View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Diposit Digital de Documents de la UAB

SORT 31 (1) January-June 2007, 75-96

Nonparametric bivariate estimation for successive survival times

Carles Serrat* and Guadalupe Gómez**

Universitat Politècnica de Catalunya

Abstract

Several aspects of the analysis of two successive survival times are considered. All the analyses take into account the dependent censoring on the second time induced by the first. Three nonparametric methods are described, implemented and applied to the data coming from a multicentre clinical trial for HIV-infected patients. Visser's and Wang and Wells methods propose an estimator for the bivariate survival function while Gómez and Serrat's method presents a conditional approach for the second time given the first. The three approaches are compared and discussed at the end of the paper.

MSC: 62N02, 62G05

Keywords: AIDS; Conditional survival; Dependent censoring; Inverse probability of censoring weighted estimators; Successive survival times.

1 Introduction

The survival experience of a population often involves two times of interest. The estimation of their joint survivor function is of intrinsic interest since it is useful in predicting the joint survival experience, in estimating the degree of dependence, in model building and testing and in strengthening marginal analysis.

* *Address for correspondence:* Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Avda. Dr. Marañón, 44–50, 08028 Barcelona, Spain. Tel: +34 93 405 44 86. Fax +34 93 401 63 06. E-mail: carles.serrat@upc.edu

** Guadalupe Gómez. Departament d'Estadística i I.O., Universitat Politècnica de Catalunya, Edifici C5, Despatx 210, Jordi Girona 1–3, 08034 Barcelona, Spain. E-mail: lupe.gomez@upc.edu

Received: November 2006
Accepted: February 2007

These two survival times of interest could be naturally paired, for instance, in twin studies when analyzing time to death of each sibling, in oncology studies when the interest is the time to cancer detection in the left and right breast, in ophthalmology when recording the time to severe visual loss on the left and right eyes. In these cases, several possibly dependent failure processes act concurrently, and henceforth, they ought to be modelled jointly.

In many other situations there is a natural ordering of the times of occurrence of events. For instance, in any clinical study, time to diagnosis precedes time to start treatment which in turns precedes times to cure. In AIDS studies, time to HIV infection precedes the time to AIDS diagnosis, which in turn precedes the time to death due to AIDS. In food science studies, when referring to climacteric fruits, the time to maturation precedes the time to senescentia.

In univariate survival studies, right censoring usually precludes the complete observation of the time to event variable. When we have two survival times of interest the censoring mechanism could either be the same for both variables or act differently on each one. For instance, when analyzing the joint behaviour of the ages of cancer diagnosis in each breast the censoring – due to loss of follow-up or end of study – acts simultaneously on each breast. However, when studying the population of twins who have suffered a heart attack, the follow-up time, and hence the censoring, could be independent for each twin. When one event precedes the second, the censoring mechanism acting on the second and subsequent times will depend not only on the total time of follow-up but also in the value of the first and preceding times. When this situation arises the methods to estimate the joint survival, or functions of the joint, have to handle the special case of dependent and informative censoring induced by the previous failure times.

The motivation of the paper comes from the Tibet (Guided-**T**reatment **I**nterruption **Be**nefi**t**) study. Tibet is a multicentre, open label clinical trial with blinded and centralized randomization conducted to investigate the safety and clinical benefits of an intermittent antiretroviral therapy guided by CD4+ T-cell counts and plasma HIV-1 RNA in patients with chronic HIV-1 infection with more than 500 CD4+ counts/mm^3 and undetectable HIV-1 RNA. Patients were randomized to follow either the intermittent guided therapy which is described below or to continue with their prior HAART (Highly Active Antiretroviral Therapy). Details of the study are described in Ruiz *et al* (2007).

In this work we restrict our attention to the interruption group in which the patients interrupt therapy until CD4+ counts reach values equal or inferior to 350 cells/mm^3, plasma viral load increases to 100000 copies/ml or a severe acute retroviral syndrome takes place or an AIDS-defining illness occurs. If any of these events occurs, the prior HAART is reinitiated and maintained until CD4+ counts increases to 500 cells/mm^3 or more and viral load reaches 80 copies/ml, at which time HAART is again discontinued as previously described.

This intermittent therapeutic strategy process defines, for each patient, a sequence of alternative stages without HAART (OFF) and with HAART (ON). Various lifetime variables can be defined within this process, for instance, T_1 is the first time OFF, that is, the time (in weeks) from randomization (and therefore interruption of HAART therapy) to first reinitiation of treatment, T_2 is the first time ON, that is the time from the first treatment reinitiation until the next interruption, T_3 is the second time OFF, that is, the number of weeks from the second HAART interruption to the second reinitiation of treatment, and so on.

Apart from the number of scientific and clinical questions that such an study poses, there are also a number of relevant statistical issues which arise due to this particular data set. In particular, the dependent censoring mechanism that affects T_2, and subsequent times, as a consequence of having an administrative censoring time C, invalidates the standard methods of survival analysis for right-censored data and requires alternative approaches.

We now review the most relevant papers concerning the estimation of the bivariate survival function as well as the estimation of the conditional survival function from pairs of random variables which might be right-censored. Campbell (1981) and Campbell and Földes (1982) propose several nonparametric estimators for the bivariate survival function in the presence of independent pairs of censoring variables that are independent of the failure times. Their main idea is based on the factorization of the bivariate distribution function as a product of the distribution function for the bivariate vector of interest and the distribution function for the censoring variables. These estimators are shown to be strongly uniform consistent at a rate of convergence equal to that of the empirical distribution function. All the estimators they propose, however, are not legitimate survival functions since they are not necessarily monotone increasing in both coordinates. Tsai, Leurgans and Crowley (1986) propose a family of closed form estimators that are always survival functions based on a decomposition of the bivariate survival in terms of identifiable survival and subsurvival functions. Their estimators are fairly complicated and have a rate of convergence slower than Campbell and Földes. Burke (1988) proposes an estimator based on the representation of the bivariate distribution function as the convolution of the subdistribution function, which can be naturally estimated by the observed data, and the inverse of the bivariate distribution function for the censoring times. Burke's approach only uses the information provided by the uncensored observations, throwing away the relevant information of censored data points. Dabrowska (1988) proposes an estimator for the bivariate survival function based on an empirical estimator for the bivariate cumulative hazard. This estimator is almost surely consistent and weakly convergent. Unfortunately, its computation is quite complicated and the covariance function of these estimators cannot be estimated analytically. More details can be found in Gómez *et al.* (2004).

Visser (1996), Wang and Wells (1998) and Gómez, Serrat and Ruiz (2007) approach different aspects of the nonparametric bivariate survival estimation problem. Visser

derives the nonparametric maximum likelihood estimate for the conditional hazard of T_2 given a fixed value of T_1 under the assumption that the two durations are discrete. Wang and Wells present an estimator for the cumulative conditional hazard of T_2 given $T_1 > t_1$ following Nelson-Aalen's construction of the cumulative hazard estimator but where each observation has been weighted using the information on the first duration to unbias the effect of dependent censoring. Due to the limited applicability of Visser's estimator, since lifetime data are genuinely continuous, and the lack of interpretability of Wang and Wells's parameter of interest in the case of two ordered times T_1 and T_2 where the observation of the second time, T_2, is conditioned on the observation of the first time, Gómez, Serrat and Ruiz (2007) propose a weighted conditional estimation for the survival of T_2 on a given category of T_1.

We introduce the notation and assumptions for the rest of the paper in Section 2 and develop the methods of Visser, Gómez and Serrat and Wang and Wells in Sections 3, 4 and 5 respectively. Each of these three sections starts with a description of the method, continues with some software considerations and ends with a specific analysis of the Tibet clinical trial. The three approaches are compared and discussed at the end of the paper.

2 Notation

Assume that T_1 and T_2 represent two consecutive duration variables corresponding to two different events at times T_1 and $T_1 + T_2$, respectively, which are measured from the start of the follow-up. The follow-up time is subject to independent right censoring by C. Note that T_1, T_2 and $T_1 + T_2$ are independent of C. However T_2, which is subject to right censoring by $C - T_1$, is not independent of $C - T_1$ unless T_1 is independent of T_2. In this situation, we cannot use conventional survival methods for independent or noninformative censorship models. Whenever the censoring random variable for a given time depends on other random times we say that we are in the framework of a dependent censoring mechanism.

Define the marginal and bivariate survival functions for (T_1, T_2) as $S_1(t_1) = \Pr\{T_1 > t_1\}$ and $S_{12}(t_1, t_2) = \Pr\{T_1 > t_1, T_2 > t_2\}$. Denote by $G(t) = \Pr\{C > t\}$ the survival function corresponding to the total time of follow-up C. If $\tau_C = \sup\{t : G(t) > 0\}$ is the maximum follow-up time, the bivariate survival function, $S_{12}(t_1, t_2)$, is only estimable for $t_1 + t_2 \leq \tau_C$. This restriction is analogous to the non-estimability of the Kaplan-Meier estimator beyond those values larger than the total follow-up time. It follows as well that the marginal distributions for T_2 cannot be estimated by the Kaplan-Meier method. Note that if T_1 and T_2 are positively correlated, even under independent censoring, persons with long T_1's are more likely to have long T_2's and hence more likely to be censored.

For a given individual we observe a vector (Y_1, Y_2, D_1, D_2) where for every $j = 1, 2$, $Y_j = \min\{T_j, C_j\}$, $D_j = 1\{T_j \leq C_j\}$, $C_1 = C$, $C_2 = (C - T_1)1\{T_1 \leq C\}$.

Note that when

i. $D_1 = 0 = D_2$: the two durations are right-censored and thus $Y_1 = C$, $Y_2 = 0$ and no information about T_2 is available

ii. $D_1 = 1, D_2 = 0$, T_1 is observed while T_2 is right-censored by $C - T_1$, which implies that T_2 is right-censored by a dependent variable if T_1 and T_2 are correlated.

iii. $D_1 = 1, D_2 = 1$, T_1 and T_2 are observed.

Our estimation problem is to be based on a random sample $\{(T_{1i}, T_{2i}, C_i), i = 1, \ldots, n\}$ of (T_1, T_2, C) from which the observed sample is $\mathcal{S} = \{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}),\ i = 1, \ldots, n\}$. We also consider $\mathcal{S}^*$, a subset of $\mathcal{S}$, consisting of those observations for which T_1 is observed, that is, $\mathcal{S}^* = \{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}) \in \mathcal{S} | D_{1i} = 1, \quad i = 1, \ldots, n\} \subset \mathcal{S}$.

Note that when $D_{1i} = 0$, no crude information about T_{2i} is available. However, these subjects provide information about T_1, which is supposed to be dependent on T_2. Thus, these missing data ($\{i : D_{1i} = 0\}$), are not at random because the probability of being observed for T_2 depends on T_1. As a consequence, inferences for T_2 cannot be only based on the subset $\mathcal{S}^*$, and we will have to use these partially observed individuals to infer about the law of T_2.

3 Visser's method. A discrete approach

3.1 Introduction to the methodology

Visser (1996) proposes a nonparametric estimator for the bivariate survival function when the two duration variables are always observed in a particular order and the censoring mechanism acts on their sum.

Visser starts assuming that T_1, T_2 and C are discrete random variables taking values in $\{0, 1, 2, \ldots, K\}$, and therefore Y_1, Y_2, defined in Section 2 as $Y_j = \min\{T_j, C_j\}$, $j = 1, 2$, are discrete as well. Due to the fact that the random variables T_1, T_2 and C are supposed to be discrete and take a finite number of values, Visser defines the corresponding survival distributions at each time t as the probability of being greater or equal than t as follows

$$\begin{aligned} S_{T_1,T_2}(k, l) &= \Pr\{T_1 \geq k, T_2 \geq l\} \\ S_{T_1}(k) &= \Pr\{T_1 \geq k\} \end{aligned}$$

$$\begin{aligned} \lambda_{T_1}(k) &= \Pr\{T_1 = k | T_1 \geq k\} \\ G(k) &= \Pr\{C \geq k\}. \end{aligned}$$

Visser factorizes $S_{T_1,T_2}(k,l)$ as the product of the conditional and the marginal as follows,

$$S_{T_1,T_2}(k,l) = S_{T_1}(k) S_{T_2|T_1}(l|k) \tag{1}$$

On the other hand, the product expression of the survival functions in terms of the hazard functions allows to write for $k, l = 1, 2, \ldots, K$:

$$S_{T_1}(k) = (1 - \lambda_{T_1}(0)) \ldots (1 - \lambda_{T_1}(k-1)) \tag{2}$$

$$S_{T_2|T_1=k}(l) = \Pr\{T_2 \geq l | T_1 = k\} = (1 - \lambda_{T_2|T_1=k}(0)) \ldots (1 - \lambda_{T_2|T_1=k}(l-1)) \tag{3}$$

where $\lambda_{T_2|T_1=k}(l) = \Pr\{T_2 = l | T_1 = k, T_2 \geq l\}$.

Remark as well that $S_{T_2|T_1}(l|k)$ can be written as follows:

$$\begin{aligned} S_{T_2|T_1}(l|k) &= \Pr\{T_2 \geq l | T_1 \geq k\} = \frac{\Pr\{T_2 \geq l, T_1 \geq k\}}{\Pr\{T_1 \geq k\}} \\ &= (S_{T_1}(k))^{-1} \sum_{j=k}^{K} S_{T_2|T_1=j}(l)(S_{T_1}(j) - S_{T_1}(j+1)) \end{aligned} \tag{4}$$

Equalities (1) and (4) imply that in order to estimate $S_{T_1,T_2}(k,l)$ we only need to estimate $S_{T_1}(k)$ and $S_{T_2|T_1=j}(l)$. The estimation of $S_{T_1}(k)$ is straightforward through the Kaplan-Meier estimator.

Denote by n_{1k}, n_{2kl}, n_{3kl} the following counting processes: $n_{1k} = \sum_{i=1}^{n} 1\{ Y_{1i} = k, \delta_i = 1\}$, $n_{2kl} = \sum_{i=1}^{n} 1\{Y_{1i} = k, Y_{2i} = l, \delta_i = 2\}$ and $n_{3kl} = \sum_{i=1}^{n} 1\{Y_{1i} = k, Y_{2i} = l, \delta_i = 3\}$. That is, n_{1k} counts the number of censored individuals at k months (for these individuals $T_1 > k$ and T_2 is not defined), n_{2kl} counts the number of individuals whose first duration is equal to k months and who are censored after $k + l$ months (for these individuals $T_1 = k$ and $T_2 > l$) and n_{3kl} counts the number of subjects with a first duration equal to k months and a second duration equal to l months (for these individuals $T_1 = k$ and $T_2 = l$). Denote as well $n_{k\cdot} = \sum_{l=1}^{K}(n_{2kl} + n_{3kl})$ which counts the total number of individuals whose $T_1 = k$ irrespective of their status on T_2.

Visser proves (see Appendix for more details) that the nonparametric MLE for $\lambda_{T_1}(k)$ is given by

$$\hat{\lambda}_{T_1}(k) = \frac{\sum_{i=1}^{n} 1\{Y_{1i} = k, \delta_i \geq 2\}}{\sum_{i=1}^{n} 1\{Y_{1i} \geq k\}} = \frac{\sum_{l=1}^{K}(n_{2kl} + n_{3kl})}{n_{1k} + \sum_{l=1}^{K}(n_{2kl} + n_{3kl})} \tag{5}$$

which yields the discrete time Kaplan-Meier estimator for $S_{T_1}(k)$ after replacing it in (2). On the other hand, the nonparametric MLE for $\lambda_{T_2|T_1=k}(l)$ is given by

$$\hat{\lambda}_{T_2|T_1=k}(l) = \sum_{i=1}^{n} \frac{1\{Y_{1i} = k, Y_{2i} = l, \delta_i = 3\}}{\sum_{i=1}^{n} 1\{Y_{1i} = k, Y_{2i} \geq l\}}. \tag{6}$$

Replacing $\hat{\lambda}_{T_2|T_1=k}(l)$ in (3) provides the MLE for $S_{T_2|T_1=k}(l)$, which in turn can be replaced in (4) to obtain an estimator for $S_{T_2|T_1}(l|k)$. Finally, everything could be replaced in (1) to get the bivariate nonparametric estimator for $S_{T_1,T_2}(k,l)$.

Visser proves that both estimators, $\hat{\lambda}_{T_1}(k)$ and $\hat{\lambda}_{T_2|T_1=k}(l)$, are consistent asymptotically normal after normalizing by $\sqrt{n}$, and asymptotically independent. These facts, together with the δ method, imply that $\sqrt{n}(\hat{S}_{T_1,T_2}(k,l) - S_{T_1,T_2}(k,l))$ is asymptotically normal, mean zero and with an asymptotic variance that can be estimated replacing the unknown functions by their estimators.

The survival function G of the censoring variable appears in the expression for the variances. It may be estimated by the product-limit method.

3.2 Implementation

We have implemented in S-PLUS a function `bwv21` that computes the conditional survival for T_2, given a value $T_1 = t_1$, according to expression (3). The function uses as parameters the observed values of T_1 and T_2, as well as, the corresponding censoring indicators (D_1 and D_2).

After estimating the conditional survival we can compute the joint survival $S_{T_1,T_2}(k,l)$ in (1), by using the function `c2jv` that implements the expression given in (4).

All the S-PLUS functions that we have implemented are available at the web page of the GRASS group at `http://www-eio06.upc.es/grass`.

3.3 TIBET project: A discrete time analysis

In the Tibet clinical trial, one hundred HIV-patients were recruited between May 2001 and January 2002 and randomly assigned to interrupt HAART. The interim closing date for the study was July 15, 2004.

Figure 1 shows the empirical survival estimator corresponding to the follow-up time of each patient. Based on this estimation, the probability of being followed 96 weeks or more is 93%, the median follow-up time is 130 weeks, the third quartile is 146 weeks and the maximum follow-up time being 188 weeks. Furthermore, the effective minimum follow-up has been 96 weeks.

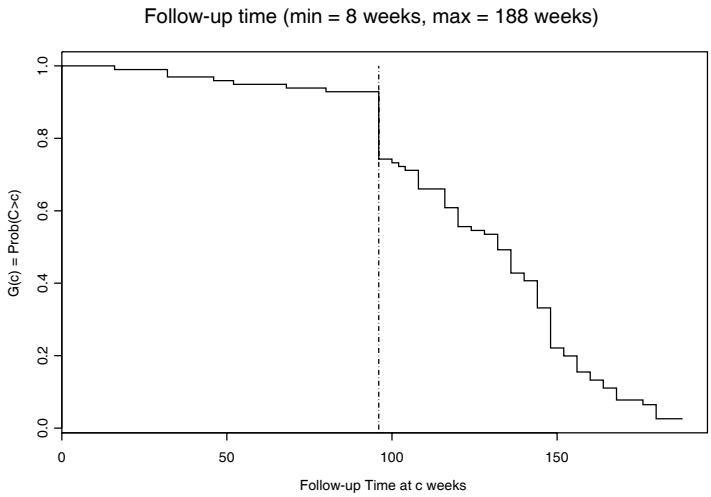

Figure 1: *Survival estimator for the time to follow-up.*

3.3.1 Conditional estimation of the first time ON given the first time OFF

Using the function `bwv21` we have obtained the results in Table 1. This table illustrates the finite grid of 30 observed times for T_1 by 23 observed times for T_2, and the corresponding estimation of $S_{T_2|T_1=t_1}(t_2)$ for each pair (t_1, t_2) of the grid. For instance, the median time of being with treatment among those who have been 8 weeks interrupted is approximately 23 weeks. Note, however, that although the median and even the third quartile are estimable for $t_1 = 8$, for longer interrupted times the estimation is either not possible or quite rough.

As a matter of fact, Visser's discrete method does not provide efficient estimates of the conditional survival function due to the drastic reduction of the sample size. This drawback is due to the fact that for a fixed time $T_1 = t_1$ the sample size is not large enough to make inferences on $T_2|T_1 = t_1$. In particular, the sample size is dramatically small for $t_1 > 40$. Furthermore, the small number of events for T_2 makes the estimation of $S_{T_2|T_1=t_1}(.)$ quite hopeless. This fact is still more problematic for high values of t_1, where the proportion of censoring for T_2 is higher (in some cases 100%).

3.3.2 Joint distribution estimation

Based on the results in Table 1 we have computed the joint survival $S_{T_1,T_2}(k, l)$ in (1). Table 2 shows the results of this estimation for a selection of times in T_1 and T_2.

It is important to remark that expression in (4) can not be directly computed from the data because, as we have seen in Table 2, $S_{T_2|T_1=j}(l)$ is not estimable for some pairs

Table 1: *Estimates for the conditional survival of T_2 given values in T_1, $\hat{S}_{T_2|T_1=t_1}(t_2)$.*

	t_1=6	8	9	10	12	13	16	20	24	26	28	32	36	40	44	48	56	60	64	68	72	76	88	92	96	100	108	116	136	144
t_2=8	1	1	1	1	1	1	1	1	1	1	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
10	0.83	1	1	1	1	1	1	1	1	1	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
14	0.83	1	1	0.50	1	1	1	1	1	0	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
15	0.83	1	0.50	0.50	1	1	1	1	1	–	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
16	0.83	0.94	0.50	0.50	1	1	0.67	1	1	–	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
19	0.83	0.88	0.50	0.50	1	1	0.67	1	1	–	1	1	1	0.67	1	1	1	1	1	1	1	1	1	1	–	1	–	1	1	1
20	0.83	0.76	0.50	0.50	1	1	0.33	1	1	–	1	1	1	0.67	1	1	1	1	1	1	1	0	0.50	1	–	1	–	1	1	1
22	0.50	0.76	0.50	0.50	1	1	0.33	1	1	–	1	1	1	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	1	–
24	0.50	0.41	0.50	0.50	1	1	0.33	1	1	–	0.75	1	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	1	–
25	0.50	0.41	0.0	0.50	1	1	0.33	1	1	–	0.75	1	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	–	–
26	0.17	0.41	–	0.50	1	1	0.33	1	1	–	0.75	1	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	–	–
28	0.17	0.29	–	0.50	1	1	0.33	0	1	–	0.50	1	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	–	–
32	0.17	0.18	–	0.50	1	1	0.33	–	0.67	–	0.25	0	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	–	–
34	0	0.18	–	0.50	1	1	0.33	–	0.67	–	0.25	–	0.50	0.67	1	1	1	1	1	1	1	–	0.50	1	–	–	–	1	–	–
36	–	0.12	–	0.50	1	1	0.33	–	0.67	–	0.00	–	0.50	0.67	1	1	1	1	1	1	1	–	0	0	–	–	–	1	–	–
39	–	0.12	–	0.50	1	0	0.33	–	0.67	–	–	–	0.50	0.67	1	1	1	1	1	1	1	–	–	–	–	–	–	1	–	–
40	–	0.12	–	0.50	1	–	0.33	–	0.33	–	–	–	0.50	0.67	0	1	1	1	1	1	1	–	–	–	–	–	–	1	–	–
44	–	0.12	–	0.50	1	–	0.33	–	0.33	–	–	–	0.50	0.67	–	1	1	0	1	1	1	–	–	–	–	–	–	1	–	–
46	–	0.12	–	0	1	–	0.33	–	0	–	–	–	0.50	0.67	–	1	1	–	1	1	1	–	–	–	–	–	–	–	–	–
48	–	0.12	–	–	1	–	0.33	–	–	–	–	–	0.50	0.67	–	1	1	–	0	1	1	–	–	–	–	–	–	–	–	–
52	–	0.12	–	–	1	–	0.33	–	–	–	–	–	0.50	0.67	–	1	0.50	–	–	1	1	–	–	–	–	–	–	–	–	–
60	–	0.06	–	–	1	–	0.33	–	–	–	–	–	0.50	0.67	–	1	0.50	–	–	1	–	–	–	–	–	–	–	–	–	–
100	–	0.06	–	–	–	–	–	–	–	–	–	–	0.50	–	–	–	–	–	–	0	–	–	–	–	–	–	–	–	–	–

Table 2: *Estimates for some selected times of the joint survival of* (T_1, T_2) *using Visser's approach.*

	$t_1 = 6$	12	24	48	72	96	120
$t_2 = 12$	0.878	0.678	0.598	0.460	0.377	0.325	0.312
24	0.428	0.315	0.241	0.164	0.080	0.054	0.054
36	0.206	0.198	0.155	0.116	0.033	0.022	0.022
48	0.100	0.098	0.084	0.062	0.010	–	–
60	0.078	0.077	0.064	0.042	–	–	–
100	0.016	0.015	0.015	–	–	–	–

(j, l). As a first approximation we have omitted the contribution of these terms in those pairs in which the estimation of the conditional survival has not been possible. This fact produces two important drawbacks: on one hand, the method is not efficient and, on the other hand, in general there is an underestimation of the corresponding bivariate survival distribution.

4 Gómez and Serrat's method. A stratified approach

Driven by the Tibet clinical trial where it is of special interest to characterize the number of weeks on treatment that a patient needs in order to recover their virological and immunological levels given that he/she has spent a certain number of weeks without treatment, we propose an estimator for the survival of T_2 on a given category of T_1.

Let $0 < \tau_1 < \tau_2 < \ldots < \tau_M$ be the M times of interest for a particular study. For convenience define $\tau_0 = 0$ and consider τ_{M+1} as an arbitrary value larger than τ_M.
Let T_1^* be a discrete version of T_1 defined as follows:

$$T_1^* = \begin{cases} \tau_k & \text{if} \quad \tau_{k-1} < T_1 \leq \tau_k \quad k = 1, \cdots, M \\ \tau_{M+1} & \text{if} \quad T_1 > \tau_M. \end{cases}$$

Note that the election of the representative of each class is not relevant for the results. Denote the conditional cumulative hazard function for T_2 given $T_1^* = \tau_k$ by $\Lambda_{T_2|T_1^*=\tau_k}(db)$ and the conditional survival function for T_2 given $T_1^* = \tau_k$ by

$$S_{T_2|T_1^*=\tau_k}(v) = \Pr(T_2 > v|T_1^* = \tau_k) = \Pr(T_2 > v|\tau_{k-1} < T_1 \leq \tau_k)$$

for $k = 1, \ldots, M$ and $v > 0$. The factorization of the survival function in terms of the conditional cumulative hazard is straightforward:

$$S_{T_2|T_1^*=\tau_k}(v) \quad = \quad \prod_{b \leq v} \{1 - \Lambda_{T_2|T_1^*=\tau_k}(db)\}. \tag{7}$$

When estimating the survival of T_2 given a certain category of T_1 we could legitimately apply the conditional Kaplan-Meier if censoring for T_2 would be non informative, that is, if individuals with different T_1 values, within a given category, had the same chances of being at risk for different values of T_2. However, this might not be the case if the categories are quite wide and, in this case, we will have to take into account the effect that the dependent censoring caused by T_1, within each strata, is producing on T_2.

In the next subsection we show how would affect the dependent censoring on the estimations and we propose how to adjust the Kaplan-Meier survival estimates for T_2 on each strata to unbias the effect produced by T_1 and we propose a weighted conditional estimator for $S_{T_2|T_1^*=\tau_k}(t_2) = \Pr(T_2 > t_2|\tau_{k-1} < T_1 \leq \tau_k)$, adjusted by the dependent censoring.

4.1 Weighted conditional methodology

Denote by $R_{T_2}(b|\tau_k)$ the risk set of T_2 at time b given $T_1^* = \tau_k$. Under the dependent censoring structure the risk set $R_{T_2}(b|\tau_k)$ for estimating $\Lambda_{T_2|T_1^*=\tau_k}(db)$ may not be homogeneous, as is shown in Theorem 1.

Theorem 1 *The probability of being at risk at time b for the second duration T_2 for an individual with first duration equal to $T_1 = t_{1i}$ depends on $G(t_{1i} + b)$.*

Proof: An observation i with the first duration $T_1 = t_{1i}$ affects the probability of the corresponding T_{2i} being included in $R_{T_2}(b|\tau_k)$, as we see in the following expression:

$$\begin{aligned}\Pr\{i \in R_{T_2}(b|\tau_k)\} &= \Pr\{Y_{1i} \in t_{1i}, \tau_{k-1} < t_{1i} \leq \tau_k, D_{1i} = 1, Y_{2i} \geq b\} \\ &= \Pr\{T_1 \in t_{1i}, \tau_{k-1} < t_{1i} \leq \tau_k, T_2 \geq b\} G(t_{1i} + b).\end{aligned}$$

where $T_1 \in t_{1i}$ is the abbreviation of $T_1 \in (t_{1i}, t_{1i} + \Delta)$ as $\Delta \to 0$. □

Therefore, the conditional Kaplan-Meier produces biased results because the value of t_{1i} affects the probability of the corresponding T_{2i} being included in $R_{T_2}(b|\tau_k)$. To adjust this heterogeneity, one can weight each observation in $R_{T_2}(b|\tau_k)$ by an estimate of the reciprocal of $G(t_{1i} + b)$.

We define the conditional cumulative hazard estimator as follows:

$$\begin{aligned}\widehat{\Lambda}_{T_2|T_1^*=\tau_k}(\Delta b) &= \frac{\sum_{i \in R_{T_2}(b|\tau_k)} 1\{Y_{2i} = b, D_{2i} = 1\}/\hat{G}(t_{1i} + b)}{\sum_{i \in R_{T_2}(b|\tau_k)} 1\{Y_{2i} \geq b\}/\hat{G}(t_{1i} + b)} \\ &= \frac{\sum_{i=1}^{n} 1\{\tau_{k-1} < Y_{1i} \leq \tau_k, D_{1i} = 1, Y_{2i} = b, D_{2i} = 1\}/\hat{G}(Y_{1i} + b)}{\sum_{i=1}^{n} 1\{\tau_{k-1} < Y_{1i} \leq \tau_k, D_{1i} = 1, Y_{2i} \geq b\}/\hat{G}(Y_{1i} + b)} \qquad (8)\end{aligned}$$

for every b such that $\max_{1\leq i\leq n} Y_{1i} + b < \hat{\tau}_C$ where $\hat{\tau}_C = sup\{t : \hat{G}(t) > 0\}$ is the observed maximum follow-up time and where $\hat{G}(\cdot)$ is the empirical survival computed from the follow-up times.

This estimator has a potential problem when $\hat{G}(\cdot) = 0$. The convention $0/0 = 0$ is used to avoid the misdefinition. However, in many clinical trials, and in particular in the one that motivated our work, the follow-up time C is a continuous variable which is observed for all the individuals and hence $\hat{G}(\cdot) \neq 0$, except for the largest follow-up time.

A nonparametric estimator, $\widehat{S}_{T_2|T_1^*=\tau_k}(v)$, for the conditional survival function is obtained by plugging (8) into (7) as follows:

$$\widehat{S}_{T_2|T_1^*=\tau_k}(v) \quad = \quad \prod_{b\leq v}\{1 - \widehat{\Lambda}_{T_2|T_1^*=\tau_k}(db)\}. \tag{9}$$

Asymptotic properties of $\widehat{\Lambda}_{T_2|T_1^*=\tau_k}(\Delta b)$ and $\widehat{S}_{T_2|T_1^*=\tau_k}(v)$, as well as related issues to the estimation of the variance of $\widehat{S}_{T_2|T_1^*=\tau_k}(v)$ via a bootstrapping methodology can be found in Gómez *et al.* (2004 and 2007). A simulation study illustrating its good behaviour when the sample size is moderate is included in Gómez *et al.* (2007).

4.2 Implementation

We have implemented in S-PLUS the inverse probability of censoring weighted (IPCW) conditional methodology introduced in the previous section. The main function in the library is called `bwwce21` and its syntax is the following:

```
bwwce21(vartimes1,varcens1,vartimes2,varcens2,breaks,wmet,vtfw,vcfw)
```

where

```
vartimes1 = first time variable (T1 by default),
varcens1 = censoring indicator for the first time (D1 by default),
vartimes2 = second time variable (T2 by default),
varcens2 = censoring indicator for the second time (D2 by default),
breaks = partition values ({12, 24, 48, 96} by default),
wmet = weighting method for the dependent censoring
   (0=no weights, 1=follow-up -default-, 2=T1+T2, 3=T1+T2+T3),
vtfw = follow-up time variable (TFW by default),
vcfw = censoring indicator for the follow-up time (DFW by default).
```

Function `bwwce21` allows to reproduce the conditional Kaplan-Meier estimator attending the categories in the variable T_1, by setting no weights (`wmet=0`) in the call.

4.3 TIBET project: Conditional estimation of the first time ON given the first time OFF

We illustrate the conditional estimator, given in Subsection 4.1, for the estimation of the survival of the first time with treatment conditioned to the first time without treatment. Clinicians were very interested in the survival pattern of the first time with treatment, T_2, for patients who had been short, medium and long times without treatment, T_1. Based on these considerations they fixed the times of interest for the conditional analysis in $\tau_1 = 12$ (one trimester), $\tau_2 = 48$ (one year) and $\tau_3 = 96$ (two years), and according to this partition we have the following three categories in T_1: $T_1 \leq 12$, $12 < T_1 \leq 48$ and $48 < T_1 \leq 96$. Among the 100 patients, there are 31 with T_1 right-censored ($D_1 = 0$). Among the 69 patients with T_1 observed ($D_1 = 1$), there are 15 patients with T_2 right–censored ($D_2 = 0$) and 51 patients with T_1 and T_2 observed ($D_1 = D_2 = 1$).

In Figure 1 we observed that, except for a few number of subjects, there is a common minimum follow-up of 96 weeks. As a consequence, if we are estimating the survival at time $T_2 = v$ in the category T_1^*, the standard Kaplan-Meier estimator would be enough if $v + T_1^* \leq 96$ (for example, for $v \leq 84$ weeks when $T_1^* = 12$ or for $v \leq 48$ weeks when $T_1^* = 48$) and, on the other hand, we will appreciate the correction of the bias due to the dependent censoring when $v + T_1^* \geq 96$ by using the proposed weighted methodology.

Table 3: *Estimates and standard errors (computed using bootstrap) for some selected times of the conditional survival of T_2 given the following three categories: T_1: $T_1 \leq 12$, $12 < T_1 \leq 48$ and $48 < T_1 \leq 96$*

t_2	(0,12]	(12,48]	(48,96]
12	0.964 (0.0328)	0.954 (0.0463)	1 (0)
24	0.464 (0.1273)	0.727 (0.1048)	0.802 (0.1265)
36	0.143 (0.0695)	0.410 (0.1028)	0.571 (0.1938)
48	0.107 (0.0588)	0.228 (0.0889)	0.386 (0.1674)
60	0.071 (0.0539)	0.228 (0.0885)	0.298 (0.1566)
100	0.071 (0.0511)	0.228 (0.0881)	0 (0.1889)

Table 3 provides the estimates and standard errors for some selected times of the conditional survival of T_2 given categories in T_1. The standard errors have been computed using bootstrap. Based on these results (see Table 4) we estimate that while a patient who needs treatment quite fast ($0 < T_1 \leq 12$) will be as well fast in recovering his/her CD4 and viral load levels (median equals to 23), those patients which are able to stay a bit longer without treatment ($12 < T_1 \leq 48$) take longer time to recover levels (median equals to 30) and those patients which are able to stay much longer without treatment ($48 < T_1 \leq 96$) take much longer time to recover levels (median equals to 42). This behaviour can be explained by introducing as a covariate the cause of treatment reinitiation (plasma viral load > 100000 copies/ml and/or CD4+ counts ≤ 350 cells/mm^3). Those patients that reinitiate treatment because viral load has become

Table 4: *Description of T_2 given categories of T_1*

	$T_1 \leq 12$	$12 < T_1 \leq 48$	$48 < T_1 \leq 96$
Size	29	22	13
Events	26	17	8
1st Q	20	23	35
Median	23	30	42
3rd Q	30	45	80

higher than 100000 copies/ml, do so quite fast and their immunological system has not had time to be deteriorated. Since treatment is design to control viral replications these patients need shorter times to reach an undetectable viral load and still CD4 > 500. On the other hand, although some patients are able to stay without treatment long enough because they can keep viral load below 100000, the immunological system is slowly, but constantly, deteriorating. As a consequence, once they start treatment they need a longer period to recover the immunological level.

We present in Figure 2 the standard conditional Kaplan-Meier estimator together with the proposed weighted estimator of the conditional survival function $\widehat{S}_{T_2|T_1^*=\tau_k}(v)$ given in (9), for each of these categories. Note that in the first two categories both estimators coincide due to the long common minimum follow-up as we have previously noticed. Figure 2 also illustrates that the time that a patient needs to recover their immunological and virological levels depends on the time that he/she has been without treatment, as we have already observed in the previous paragraph. We can also clearly see a different behaviour between the survival of the times on treatment for patients that

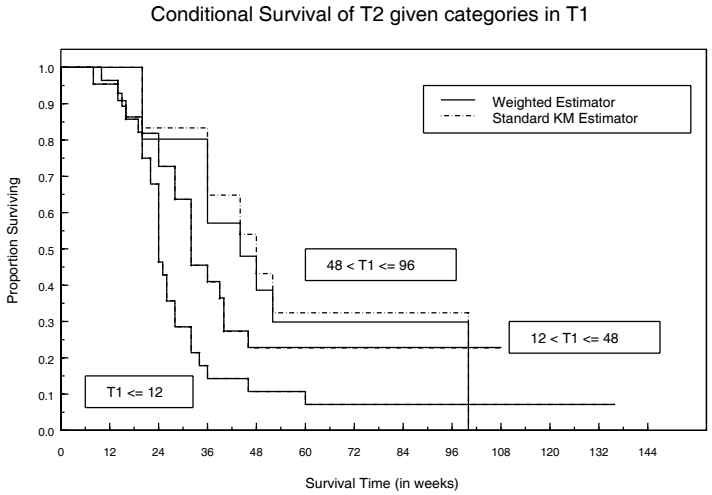

Figure 2: *Plots of the conditional survival function of T_2 given T_1 in the following three categories: $T_1 \leq 12$, $12 < T_1 \leq 48$ and $48 < T_1 \leq 96$ for both the standard conditional Kaplan-Meier and the proposed weighted estimator given in (9).*

stay without treatment less than 12 weeks as opposed to those that stay without treatment more than 48 weeks.

5 Wang and Wells' method. A continuous approach

5.1 Wang and Wells' estimator

Wang and Wells (1998) propose a path-dependent (nonparametric) estimate for the joint survival function of two duration variables.

According to the notation introduced in Section 2, let the observed sample be $\mathcal{S} = \{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}), i = 1, \ldots, n\}$, and $\mathcal{S}^* = \{(Y_{1i}, Y_{2i}, D_{1i} = 1, D_{2i}), i = 1, \ldots, n\}$ the subset of $\mathcal{S}$ consisting of those observations for which T_1 is observed. Wang and Wells consider the following path-dependent decomposition

$$\begin{aligned} S_{12}(t_1, t_2) &= \Pr\{T_2 > t_2 | T_1 > t_1\} \Pr\{T_1 > t_1\} \\ &= \prod_{v \leq t_2} \{1 - \Lambda_{T_2|T_1>t_1}(dv)\} S_1(t_1) \end{aligned} \tag{10}$$

where $\Lambda_{T_2|T_1>t_1}(dv)$ is the cumulative conditional hazard of T_2 given $T_1 > t_1$. Wang and Wells propose to estimate $S_{12}(t_1, t_2)$ via estimable components for $\Lambda_{T_2|T_1>t_1}(dv)$ and for $S_1(t_1)$ and to plug them into (10). The estimation of the marginal $S_1(t_1)$ is accomplished using the Kaplan-Meier estimator based on the observables (Y_{1i}, D_{1i}) $(i = 1, \ldots, n)$.

The estimator for $\Lambda_{T_2|T_1>t_1}(dv)$ extends Campbell and Földes estimator so that dependent censoring is taking into account. First note that if we let $R_{T_2}(v|t_1)$ be the risk set of T_2 at time v given $T_1 > t_1$, if $v > 0$ then $R_{T_2}(v|t_1) \subset \mathcal{S}^*$. An observation i with the first duration $T_1 = t_{1i}$ affects the probability of the corresponding T_{2i} being included in $R_{T_2}(v|t_1)$ as we see in the following expression

$$\begin{aligned} \Pr\{i \in R_{T_2}(v|t_1)\} &= \Pr\{Y_{1i} \in t_{1i}, t_{1i} > t_1, D_{1i} = 1, Y_{2i} \geq v\} \\ &= \Pr\{T_1 \in t_{1i}, t_{1i} > t_1, T_2 \geq v\} G(t_{1i} + v). \end{aligned}$$

Hence they adjust this heterogeneity by weighting each observation in $R_{T_2}(v|t_1)$ by an estimate of $1/G(t_{1i} + v)$.

Wang and Wells' estimator for $\Lambda_{T_2|T_1>t_1}(dv)$ can be expressed as follows:

$$\begin{aligned} \hat{\Lambda}^{WW}_{T_2|T_1>t_1}(\Delta v) &= \frac{\sum_{i \in R_{T_2}(v|t_1)} 1\{Y_{2i} = v, D_{2i} = 1\} / \hat{G}(t_{1i} + v)}{\sum_{i \in R_{T_2}(v|t_1)} 1\{Y_{2i} \geq v\} / \hat{G}(t_{1i} + v)} \\ &= \frac{\sum_{i=1}^{n} 1\{Y_{1i} > t_1, D_{1i} = 1, Y_{2i} = v, D_{2i} = 1\} / \hat{G}(Y_{1i} + v)}{\sum_{i=1}^{n} 1\{Y_{1i} > t_1, D_{1i} = 1, Y_{2i} \geq v\} / \hat{G}(Y_{1i} + v)} \end{aligned} \tag{11}$$

where $\hat{G}(\cdot)$ is an appropriate estimator of $G(\cdot)$ computed from the follow-up data. For example, $\hat{G}(\cdot)$ can be the Kaplan-Meier estimator of $G(\cdot)$ computed from the data $(Y_{1i} + Y_{2i}, 1 - D_{1i}D_{2i})$ $(i = 1, \ldots, n)$.

Wang and Wells' estimator for $S_{T_2|T_1>t_1}(t_2) = \Pr\{T_2 > t_2 | T_1 > t_1\}$ is given by plugging $\hat{\Lambda}^{WW}_{T_2|T_1>t_1}(\Delta v)$ into (10):

$$\widehat{S}^{WW}_{T_2|T_1>t_1}(t_2) \quad = \quad \prod_{v \le t_2} \{1 - \hat{\Lambda}^{WW}_{T_2|T_1>t_1}(dv)\}. \tag{12}$$

and the corresponding estimator for $S_{12}(t_1, t_2)$ is given by

$$\hat{S}_{12}(t_1, t_2) \quad = \quad \widehat{S}^{WW}_{T_2|T_1>t_1}(t_2)\hat{S}_1(t_1) \tag{13}$$

Their estimator uses the information on the first duration to weight each observation to unbias the effect of dependent censoring. This estimator has a potential problem with the existence of $\hat{S}_{12}(t_1, t_2)$ when $\hat{G}(\cdot) = 0$. If the largest value of $Y_{1i} + Y_{2i}$, say $c_{(n)}$, is censored $(D_{1i}D_{2i} = 0)$, then the largest observation of the censoring variables is observed $(1 - D_{1i}D_{2i} = 1)$ and hence $\hat{G}(c_{(n)}) = 0$. However, in this case the numerator in (11) is also 0 and the convention $0/0 = 0$ can be used. Note that the marginal survivor function can be estimated by $\hat{S}_2(t_2) = \hat{S}_{12}(0, t_2)$.

Wang and Wells show that $\hat{S}_{12}(t_1, t_2)$ converges in probability to $S_{12}(t_1, t_2)$ and claim that the limit distribution of $\sqrt{n}(\hat{S}_{12}(t_1, t_2) - S_{12}(t_1, t_2))$ converges weakly to a zero-mean Gaussian process, but the variance of the limiting process is quite complex and is not given.

5.2 Joint survival considerations for (T_1, T_2)

The bivariate estimator is useful in predicting the joint survival experience, in estimating the degree of dependence, in model building and testing and in strengthening marginal analysis. Furthermore, it is a necessary step if we want to compare $\widehat{S}_{T_2|T_1^*=\tau_k}(v)$, given in (9), to the estimator of $S_{T_2|T_1^*=\tau_k}(v)$ obtained from Wang and Wells' approach.

An estimator, $\hat{S}^{WW}_{12}(t_1, t_2)$, for the bivariate survival function of (T_1, T_2) is obtained plugging $\widehat{S}_1(t)$ and $\widehat{S}^{WW}_{T_2|T_1>t_1}(t_2)$, given in (12), into (13). This estimator suffers from two drawbacks: it is not a legitimate survival function and is dependent on the selected path and ordering of the components. We propose to isotonize $\hat{S}^{WW}_{12}$ so that the survival function is monotone in both components. Denote by $\hat{S}^{isot}_{12}$ the isotonic version of $\hat{S}^{WW}_{12}$.

5.3 Related issues

On one hand, by using the joint survival $\hat{S}^{isot}_{12}$ introduced in the previous subsection, we can also derive an estimator for the survival of T_2 conditioned on the categories in T_1

just defining $\hat{S}^{isot}_{T_2|T_1^*=\tau_k}(v)$ as follows:

$$\hat{S}^{isot}_{T_2|T_1^*=\tau_k}(v) = \frac{\hat{S}^{isot}_{12}(\tau_{k-1}, v) - \hat{S}^{isot}_{12}(\tau_k, v)}{\hat{S}_1(\tau_{k-1}) - \hat{S}_1(\tau_k)}, \tag{14}$$

This is an alternative estimator for $S_{T_2|T_1^*=\tau_k}(v)$.

On the other hand, when the investigator is interested in the bivariate survival distribution in some prefixed intervals of time (for instance, we might be interested in the survival behaviour for every year) the estimation of the following joint function $f_{T_1^*,T_2}(\tau_k, v) = \Pr(T_1^* = \tau_k, T_2 > v)$ is particularly appealing. $f_{T_1^*,T_2}(\tau_k, v)$ can be factorized following the path-dependent decomposition: $f_{T_1^*,T_2}(\tau_k, v) = \Pr\{T_2 > v | T_1^* = \tau_k\} \cdot \Pr\{T_1^* = \tau_k\} = S_{T_2|T_1^*=\tau_k}(v) \cdot \Pr\{T_1^* = \tau_k\}$.

As a consequence, $f_{T_1^*,T_2}(\tau_k, v)$ is estimated straightforwardly using the nonparametric estimation of $S_{T_2|T_1^*=\tau_k}(v)$ provided in (9) and from which we know asymptotic properties, and an estimator, $\widehat{\Pr}\{T_1^* = \tau_k\}$, for $\Pr\{T_1^* = \tau_k\}$. To estimate $\Pr\{T_1^* = \tau_k\}$ we simply estimate the marginal survival function of T_1, $S_1(\tau_k)$, using the Kaplan-Meier estimator, $\widehat{S}_1(\tau_k)$, and replace accordingly, that is, $\widehat{\Pr}\{T_1^* = \tau_k\} = \hat{S}_1(\tau_{k-1}) - \hat{S}_1(\tau_k)$.

5.4 Implementation

In a similar way that for the weighted conditional estimator in Section 4 we implemented in S-PLUS the function `bwww21` to estimate the joint survival distribution of (T_1, T_2) according to the Wang and Wells estimator in (13). The basic syntax of `bwww21` is: `bwww21(vartimes1, varcens1, vartimes2, varcens2, wmet, vtfw, vcfw)` where the parameters for the function are the same as the ones described for the `bwwce21` function in Subsection 4.2.

Specific computations for $\hat{\Lambda}^{WW}_{T_2|T_1>t_1}(dv)$ in (12) have been implemented in the function `lww21`. The function `bwww21` also uses the function `isoton` that performs the isotonization of a matrix so that the corresponding survival function is monotone in both components. It is important to note that, in order to avoid successive steps isotonizing by rows and columns alternatively, with non-unique results, our algorithm applies in one single-step an upper-left triangular minimization (see the corresponding code below). Related functions are `is.isoton` and `isotonv` that have been implemented to check if a matrix is isotonic and to isotonize a vector, respectively.

```
isoton <- function(mat) {
 n.r <- dim(mat)[1]
 n.c <- dim(mat)[2]
 mati <- mat
 for(j in 2:n.c) mati[1,j] <- min(mati[1, c(j-1,j)])
 for(i in 2:n.r) mati[i,1] <- min(mati[c(i-1, i),1])
```

```
 for(j in 2:n.c) for(i in 2:n.r)
   mati[i,j] <- min(mati[i-1,j], mati[i,j], mati[i,j-1])
 mati
}
```

Finally, the conditional survival $\hat{S}^{isot}_{T_2|T_1^*}$ in (14) has been implemented in the function `j2c`.

5.5 TIBET project: Joint survival estimation

In the same way that in Section 4 the follow-up time variable, TFW, is the information on the censoring that we have for each patient.

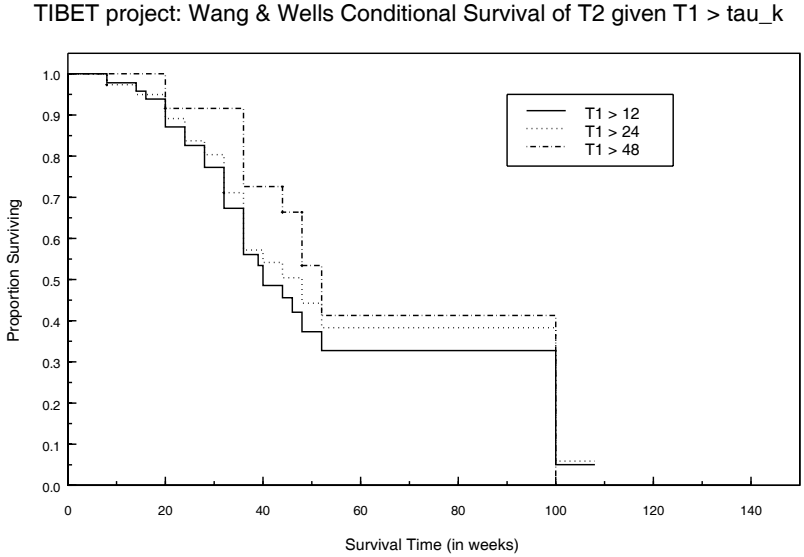

Figure 3: *Plots of the conditional survival function of T_2 given categories in T_1. T_1 represents the number of weeks without treatment and is splitted into three categories: $T_1 > 12$, $T_1 > 24$ and $T_1 > 48$. Each curve represents the survival function of the first time with treatment on each of the categories.*

Figure 3 illustrates Wang and Wells estimator $\widehat{S}^{WW}_{T_2|T_1>t_1}(t_2)$ given in (12) for the following three categories defined by T_1: $T_1 > 12$, $T_1 > 24$ and $T_1 > 48$. We see from these curves that patients who stay without treatment more than 48 weeks, will stay with treatment longer times than those patients who stayed OFF more than 24 weeks.

In Table 5 we illustrate the isotonic joint survival estimator for $\hat{S}_{12}(t_1, t_2)$, $\hat{S}^{isot}_{12}$, proposed in Subsection 5.2. We can see that the joint estimation it is not feasible for those pairs (t_1, t_2) with no events for T_2, with $T_2 > t_2$, in the category $T_1 > t_1$.

Table 5: *Estimates for some selected times of the joint survival of* (T_1, T_2) *using Wang and Wells method.*

	$t_1 = 0$	6	12	24	48	72	96	120
$t_2 = 12$	0.972	0.926	0.694	0.613	0.489	0.405	0.353	0.312
24	0.691	0.666	0.586	0.527	0.448	0.360	0.353	0.312
36	0.394	0.394	0.394	0.360	0.355	0.207	0.207	–
48	0.266	0.266	0.265	0.256	0.256	–	–	–
60	0.219	0.219	0.219	0.219	0.202	–	–	–
100	0.050	0.050	0.036	0.033	0.000	–	–	–

With respect to isotonizing the resulting joint survival in (13), note that we did not need to isotonize in more than 60% of the points. On the other hand, the resulting differences in the rest of the points have −0.157, −0.025 and −0.010 as quartiles.

Table 6 gives the estimates for the conditional survival of T_2 given categories in T_1, $\hat{S}^{isot}_{T_2|T_1^*=\tau_k}(v)$, derived from Wang and Wells method. In more than 70% of the points isotonization has not been necessary.

Table 6: *Estimates for some selected times of the conditional survival of* T_2 *given categories in* T_1 *after estimating the joint survival distribution of* (T_1, T_2) *using Wang and Wells method.*

t_2	(0,12]	(12,48]	(48,96]
12	0.960	0.930	1
24	0.360	0.627	0.697
36	0.000	0.130	0.697
48	0.000	0.005	–
60	0.000	0.005	–
100	0.000	0.005	–

After comparing with the weighted conditional estimator that we have proposed in the Section 4 (see Table 3), we can see that the conditional estimator derived from Wang and Wells approach underestimates, in general, the corresponding survival.

6 Discussion

In this paper we have illustrated three different approaches to analyze two successive survival times. The main difficulty in this type of study is the presence of the dependent censoring induced by the potential relationship between both times of interest. All the approaches consider the estimation either of the joint distribution of (T_1, T_2) or the conditional distribution of T_2 given T_1. The main difference between the proposed methods is on the conditioning strategy and the way of considering the correction of the bias due to the dependent censoring.

Visser's method is based on the direct estimation of the conditional survival given an specific value for T_1, and it does not correct for the effect of the dependent censoring. As we mentioned in Section 3 the main restriction of this methodology is that it needs an important initial sample in order to obtain, after conditioning, a sample size that is large enough to estimate efficiently the conditional survival.

On the other hand, the weighted conditional estimation proposed by Gómez and Serrat provides an unbiased estimator for the conditional survival function for the second survival time given the categories in the first survival time. The main interest of this approach is that it takes into account the heterogeneity due to the dependent censoring by using all the information provided by T_1 to weight the observed data. In this sense, Gómez and Serrat's estimator is a good alternative to Visser's method because it does not need a discretization of the time variables and it allows to perform the estimation when the sample size is not very large. We remark here that although our parameter of interest is based on the categories of a first survival time, we use the continuous survival times, t_{1i}, without discretizing them, to contribute to the inverse weight, $G(t_{1i}+b)^{-1}$, and furthermore we do not need to discretize the second survival time T_2.

Concerning Wang and Wells' estimator, it is important to remark that the conditioning part is based on the subsets $T_1 > t_1$ and that the methodology corrects for the dependent censoring. However, the resulting estimator for the joint survival is not isotonic and, as a consequence, it is not a proper distribution. Hence the derivation of other functions of the bivariate survival distribution, for instance the conditional survival in the Tibet clinical trial, is questionable. As we noticed in Subsection 5.2 an alternative could be to isotonize the resulting estimates, however, as we can see after comparing Tables 3 and 6, this strategy provides a quick-to-zero survival distribution that underestimates the parameter of interest. In this sense, the weighted conditional estimation is also an interesting alternative to Wang and Wells' estimator because it avoids the non-desirable effects of the isotonization.

It is important to note that the proposed Gómez and Serrat's estimator can depend on the partition and the resulting estimates can be sensitive to the sample size in each category as well as to the number of different observed times T_2 in the category. In practice and for the Tibet clinical trial study, we have also analyzed the dataset using other partitions, for instance splitting T_1 into the following four categories: $T_1 < 12$, $12 < T_1 \leq 24$, $24 < T_1 \leq 48$ and $48 < T_1 \leq 96$, and similar results are obtained. In fact, in order to choose a partition for the analysis, it is necessary to take into account not only the resulting sample size in each category but also the number of different observed times for T_2 in the category.

Extensions of Gómez and Serrat's approach to the estimation of the survival function for other successive duration times given the information on the first are under consideration for the authors, by studying the effect of the intermediate events in the estimation of the appropriated weights for each subject. In the Tibet study it could be of interest, for instance, to estimate the duration of the second period OFF, T_3, given the category of the duration of the first period OFF, T_1.

All the approaches in this paper have been implemented in S-PLUS and they are easily exportable to other available software or platforms. The respective functions are available at the web page of the GRASS group at `http://www-eio06.upc.es/grass`.

Acknowledgements

This work has been partially supported by grant MTM2005–08886 from the Ministerio de Ciencia y Tecnología. Authors are grateful to the GRASS group for fruitful discussions and interesting suggestions, and to the Fundació Lluita contra la SIDA for providing the data set. We thank the reviewers and the editor for constructive suggestions in the previous version of this manuscript.

Appendix: Visser's likelihood

To simplify the expressions we introduce the following complementary notation. For a given individual a different way of representing the observables is using (Y_1, Y_2, δ), where

$$\delta = \begin{cases} 1 & \text{if} \quad T_1 > C \qquad (i.e. \quad D_1 = 0 = D_2) \\ 2 & \text{if} \quad T_1 \leq C < T_1 + T_2 \quad (i.e. \quad D_1 = 1, D_2 = 0) \\ 3 & \text{if} \quad T_1 + T_2 \leq C \qquad (i.e. \quad D_1 = 1, D_2 = 1). \end{cases}$$

The likelihood for the n observations is as follows:

$$\begin{aligned} L \;=\; & \prod_{i=1}^{n} \Big\{ \Pr\{T_1 > y_{1i}, C = y_{1i}\}^{1\{\delta_i=1\}} \\ & \Pr\{T_1 = y_{1i}, T_2 > y_{2i}, C - y_{1i} = y_{2i}\}^{1\{\delta_i=2\}} \\ & \Pr\{T_1 = y_{1i}, T_2 = y_{2i}, C > y_{1i} + y_{2i}\}^{1\{\delta_i=3\}} \Big\} \end{aligned}$$

and the corresponding log-likelihood, $\mathcal{L} = \log L$, since variables are discrete, the possible values for y_{1i}, y_{2i} and C are only $\{0, 1, 2, \ldots, K\}$ and C is independent of (T_1, T_2), looks like as

$$\begin{aligned} \mathcal{L} \;=\; & \sum_{k=1}^{K} \{n_{k\cdot} \log \Pr\{T_1 = k\} + n_{1k} \log \Pr\{T_1 > k\}\} + \\ & \sum_{k=1}^{K} \sum_{l=1}^{K} \{n_{3kl} \log \Pr\{T_2 = l | T_1 = k\} + n_{2kl} \log \Pr\{T_2 > l | T_1 = k\}\} + \end{aligned}$$

$$\sum_{k=1}^{K}\{n_{1k}\log\Pr\{C=k\}+\sum_{l=1}^{K}\{n_{2kl}\log\Pr\{C=k+l\}+n_{3kl}\log\Pr\{C>k+l\}\}\}$$
$$= \mathcal{L}_{T_1}+\mathcal{L}_{T_2|T_1}+\mathcal{L}_C.$$

All the expressions for the probabilities can be replaced by functions containing uniquely $\lambda_{T_1}(k)$ and to $\lambda_{T_2|T_1=k}(l)$. For instance,

$$\begin{aligned}\mathcal{L}_{T_1} &= \sum_{k=1}^{K}\{n_{k\cdot}(\log\lambda_{T_1}(k)+\log\prod_{j=0}^{k-1}(1-\lambda_{T_1}(j)))+n_{1k}\log\prod_{j=0}^{k}(1-\lambda_{T_1}(j))\}\\ &= \sum_{k=1}^{K}n_{k\cdot}\log\lambda_{T_1}(k)+\sum_{k=1}^{K}(n_{1k}+n_{k\cdot})\sum_{j=0}^{K-1}\log(1-\lambda_{T_1}(j))+\sum_{k=1}^{K}n_{1k}\log(1-\lambda_{T_1}(k)).\end{aligned}$$

The nonparametric estimators for the hazard functions are obtained after maximizing the log-likelihood $\mathcal{L}=\mathcal{L}_{T_1}+\mathcal{L}_{T_2|T_1}+\mathcal{L}_C$. Note that we are in fact maximizing $\log L$ with respect to $\lambda_{T_1}(k)$ and to $\lambda_{T_2|T_1=k}(l)$, and because the terms act additively we can maximize first with respect to $\lambda_{T_1}(k)$ and then with respect to $\lambda_{T_2|T_1=k}(l)$.

References

Burke, M. D. (1988). Estimation of a bivariate distribution function under random censorship. *Biometrika*, 75, 379-382.

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika*, 68, 417-422.

Campbell, G. and Földes, A. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. *Colloquia Mathematica Societatis Janos Bolyai*, 32, 103-121.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics*, 16, 1475-1489.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Boca Raton: Chapman & Hall/CRC.

Gómez, G., Calle, M. L., Serrat, C. and Espinal, A. (2004). Review of Multivariate Survival Data. Technical Report DR2004/15. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.

Gómez, G., Serrat, C. and Ruiz, L. (2007). Weighted conditional survival estimator for ordered failure times subject to a common censoring process. *Submitted.*

Ruiz, L., Paredes, R., Gómez, G. *et al*. (2007). Antiretroviral Therapy Interruption Guided by CD4 T cell Counts and Plasma HIV-1 RNA Levels in chronically HIV-1 Infected Patients. *AIDS*, 21, 169-178.

Tsai,W-Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics*, 14, 1351-1365.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika*, 83, 507-518.

Wang, M.-C. and Wells, M. T. (1998). Nonparametric Estimation of successive duration times under dependent censoring. *Biometrika*, 85, 561-572.