted regression:

1. $W(z) > 0$ for $|z| < 1$;
2. $W(-z) = W(z)$;
3. $W(z)$ is a non-increasing function for $z \geq 0$;
4. $W(z) = 0$ for $|z| \geq 1$.

Following Cleveland & Devlin (1988), we introduce the tricube weight function

$$W(z) = \begin{cases} (1 - |z|^3)^3, & \text{if } |z| \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

We must now specify the neighbourhood size. Following Loader (1999a), we define the *span* or *bandwidth* (these terms will be used interchangeably) as $h(x)$ and a smoothing window $(x - h(x), x + h(x))$ for a fitting point x . To find the smooth function $\check{f}(x)$, only observations within the smoothing window are used. The weights for each observation (x_i, y_i) are then

$$w_i(x) = W\left(\frac{x_i - x}{h(x)}\right) \tag{14}$$

Using a LOESS smoothing model at each iteration of the SC-CR algorithm, the smoothed estimate for the K th iteration is

$$\check{f}_K(x) = \frac{1}{n} \sum_{i=1}^n E_{\check{f}_{K-1}} \left[W\left(\frac{x_i - x}{h(x)}\right) \middle| \Psi \right] \tag{15}$$

conditioning on all observations, Ψ , since the LOESS smoother can use all of the observed data, depending on the degree of smoothing that is desired.

For the competing risks setting, an expression analogous to equation (15) can be developed. For modes of failure indexed by $j = 1, \dots, k$, assume independence between the modes of failure. Then, the cause-specific LOESS smoothed estimate for iteration K would be expressed as

$$\check{f}_K^{(j)}(x) = \frac{1}{n} \sum_{i=1}^n E_{\check{f}_{K-1}^{(j)}} \left[W\left(\frac{x_i^{(j)} - x}{h(x)_j}\right) \middle| \Psi \right] \tag{16}$$

where an optimal value for $h(x)$, the span, should be chosen for the failure density of each independent risk j . Although the independence assumption is not required for the SC-CR algorithm, it is useful here to avoid the complication of having to invoke multivariate LOESS techniques that reflect the joint distribution between the various competing risks.

The following theorem shows that as the bandwidths for the local fit of each risk tend to 0, the smoothed estimator of the density approaches the self-consistent density. The theorem statement, steps, and logic of the proof are direct generalisations of an analogous univariate proof from Braun *et al.* (2005) regarding kernel density estimation. The tricube weight function, $W(z)$, is also a commonly used kernel function and the theorem can be applied to the LOESS setting.

Theorem 4.1 Let $\check{F}(x)_K^{(j)}(x) = \int_{-\infty}^x \check{f}(x)_K^{(j)}(t) dt$ be the LOESS smoothed estimate of the corresponding CIF for risk j at the K th iteration with a cause-specific span $h(x)_j$, under an interval-censoring scheme with the possibility of masked failure modes. Then, assuming independent tricube weight functions for each risk j

$$\lim_{h(x)_j \rightarrow 0} \check{F}_K^{(j)}(x) = \hat{F}_K^{(j)}(x) \quad \forall x, j, K$$

Proof. The self-consistent CIF is

$$\begin{aligned} \hat{F}_K^{(j)}(x) &= E_{K-1} \left[F^{(j)}(x) \middle| \Psi \right] \\ &= E_{K-1} \left[\frac{1}{n} \sum_{i=1}^n I(X_i^{(j)} \leq x) \middle| \Psi \right] \end{aligned}$$

since $F^{(j)}(x)$ captures the relative proportion of observations with $X_i \leq x$ and $J = j$, and is also the NPMLE when using complete data. Bringing the expectation into the sum yields

$$= \frac{1}{n} \sum_{i=1}^n E_{K-1} \left[I \left(X_i^{(j)} \leq x \right) \middle| \Psi \right]$$

Now, since the tricube weight function is a kernel function it is also a valid probability distribution function that approaches the empirical distribution as the bandwidth tends to 0. And since each competing risk is smoothed with its own unique span $h(x)_j$

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n E_{K-1} \left[\lim_{h(x)_j \rightarrow 0} \int_{-\infty}^x W \left(\frac{x_i^{(j)} - u}{h(x)_j} \right) du \middle| \Psi \right] \\ &= \lim_{h(x)_j \rightarrow 0} \frac{1}{n} \sum_{i=1}^n E_{K-1} \left[\int_{-\infty}^x W \left(\frac{x_i^{(j)} - u}{h(x)_j} \right) du \middle| \Psi \right] \end{aligned}$$

since $\int W \left(\frac{x_i^{(j)} - u}{h(x)_j} \right) \leq 1$, and therefore the limit can be taken outside of the expectation. By the Fubini–Tonelli theorem, the order of expectation and integration may be interchanged in this case, producing

$$\begin{aligned} &= \lim_{h(x)_j \rightarrow 0} \int_{-\infty}^x \frac{1}{n} \sum_{i=1}^n E_{K-1} \left[W \left(\frac{x_i^{(j)} - u}{h(x)_j} \right) \middle| \Psi \right] du \\ &= \lim_{h(x)_j \rightarrow 0} \int_{-\infty}^x \check{f}_K^{(j)}(u) du \quad \square \\ &= \lim_{h(x)_j \rightarrow 0} \check{F}_K^{(j)}(x) \end{aligned}$$

4.2. Span selection

A key issue when employing local regression is the optimal choice of span. The magnitude of the span is always a tradeoff between the bias and variability of the local regression fit and therefore a compromise is always necessary. Values for the span that are too small will produce an estimate for the smooth function, $\check{f}(x)$, which shows too many spurious features, reflecting large variance due to the fact that an insufficient amount of data falls into the smoothing window. In the case where the value for the span is too large, the estimate within the smoothing window may be over-smoothed and hide key structural features of the real smooth function, $f(x)$, reflecting a large amount of bias.

There are a number of methods for finding the optimal span for a data set. The first method, which is simplistic in nature, is fixed bandwidth selection. In this case, the span $h(x)$ is chosen to be a constant value, h . However, a constant choice for the bandwidth is rarely ideal because the estimate resulting from a fixed bandwidth will often behave particularly poorly in boundary or tail regions and have large variance due to differences in the density of the data, producing fits that are too noisy or in some cases undefined if the neighbourhoods are empty. Methods to determine a single fixed bandwidth in local regression include modern plug-in methods and classical methods such as cross-validation and Akaike’s information criterion. Loader (1999b) examines these methods, challenging the superiority of plug-in methods over their older counterparts, the classical methods. Four bandwidth selectors are considered by Loader (1999b): generalised cross-validation, Mallows (1973) C_p method, the plug-in algorithm of Gasser *et al.* (1991) (GKK), and a hybrid of C_p and plug-in methods proposed by Ruppert *et al.* (1995) (RSW). See Loader (1999b) for further details on implementing these methods, the results from simulations, and extensions to locally quadratic regression.

However, the problems engendered due to data sparsity can be resolved by ensuring that all neighbourhoods contain sufficient data. Using nearest neighbour bandwidth selection the bandwidth $b(x)$ is chosen so that the local neighbourhood will always contain a specific number of points. To compute the nearest neighbour bandwidth $b(x)$, a bandwidth parameter $\alpha > 0$ needs to be considered. The parameter α gives the proportion of observations that are used to find the smoothed estimate and as α increases $\hat{f}(x)$ becomes smoother. Following Cleveland & Loader (1996), Loader (1999a), and Irizarry (2001) when $0 < \alpha \leq 1$ let q be equal to αn truncated to an integer. Then $\Delta_i(x) = |x - x_i|$ is the distance between fitting point x and the data points x_i and $\Delta_{(j)}(x)$ is the value of these distances ordered from largest to smallest. Finally, the nearest neighbourhood bandwidth $b(x) = \Delta_{(q)}(x)$, where $\Delta_{(q)}(x)$ is the q th largest value of $\Delta_{(i)}(x)$, $i = 1, \dots, n$. For $\alpha > 1$, the nearest neighbour bandwidth is $b(x) = \alpha^{1/p} \Delta_{(m)}(x)$ where p is the number of numeric independent variables involved in the fit. Cleveland & Devlin (1988) propose using an M plot, an adaptation of the C_p procedure developed by Mallows (1966, 1973) to select the bandwidth parameter, α , of nearest neighbour fitting. The M plot is a graph of an estimate of the expected mean squared error summed over the x_i in the sample and divided by σ^2 , \hat{M}_α , against the contribution of variance, V_α , for a selection of values of α . Cleveland and Devlin admit that the M plot is not a definite method of determining the span but because the M plot graphically represents the tradeoff between variance and bias as the bandwidth parameter changes, it can aid in choosing the amount of smoothing required for a data set.

Another method for finding an appropriate span is graphical diagnostics. Lack of fit due to under smoothing or over smoothing can often be detected using graphical representations to compare the fit of models with varying magnitudes of spans. After the initial examination of the fitted curves and the original data it is essential to refer to other graphical aids to view the whole picture of a data set. Cleveland & Devlin (1988) suggest that other diagnostic procedures for regression models (see Daniel & Wood, 1971; Belsley *et al.*, 1980; Cook & Weisberg, 1982; Chambers *et al.*, 1983) can be applied to local regression. A normal probability plot of $|\varepsilon_i|$ can be made to check the normality assumption, a plot of $|\varepsilon_i|$ against \hat{y}_i can be made to check the assumption of constant variance, and a plot of $|\varepsilon_i|$ against the independent variables can be used to check for bias in the model.

Selection of an appropriate bandwidth has a critical effect on the smoothed LOESS estimate $\hat{f}(x)$. The advantage of more formal model selection criteria such as C_p or cross-validation is that they are automated. The disadvantage is that these methods can easily give bandwidths that are not optimal depending on the data set they are applied to. Graphical displays are powerful tools that show the bias-variability tradeoff readily and aid in making decisions on the importance of each factor. However, as noted by Cleveland & Loader (1996), graphical diagnostics are labour intensive and therefore not the most efficient method for finding the optimal smoothed estimate. In practise, model selection criteria can be used to decide on a fit and that fit should then be analysed using graphical representations to study the goodness of fit. The importance of graphical representations when completing LOESS smoothing and choosing an appropriate value for the span cannot be emphasised enough – both methods are more powerful when used in conjunction.

4.3. Polynomial degree selection

Another important decision when fitting a LOESS model is the choice of polynomial degree, λ , which also affects the bias-variability tradeoff. Polynomials with higher degrees will produce estimates that are generally less biased but more variable than their lower degree counterparts. In R² the *loess()* function

² R is a free command line statistical software and implementation of S that can be downloaded from www.r-project.org

allows for the fitting of polynomials of degree 0, 1, and 2. Cleveland & Loader (1996) state that polynomials with degree 0, or locally constant fits, are rarely if ever the best choice because they cannot exactly fit any mathematical curve except for a straight line, which can only be fit in special cases. Loader (1999a) affirms that the most common choices in practise are local linear, $\lambda = 1$, and local quadratic, $\lambda = 2$, fits. By using polynomials of degrees greater than 0, the bandwidth can be increased without introducing intolerable bias. Higher-order polynomials have a larger number of coefficients to estimate, resulting in higher variance. However, the final fit will be smoother due to a larger neighbourhood size. In practice, fitting local polynomials of third degree and higher is rarely beneficial. Often a choice between a local linear fit and a local quadratic fit, the default in R, can be made by examining the scatterplot of the data (see Jacoby, 2000). A LOESS modification was integrated into the SC-CR algorithm that was programmed in R. The program was then used to model the trypanosomiasis data.

5. Application to a Real Data Set

The *Trypanosoma brucei* is a parasite that causes the rare disease *African trypanosomiasis*, colloquially referred to as African sleeping sickness. The parasite is transferred to humans by the bite of the tsetse fly. A person can be infected for long periods without exhibiting any symptoms. Indeed, when symptoms do appear, the disease is often already at an advanced stage of development. There are two forms that the disease can take: the neurological form (N) and the lymphatic-sanguine (LS) form. These will comprise the two competing modes of failure, where the failure time, t_i , is the age at which a person is first infected by the parasite. This diagnosis phase is crucial because it determines the appropriate type of treatment, which is highly toxic.

The data is from the *Bulletin Épidémiologique Hebdomadaire* that was published by the Institut de Veille Sanitaire as found in Legros *et al.* (2006), and is summarised in Table 4. There were 26 cases of the disease reported in France during the years 1980 and 2004. Due to the long incubation period where the disease was not detected, the exact age at infection was not always known precisely, resulting in numerous left-censored observations. Also, in few cases, the exact form of the disease (N or LS) was not able to be determined. This resulted in observations that can be viewed as masked competing risks. As such, the SC-CR algorithm of Adamic (2010) can therefore be used to model the cumulative incidence failure times under these two forms of the trypanosomiasis disease. The SC-CR algorithm was run on the data, assuming an ε of 0.0000001. The results are given in the second and third columns of Table 5.

Note that <50% of the possible data points have cause-specific innermost intervals for the two competing risks. Smoothing these two vectors of probability mass would not be advisable since there are so many gaps between the innermost intervals, making the probability landscape sparse and multimodal. Use of the LOESS modification at each iteration of the SC-CR algorithm is therefore opportune. To implement the smoothing modification, a LOESS function with a span of 1.6 and degree of 2 was used. Independent local fits for each risk seemed justified, since the correlation coefficient of the two competing densities before smoothing was only -0.1473877 . Thus, a multivariate LOESS modification was not necessary in this case. The converged results after 32 iterations are also shown in Table 5, where an ε of 0.00000001 was again used.

The LOESS modification to the SC-CR algorithm produced revealing results. Note that the gaps between the innermost intervals were successfully filled in an optimal manner. In turn, this produces valuable information that was not previously available. First, there are obviously non-zero probabilities of decrement at each of the possible ages in reality. This is immediately rectified once the LOESS modification is employed. Second, the magnitudes of the failure probabilities are

Table 4. Summary of the trypanosomiasis data.

Case	Year	Age at infection	Form of disease
1	1980	[20,20]	N
2	1981	(0,29]	LS
3	1981	[29,29]	LS
4	1981	(0,32]	LS
5	1981	[43,43]	N
6	1982	(0,19]	LS
7	1982	[23,23]	N
8	1983	(0,40]	LS or N
9	1984	[24,24]	N
10	1984	(0,31]	N
11	1986	[13,13]	N
12	1988	[2,2]	N
13	1988	(0,34]	N
14	1990	[27,27]	N
15	1991	[34,34]	N
16	1992	(0,62]	LS or N
17	1992	(0,50]	N
18	1993	(0,30]	N
19	1993	[27,27]	LS
20	1993	[30,30]	N
21	1996	[53,53]	N
22	1997	(0,53]	LS
23	1998	[45,45]	LS
24	1999	(0,50]	LS
25	1999	(0,34]	LS or N
26	2001	[28,28]	N

Note: N, neurological form; LS, lymphatic-sanguine form.

themselves revealing. For example, consider age 50. Before the LOESS modification, it was not known whether failure due to cause N or cause LS was more likely (since they were both 0). Now, we can estimate that infection at age 50 is over twice as likely to be due to form N than form LS. Figure 1 depicts the final CIF curves for each of the two competing risks.

6. Conclusion and Future Work

There are other advantages that can be realised from using the LOESS modification to the SC-CR algorithm. First, local regression adapts well to bias problems at boundaries and in regions of high curvature. We note that the issue of larger bias introduced from arbitrarily placing probability mass at the right end of the innermost intervals, a dynamic outlined in Duchesne & Stafford (2001), is even more relevant in the competing risks framework due to the larger number of innermost intervals involved. The LOESS modification to the SC-CR algorithm is unarbitrarily data driven, thereby making the cause-specific innermost intervals, and the problems they engender, irrelevant, as was also stated in Braun *et al.* (2005) for kernel smoothing. Second, the use of the LOESS modification tends to reduce the number of iterations required for convergence. Many methods have been developed to provide fast computation of a smoothed estimate for one or more independent variables, partially because of the conceptual simplicity and excellent theoretical properties of local regression, a dynamic

Table 5. Converged failure rates for the trypanosomiasis data set.

t_i	$\hat{f}^{(N)}(t_i)$	$\hat{f}^{(LS)}(t_i)$	$\check{f}^{(N)}(t_i)$	$\check{f}^{(LS)}(t_i)$
2	0.11176990	0.00000000	0.09532087	0.00000000
13	0.11176990	0.00000000	0.08710922	0.00066400
19	0.00000000	0.00000000	0.07929145	0.00542837
20	0.07451327	0.00000000	0.07187282	0.00948088
23	0.07451327	0.00000000	0.06485857	0.01279662
24	0.07451327	0.00000000	0.05838599	0.01534092
27	0.07451327	0.07451327	0.05250641	0.01712799
28	0.07451327	0.00000000	0.04709622	0.01820631
29	0.00000000	0.07451327	0.04203177	0.01862433
30	0.06773934	0.00000000	0.03718945	0.01843050
31	0.00000000	0.00000000	0.03252463	0.01765627
32	0.00000000	0.00000000	0.02814704	0.01636821
34	0.05419147	0.00000000	0.02416456	0.01466751
40	0.00000000	0.00000000	0.02068509	0.01265536
43	0.04563492	0.00000000	0.01781653	0.01043294
45	0.00000000	0.04563492	0.01554459	0.00800064
50	0.00000000	0.00000000	0.01377329	0.00527441
53	0.04166667	0.00000000	0.01248798	0.00222874
62	0.00000000	0.00000000	0.01167401	0.00000000

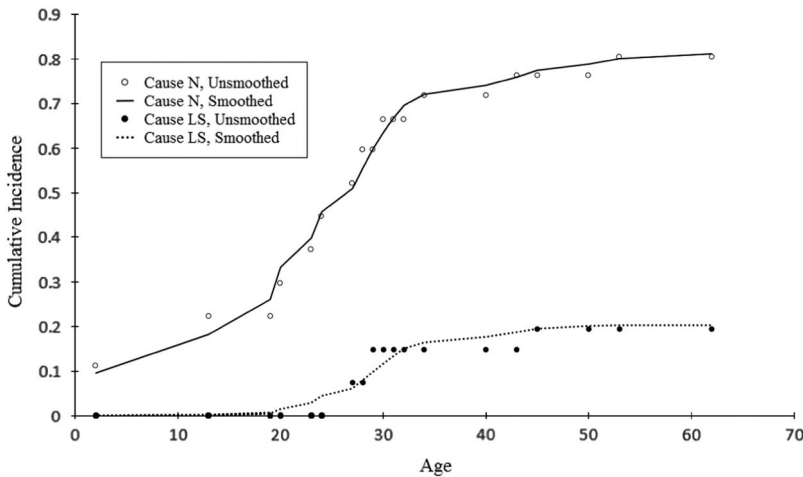


Figure 1. Converged cumulative incidence functions for the trypanosomiasis data. N, neurological form; LS, lymphatic-sanguine form.

that also allows it to work for many different distributional assumptions. Finally, local regression does not require the smoothness and regularity conditions required by other smoothing methods, such as boundary kernels. These and other strengths of local regression, as outlined in Hastie & Loader (1993), make local regression an attractive supplement to the SC-CR algorithm.

In terms of future work directly related to the LOESS modification, we note that the modification can be applied, not only to left and right-censored failures, but to other censoring mechanisms as well. It

can be applied to a variety of self-consistent models that include left-censored, doubly-censored, truncated, and other more elaborate censoring schemes that are very much at the forefront of contemporary scholarship in survival analysis.

Acknowledgement

This work was supported by the Government of Canada: Natural Sciences and Engineering Research Council of Canada under Discovery Grant No. 356028-2010 RGPIN.

References

- Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, **3**, 15–27.
- Adamic, P. (2010). Modeling multiple risks in the presence of double censoring. *Scandinavian Actuarial Journal*, **1**, 68–81.
- Adamic, P. (2012). A model for time-dependent competing risks. Contributed paper presented at the 40th Annual Meeting of the Statistical Society of Canada, June 4, 2012, University of Guelph, Guelph, Ontario, Canada.
- Adamic, P., Dixon, S. & Gillis, D. (2010). Multiple decrement modeling in the presence of interval censoring and masking. *Scandinavian Actuarial Journal*, **4**, 312–327.
- Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). *Regression Diagnostics*. John Wiley, New York.
- Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. & Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd edition. The Society of Actuaries, Schaumburg, IL.
- Braun, J., Duchesne, T. & Stafford, J.E. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics*, **33**(1), 39–60.
- Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). *Graphical Methods for Data Analysis*. Wadsworth, Monterey, CA.
- Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.
- Cleveland, W.S. & Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596–610.
- Cleveland, W.S. & Loader, C.L. (1996). Smoothing by local regression: principles and methods. In W. Härdle & M.G. Schimek (Eds.), *Statistical Theory and Computational Aspects of Smoothing* (pp. 10–49). Springer, New York.
- Cook, R.D. & Weisberg, S. (1982). *Influence and Residuals in Regression*. Chapman & Hall, New York.
- Cox, D.R. (1959). The analysis of exponentially distributed lifetimes with two types of failures. *Journal of the Royal Statistical Society, Series B*, **21**, 411–421.
- Craiu, R.V. & Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Life Time Data Analysis*, **12**, 21–33.
- Daniel, C. & Wood, F. (1971). *Fitting Equations to Data*. John Wiley, New York.
- Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.
- Dinse, G.E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics*, **38**, 417–431.
- Duchesne, T. & Stafford, J.E. (2001). A kernel density estimate for interval censored data, technical report no. 0106, University of Toronto, Toronto, Ontario, Canada.
- Efron, B. (1967). The two-sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: University of California Press*, **4**, 831–853.

- Escarela, G. & Carriere, J.F. (2003). Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, **12**(4), 333–349.
- Finkelstein, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, **42**, 845–854.
- Flehinger, B.J., Reiser, B. & Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika*, **85**, 151–164.
- Gasser, T., Kneip, A. & Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, **86**, 643–652.
- Gichangi, A. & Vach, W. (2005). The analysis of competing risks data: a guided tour. *Statistics in Medicine*, **132**(4), 1–41.
- Goutis, C. (1997). Nonparametric estimation of a mixing density via the kernel method. *Journal of the American Statistical Association*, **92**, 1445–1450.
- Groeneboom, P., Maathuis, M.H. & Wellner, J.A. (2008). Current status data with competing risks: consistency and rates of convergence of the MLE. *The Annals of Statistics*, **36**, 1031–1063.
- Hastie, T. & Loader, C.R. (1993). Local regression: automatic kernel carpentry. *Statistical Science*, **8**, 120–129.
- Huang, J. (2000). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, **9**, 501–519.
- Hudgens, M.G., Satten, G.A. & Longini, I.M. Jr (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, **57**, 74–80.
- Irizarry, R.A. (2001). Chapter 3: local regression. Available online at the address <http://www.biostat.jhsph.edu/~ririzarr/Teaching/754/section-03.pdf> [accessed Jun-2015].
- Jacoby, W.G. (2000). LOESS: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, **19**, 577–613.
- Jewell, N.P., Van Der Laan, M. & Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika*, **90**, 183–197.
- Kalbfleisch, J.D. & Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- Klein, J.P. & Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.
- Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. John Wiley and Sons, New York.
- Legros, F. & Ancelle, T. (2006). Trypanosomiase humaine africaine: recensement des cas d'importation observés en France, 1980–2004. *Bulletin épidémiologique hebdomadaire*, **7**, 57–59.
- Li, J. & Yu, Q. (2014). A consistent NPMLE of the joint distribution function with competing risks data under the dependent masking and right-censoring model. *Life Time Data Analysis*, **22**, 63–99.
- Li, L., Watkins, T. & Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, **24**, 531–542.
- Loader, C.R. (1999a). *Local Regression and Likelihood*. Springer-Verlag, New York.
- Loader, C.R. (1999b). Bandwidth selection: classical or plug-in? *The Annals of Statistics*, **27**, 415–438.
- Mallows, C.L. (1966). Choosing a subset regression. Presented at the Annual Meeting of the American Statistical Association, Los Angeles, California, August 15–19.
- Mallows, C.L. (1973). Some comments on C_p . *Technometrics*, **15**, 661–675.
- Pan, W. (2000). Smooth estimation of the survival function for interval censored data. *Statistics in Medicine*, **24**, 531–542.
- Park, C. & Padgett, W.J. (2006). Analysis of strength distributions of multi-modal failures using the EM algorithm. *Journal of Statistical Computation and Simulation*, **76**, 619–636.

- Ruppert, D., Sheather, S.J. & Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **901**, 1257–1270.
- Stafford, J. (2005). “Exact” local likelihood estimators. Available online at the address <http://www.gla.ac.uk/external/RSS/RSScomp/Stafford.pdf> [accessed Jun-2015].
- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences USA*, **72**, 20–22.
- Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, **69**, 169–173.
- Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society*, **38**, 290–295.
- Wang, J., Yu, Q. & Wong, G.Y.C. (2012). The random partition masking model for interval-censored and masked competing risks data. *Journal of Statistical Computation and Simulation*, **82**, 981–1002.
- Yu, Q. & Li, J. (2012). The NPMLE of the joint distribution function with right-censored and masked competing risks data. *Journal of Nonparametric Statistics*, **24**, 753–764.
- Yu, Q., Li, L. & Wong, G.Y.C. (2000). On consistency of the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, **27**, 35–44.
- Zhou, M. (2004). Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Statistica Sinica*, **14**, 533–546.

Appendix

Note that the definition and notation for univariate self-consistency, as found in the first two lines of the proof, come from Braun *et al.* (2005).

Proof for the self-consistency of the CIFs derived from the SC-CR algorithms

$$\begin{aligned}
 \hat{F}_K(x) &= E_{K-1}[F_n(x)|\Psi] \\
 &= E_{K-1}\left[\frac{1}{n}\sum_{i=1}^n I(X_i \leq x)|\Psi\right] \\
 &= E_{K-1}\left[\frac{1}{n}\sum_{i=1}^n \sum_j I(X_i^{(j)} \leq x)|\Psi\right] \\
 &= \sum_j E_{K-1}\left[\frac{1}{n}\sum_{i=1}^n I(X_i^{(j)} \leq x)|\Psi\right] \\
 &= \sum_j \hat{F}_K^{(j)}(x), \text{ if and only if } \hat{F}_K^{(j)}(x) \text{ is a self-consistent estimator of } F_K^{(j)}(x) \\
 &= F_K^{(j)}(x), \text{ a property of the SC-CR algorithm}
 \end{aligned}$$

Thus, $\hat{F}_K^{(j)}(x)$ must be a self-consistent estimator of $F_K^{(j)}(x)$, otherwise a contradiction would ensue.