Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Semi-Competing Risks Data - Application to LTC Insurance*

Quentin Guibert[†1,2] and Frédéric Planchet[‡2,3]

[1] *Autorité de Contrôle Prudentiel et de Résolution*[§]
[2] *ISFA, Laboratoire SAF, Université Claude Bernard Lyon 1*[¶]
[3] *Prim'Act*[‖]

March 23, 2015

Abstract. The studying of Long Term Care insurance problems requires to model the lifetime of individual in presence of both terminal and non-terminal events which are concurrent. This corresponds to semi-competing risks data where there are several intermediate endpoints subject to dependent censoring. In this paper, we regard this situation deploying multi-state approach and we exhibit non-parametric estimators of transition probabilities considering the Markov assumption is not rely. The proposed estimators can be seen as Aalen-Johansen integrals for competing risks data which are obtained by re-setting the system with two competing risks blocks. As little attention has been given to this, we derive asymptotic results for this type of estimator under non-dependent random right censorship when covariates are present and discuss its possible outlooks. We also develop a methodology to regard times dependence association measure between failure times for cause-specific. For key transition probabilities, we conduct simulations to analyze the performance of our estimators versus the classical Aalen-Johansen estimators. Finally, numerical application with LTC insurance data, traditionally analyzed with semi-Markov model, is proposed.

Keywords. Semi-Competing Risks Data, Aalen-Johansen Integral, Non-parametric Estimator, Non-Markov Process, Multi-state Model, Simulation, Time-Dependence Association Measure, LTC Insurance.

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Autorité de Contrôle Prudentiel et de Résolution (ACPR), neither those of the Banque de France.

[†]Email: quentin.guibert@acpr.banque-france.fr
[‡]Email: frederic.planchet@univ-lyon1.fr
[§]ACPR-CC 66 2771, 61 Rue Taitbout, 75436 Paris Cedex 09, France.
[¶]Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances, 50 Avenue Tony Garnier, F-69007 Lyon, France.
[‖]Prim'Act, 42 Avenue de la Grande Armée, 75017 Paris, France.

1 Introduction

1.1 About LTC insurance context

Long Term Care (LTC) insurance is a mix of social and health care provided on a daily basis, formally or informally, at home or in institutions, to people suffering from a loss of mobility and autonomy in their activity of daily living. In France, this guarantee is dedicated to elderly people who are partially or totally dependent and benefits are mainly paid as an annuity depending on the lifetime of policyholders and possibly to the degree of dependency, see e.g. Courbage and Roudaut (2011) or Plisson (2009) for general discussions on the french LTC insurance market.

Multi-state models, see Hougaard (1999), Hougaard (2001, Chapter 5 and 6) and Andersen and Keiding (2002), have known growing interest for fifteen years in the actuarial literature, since they offer sound modeling for the random pattern of states experiences by a policyholder. Lots of further applications exist in biostatistic (see Commenges, 2002; Putter *et al.*, 2007), credit risk and reliability areas (see Lando and Skødeberg, 2002; Janssen and Manca, 2007). For health and life insurance modeling purposes, many papers have been developed comprehensive frameworks for pricing and reserving both with Markov or semi-Markov assumptions. Among them, we recommend the classical books of Haberman and Pitacco (1998) and Denuit and Robert (2007) and the cited literature. Recently, Christiansen (2012), based on the PhD thesis of Helwich (2008), give a wide overview of how using multi-state models in health insurance, including LTC insurance, for academics interests.

Fitting multi-state models related to LTC insurance with available longitudinal data involves various issues. The easier way followed by practitioners to do this is assuming the data fulfill Markov assumption, i.e. the transition to the next state depends only on the current state (see e.g. Gauzère *et al.*, 1999; Pritchard, 2006; Deléglise *et al.*, 2009; Levantesi and Menzietti, 2012). That implies the process ignores the effect of any previous life-history. In this context, non-parametric inference can be easily performed with the so-called Aalen-Johansen estimator (Aalen and Johansen, 1978) to compute transitions probabilities. However, this assumption is unappropriated when modeling LTC claimants mortality as the transition probabilities depend to the age and the duration (or sojourn time) in each disease, see evidence in Czado and Rudolph (2002), and semi-Markov model is required. In our framework, the residual lifetime once the individual is entered into dependency becomes very different from each other according to the type of disease.

In actuarial science, research about fitting non-Markov models is relatively scarce. The most explored area is about disability data which are generally fitted with parametric models, see e.g. the so-called Poisson model (Haberman and Pitacco, 1998) based on the number of deaths over the age and age of entry dimensions, or semi-parametric, e.g. Cox semi-Markov model in Czado and Rudolph (2002). These approaches rely on possibly strong assumptions on the parametric specification and there is a need to develop non-parametric approaches to check how these assumptions are realistic especially for the construction of biometric tables

in insurance. This last need of using realistic tables is particularly driven by the development of Solvency II and IFRS frameworks. One can find a similar motivation for LTC insurance in the recent papers of Guibert and Planchet (2014) and Tomas and Planchet (2013).

1.2 Statistical motivations

Multi-state models offer sound modeling for longitudinal data with both multiple event times. The analysis is complicated for actuaries with right-censoring by multiple causes of entry in non-terminal event and other direct failure types (e.g. cancellation of the contract) since these causes may be strongly dependently censored by death or even multiple causes of death. If all failure causes can occur in any order, this correspond to the case of competing risks data. However, we consider in this article that the lifetime ends due to different failure events but the individual can experience before a set of possible non-terminal events, such as different diseases, which cause ultimately death. When there is only one terminal event which dependently censors a terminal event, such situation refers to the framework of semi-competing risks data (see Fine *et al.*, 2001). In such a case, many works mainly focus on estimating the survival function of the latent failure time to the non-terminal event using for instance (see Lakhal *et al.*, 2008) the so-called copula-graphic estimator (Zheng and Klein, 1995). This is not sufficient to have a clear sketch of the risk and there is a need to construct non-parametric estimators for transition probabilities with useful theoretical results to provide for instance goodness-of-fits tests.

Although non-parametric estimation is covered for such multi-state data when Markov assumption is rely (see Andersen *et al.*, 1993), the complexity of inferring multi-state model without this assumption greatly depends on the number of states defined, the nature of data observed (censoring and truncation processes) and by transition between the states. With non-parametric techniques in presence of censoring, there is no general estimator for the transition probabilities in continuous times to the best of our knowledge and the literature about this subject is a field under development. Regarding to bio-statistic literature dedicated to non-Markov model, research has mainly focused on the non-parametric estimation techniques for particular probabilities. For instance, Datta and Satten (2002) defined estimators for the state occupation probabilities and Meira-Machado *et al.* (2006) have considered the transition probabilities estimators for the illness-death (or disability) model. Note also that Datta and Satten (2001) (consistency results) and Glidden (2002) (weak convergence results) show that Aalen-Johansen estimor can be used to infer state occupation probabilities without Markov assumption. An overview of inference methods used in medical studies when the Markov assumptions is not rely is available in Meira-Machado *et al.* (2009). Using non-parametric competing risks techniques Peng and Fine (2006) propose non-parametric estimator for sub-distribution functions with censoring and truncation processes.

In this paper, we consider multi-state approach to model situations with both multiple terminal and non-terminal events as it offer better modeling than using classical techniques developed for semi-competing risks data by grouping states. Similar remark is given by Xu *et al.* (2010). Conversely to general medical studies, our interest differs from a basic illness-

death model in that we wish distinguishing the mixture of illness which can cause entry in non-terminal stage and studying the lifetime after this entry. This situation is adapted to the studying of LTC insurance guarantees but we believe it may be also applicable for multimorbidity purpose (Varadhan *et al.*, 2014).

Our approach is derived and extended the work proposed by Meira-Machado *et al.* (2006) applying model based on two competing risks blocks which are nested to consider the progressive form of the model. Regardless to our research, Allignol *et al.* (2013) have also recently emphasized the interest of competing risks approaches to explain the Meira-Machado et al. estimators for illness-death model. With such structure our model can be viewed as a particular case of bivariate competing risks data problem for which Cheng *et al.* (2007) have exhibited non-parametric estimators for bivariate cause specific hazard and bivariate cumulative incidence functions under independent censoring. To the best of our knowledge, there is a gap in the literature regarding to the expectation of a random function under the estimator of bivariate cumulative incidence function contrary to the case of bivariate survival data (see e.g. Lopez, 2012). In our context, considering some particular random functions allows providing useful estimators for transition probabilities. Hence, we develop non-parametric methodology based on Aalen-Johansen integrals for competing risks data (see Suzukawa, 2002) and investigate the consistency and weak convergence of such estimators in presence of independent right-censoring and covariates. This approach encompasses the case of classic illness-death model and progressive model. It can be used to estimate transition probabilities in insurance for critical illness data where death due to dread disease should be distinguished to other causes.

Otherwise, it is quite natural in such context to study association measure among the cause-specific failure times. By considering recent advances for bivariate competing risks data (see Bandeen-Roche and Liang, 2002; Bandeen-Roche and Ning, 2008; Cheng *et al.*, 2007, 2010; Scheike and Sun, 2012), we exhibit non-parametric estimator in our framework for the cross-odds ratio which is valued association measures which can be used for instance as a preliminary estimator before applying regression models. This non-parametric approach is motivated as local version of Kendall'tau (Oakes, 1989) is complex to apply for competing risks data.

1.3 Outline

This paper is organized as follow: Section 2 introduces our non-Markov multi-state model and defines quantities of interest for assessing transition probabilities and association. An acyclic multi-state model with a set of intermediary and terminal states is re-parametrized to a system of two competing risks blocks which are nested. This is the main idea of this paper that we exploit in Section 3 to provide non-parametric estimators of interest quantities. We focus in this section on showing convergence properties of these estimators under independent right-censoring. Section 4 is devoting to a simple simulation analysis to assess the performance of our non-parametric transition probabilities estimators against Aalen-Johansen estimator. Application to real french LTC insurance data is proposed in

Section 5.

2 Competing risks model setup

This section outlines in 2.1 our notation and the general multi-state structure re-set with two competing risks blocks to encompass semi-competing risks data. This formulation allows modeling the individual lifetime who can experience both terminal and non-terminal events. In subsection 2.2, we then describe the transition probabilities usefully in actuarial science for pricing and reserving in 2.2. In 2.3, we present a version of cross-odd ratio to account for dependence between the lifetime in healthy state and the overall lifetime for each non-terminal and terminal causes.

2.1 Model and basic assumptions

On a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we consider a time-continuous stochastic process $(X_t)_{t\geq 0}$ with finite state space $\mathcal{S} = \{a_0, e_1, \ldots, e_{m_1}, d_1, \ldots, d_{m_2}\}$ and right-continuous paths with left-hand limits. This process represents the state of the individual, e.g. in insurance the policyholder, at time $t \geq 0$. The set $\{e_1, \ldots, e_{m_1}\}$ represents m_1 intermediary states or non-terminal events, e.g. disability competing caused of entry in dependency, and the set $\{d_1, \ldots, d_{m_2}\}$ are terminal events, i.e. absorbing states such as direct death, lapse or death after entry in dependency. The state a_0 corresponds to the healthy state for which we define in the same set-notations used by (Rotolo *et al.*, 2013), the set of its children $\mathcal{C}(a_0) \subset \mathcal{S}$ as the set of the states to which a direct transition from a_0 is possible. For each state $e \in \{e_1, \ldots, e_{m_1}\}$, we also introduce the set of its children $\mathcal{C}(e) \subset \{d_1, \ldots, d_{m_2}\}$. Of course, the set of children for one terminal event $d \in \{d_1, \ldots, d_{m_2}\}$ is $\mathcal{C}(d) = \emptyset$. Hence, an individual is likely to take two types of lifetime paths depending if a intermediate event occurring or not. The Figure 1 depicts an example of a such acyclical multi-state structure.

[Figure 1 about here.]

This multi-state structure can be expressed with latent failure times and we denote by $T_{0e}, e \in \{e_1, \ldots, e_{m_1}\}$, and $T_{0d}, d \in \{d_1, \ldots, d_{m_2}\}$, the potential latent times in healthy state in terms of arrival state where we distinguishes transitions to a non-terminal event and transitions to terminal event. Similarly we set T_{ed} the potential latent time since the individual became disabled respectively for intermediate endpoint e in terms of terminal event d. With this notation, we distinguish two steps for the overall lifetime corresponding to two competing risks schemes. At the first step, the lifetime is composed of $Card(\mathcal{C}(a_0))$-competing causes of exit from the healthy state i.e. the non-terminal events $(a_0 \to e)$ and the terminal events $(a_0 \to d)$. We write $\mathcal{C}(a_0) = \mathcal{C}_E(a_0) \cup \mathcal{C}_D(a_0)$ the split of arrival states

between non-terminal and terminal events. Hence, we set-up the competing risks process (S, V_1) such as

$$\begin{aligned} S &= \inf\{t : X_t \neq a_0\} \\ &= \sum_{e \in \mathcal{C}_E(a_0)} T_{0e} \mathbb{1}_{\{V_1 = e\}} + \sum_{d \in \mathcal{C}_D(a_0)} T_{0d} \mathbb{1}_{\{V_1 = d\}}, \end{aligned}$$

where V_1 taking its values in $\mathcal{C}(a_0)$. We denote with H the distribution function of the sojourn time S in healthy state.

At the second step, we consider the lifetime is exposed to m_2-causes of exit, i.e. direct exit $(a \to d)$ and exit with experience of one of the m_1 intermediary states $(a \to e \to d)$. This lifetime T is 'forced' into a competing risks setup (T, V) where $V = (V_1, V_2)$ is an indicator of the path followed by the individual and V_2 depends to the value taken by V_1 such as

$$\begin{cases} V = (e, d) \text{ with } d \in \mathcal{C}(e) \text{ if } e \in \mathcal{C}_E(a_0), \\ V = (0, d) \text{ otherwise.} \end{cases}$$

Hereafter, we denote $\mathcal{V}$ the set of values taken by V. With this definition, the variable V identifies cause-specific to exit by distinguishing those for which non-terminal event occurs and those without non-terminal event. With this notation, we have

$$\begin{aligned} T &= \inf\{t : X_t \in \{d_1, \ldots, d_{m_2}\}\} \\ &= \sum_{e \in \mathcal{C}_E(a_0)} \left[T_{0e} \mathbb{1}_{\{V_1 = e\}} + \sum_{d \in \mathcal{C}(e)} T_{ed} \mathbb{1}_{\{V_2 = d\}} \right] + \sum_{d \in \mathcal{C}_D(a_0)} T_{0d} \mathbb{1}_{\{V_2 = d\}}, \end{aligned}$$

where the survival time of the individual T has a distribution function F. These distribution functions are assumed to be continuous. Further, we set $F^{(v)}(t) = \mathbb{P}(T \leq t, V = v)$ the continuous sub-distribution function[1] for $t \geq 0$ and $v \in \mathcal{V}$.

As noted by Meira-Machado *et al.* (2006) with only one non-terminal event and one terminal event, the variables T_{ed}, representing the residual lifetime after the occurrence of one non-terminal event, are not observed in case of direct transition, so $S = T$. Otherwise, we have $S < T$. Similarly, the variables (T_{0e}), $e \in \{e_1, \ldots, e_{m_1}\} \setminus \{e'\}$, and (T_{0d}), $d \in \{d_1, \ldots, d_{m_2}\}$, are automatically censored if the individual experiences the intermediary state e'. This corresponds to a general definition of semi-competing risks data although applications with more than one terminal are scarce in the literature. Our formulation allows considering dependence between non-terminal events occurrences.

As is it classical for the lifetime data analysis, we consider now a right censoring variable C with a distribution function G in respect to the following assumption

Assumption 1. C is independent of the vector (S, T, V).

[1] Also called cumulative incidence distribution.

Independence Assumption 1 is widely used for simplicity in practice and seems to be consistent for LTC insurance observations. It is important to note that the censoring variable is unique for all the latent times. Because of the right censoring mechanism, the two lifetime variables (S, T) are not directly observed. Instead, the following variables are available

$$\begin{cases} Y = \min(S, C) \text{ and } \gamma = \mathbb{1}_{\{S \leq C\}}, \\ Z = \min(T, C) \text{ and } \delta = \mathbb{1}_{\{T \leq C\}}. \end{cases}$$

Additionally for the sake of generality, we incorporate in this paper a vector $\boldsymbol{U} = (U_i)_{i=1,\dots,p}$ of p-covariates. In practice, they may be discrete (as sex, geographical location, social situation ...) or continuous components (as smoker/non-smoker state, the generation ...) but we consider for this presentation only continuous covariates for the sake of clarity. No assumption about the dependence structure between $(C, S, T, \boldsymbol{U}, V)$ is defined here. However, we assume following Stute (1993) that these covariables do not provide any further information as to whether censoring will take place or not. So, we have the following assumptions throughout this paper

Assumption 2. We assume that:

i. $\mathbb{P}(S \leq C \mid S, \boldsymbol{U}, V_1) = \mathbb{P}(S \leq C \mid S, V_1)$,

ii. $\mathbb{P}(T \leq C \mid S, T, \boldsymbol{U}, V) = \mathbb{P}(T \leq C \mid T, V)$.

As pointed out above, the equality **ii.** of Assumption 2 is explained by the fact the pair (S, T) is subject by construction to censor and S is uncensored whenever T is.

For estimation purpose, we introduce the distribution function of $(S, \boldsymbol{U})$ noted H_0 and $F_0^{(v)}$ the sub-distribution function of $(S, T, \boldsymbol{U})$ where the cause is $V = v$, $v \in \mathcal{V}$. Moreover, we have $F_0 = \sum_v F_0^{(v)}$.

2.2 Transition probabilities

With the notation of subsection 2.1, the multi-state process $(X_t)_{t \geq 0}$ is specified in terms of transition probabilities with $0 \leq s \leq t$ and $i, j \in \mathcal{S}$

$$p_{ij}(s, t \mid \boldsymbol{u}) = \mathbb{P}(X_t = j \mid X_s = i, \boldsymbol{U} = \boldsymbol{u}).$$

For the risk assessment purpose in actuarial science, quantities of interest relate to the probabilities of paying or receiving cash-flows. In particular, it is needed to distinguish each intermediary state since the related payment function may be different. So, we want to estimate:

i. $p_{00}(s,t \mid \boldsymbol{u})$, the survival probability in healthy state with $0 \leq s \leq t$,

ii. $p_{0e}(s,t,\eta \mid \boldsymbol{u})$, the entry probability in a no-terminal state due to cause $e \in \mathcal{C}_E(a_0)$ with $0 \leq s \leq t$ and the sojourn time in state is e upper than $\eta \geq 0$ at time t,

iii. $p_{ee}(s,t,\mid \boldsymbol{u})$, the non-exit probability of the non-terminal event $e \in \mathcal{C}_E(a_0)$ between times s and t with $0 \leq s \leq t$,

iv. $p_{ed}(s,t,\eta,\zeta \mid \boldsymbol{u})$, the exit probability of the non-terminal event $e \in \mathcal{C}_E(a_0)$ due to cause $d \in \mathcal{C}(e)$ when the sojourn time in state e is lower than $\zeta \geq 0$ knowing the sojourn time is upper than $0 \leq \eta \leq \zeta$ and the time of entry in non-terminal event is between s and t with $0 \leq s \leq t$,

v. $p_{0d}(s,t \mid \boldsymbol{u})$, the direct transition probability from the healthy state to terminal event $d \in \mathcal{C}_D(a_0)$ without suffering of any disease with $0 \leq s \leq t$.

This list can be easily extend or modify considering some particular paths or if it is needed to exhibit exit probabilities from a non-terminal state to some specific sets of terminal events. Anyway, these quantities can be expressed in terms of the joint distributions of $(S, \boldsymbol{U})$ and $(S, T, \boldsymbol{U}, V)$ as follows

$$p_{00}(s,t \mid \boldsymbol{u}) = \frac{\mathbb{P}(S > t, \boldsymbol{U} = \boldsymbol{u})}{\mathbb{P}(S > s, \boldsymbol{U} = \boldsymbol{u})}, \tag{2.1}$$

$$p_{0e}(s,t,\eta \mid \boldsymbol{u}) = \frac{\mathbb{P}(s < S \leq \min(t, t-\eta), T > t, \boldsymbol{U} = \boldsymbol{u}, V_1 = e)}{\mathbb{P}(S > s, \boldsymbol{U} = \boldsymbol{u})}, \tag{2.2}$$

$$p_{ee}(s,t \mid \boldsymbol{u}) = \frac{\mathbb{P}(S \leq s, T > t, \boldsymbol{U} = \boldsymbol{u}, V_1 = e)}{\mathbb{P}(S \leq s, T > s, \boldsymbol{U} = \boldsymbol{u}, V_1 = e)}, \tag{2.3}$$

$$p_{ed}(s,t,\eta,\zeta \mid \boldsymbol{u}) = \frac{\mathbb{P}(\eta < T - S \leq \zeta, s < S \leq t, \boldsymbol{U} = \boldsymbol{u}, V = (e,d))}{\mathbb{P}(T - S > \eta, s < S \leq t, \boldsymbol{U} = \boldsymbol{u}, V_1 = e)}, \tag{2.4}$$

$$p_{0d}(s,t \mid \boldsymbol{u}) = \frac{\mathbb{P}(s < S, T \leq t, \boldsymbol{U} = \boldsymbol{u}, V = (0,d))}{\mathbb{P}(S > s, \boldsymbol{U} = \boldsymbol{u})}. \tag{2.5}$$

In Equations (2.2), (2.3) and (2.4), note that the event $\{V_1 = e\} = \{V_1 = e, V_2 \in \mathcal{C}(e)\}$. Regarding to numerator of Equation (2.5), we remark that considering $\mathbb{P}(T \leq t, \boldsymbol{U} = \boldsymbol{u}, V = (0,d))$ and $\mathbb{P}(S \leq t, \boldsymbol{U} = \boldsymbol{u}, V_1 = d)$ conducts to different results as the former assumes competition between all the terminal states whereas the latter is equivalent applying right-truncation on observations which experience non-terminal events.

2.3 Association measures

Consider the general model introduced above, the sub-distribution function $F_0^{(v)}$, $v \in \mathcal{V}$, is a key quantity that we have defined. This formulation is a particular case of general bivariate competing risks model with covariates and unique right-censoring process.

There has been little work on analyzing dependence for multivariate competing risks data contrary to survival data. In particular, it seems that the so-called Kendall's tau is not easily extendable for measuring association between several causes whereas this can be computed non-parametrically for bivariate survival data (see e.g. Lopez, 2012). However, this issue has recently received some attention (see Bandeen-Roche and Liang, 2002; Bandeen-Roche and Ning, 2008; Cheng and Fine, 2008; Cheng *et al.*, 2010) with independent censoring scheme. In particular, this recent literature has focused on estimating a cause-specific version of localized Kendall's tau (see Bandeen-Roche and Ning, 2008). However, they analyze association between observed failure times along the lines of Oakes's non-parametric estimator (Oakes, 1989). Estimation is potentially biased since it is evaluated with observable failure times. Moreover, this ratio involves smoothing and the authors consider piecewise constant assumption in disjoint rectangular regions.

Scheike *et al.* (2010) and Scheike and Sun (2012) propose local association measures based on cross-odds ratio which has the advantages to take account for covariates. We follow this approach and consider non-parametric estimators of the cross-odd ratio. Contrary to standard bivariate competing risks model, there is strong dependence between the both indicators V_1 and V_2 induced by the structure of the model. Otherwise, we consider here that the Markov assumption is clearly violated and trying to measure how the process is Markov has limited interest. The interested reader for such a test could refer to Rodríguez-Girondo and de Uña-Álvarez (2012).

Let the cross-odds ratio for $(e,d) \in \mathcal{V}$ and $0 \leq s \leq t$ adapted to our model given by

$$\pi_0^{(e,d)}(s,t \mid \boldsymbol{u}) = \frac{\text{odds}\left(T \leq t, V_2 = d \mid S \leq s, V_1 = e, \boldsymbol{U} = \boldsymbol{u}\right)}{\text{odds}\left(T \leq t, V2 = d \mid V_1 = e, \boldsymbol{U} = \boldsymbol{u}\right)}, \tag{2.6}$$

where $\text{odds}(A) = \dfrac{\mathbb{P}(A)}{1-\mathbb{P}(A)}$. This measure is based on comparison between the conditional odd of the occurrence of the event $\{T \leq t, V = (e,d)\}$, given individual exits from healthy state before time s due to specific cause e, divided by the odds unconditionnaly the event $\{S \leq s\}$. Due to the definition of S and T, it is not possible to exchange s and t in $\pi_0^{(e,d)}(s,t \mid \boldsymbol{u})$ which is defined on the upper wedge $\{0 \leq s \leq t\}$. In case of independence for the event $\{T \leq t, V_2 = d\}$ and $\{S \leq s, V_1 = e\}$ conditionally to $\{V_1 = e, \boldsymbol{U} = \boldsymbol{u}\}$, i.e. the time and the cause of entry in a non-terminal event has no effect, we have $\pi_0^{(e,d)}(s,t \mid \boldsymbol{u}) = 1$. Positive (respectively negative) association is observed if $\pi_0^{(e,d)}(s,t \mid \boldsymbol{u}) > (<)\, 1$ and this measure takes its values in $[0,\infty[$. This quantity is easier to handle and is a valuable tool to analyze duration dependence contrary to the cumulative incidence function $F_0^{(v)}$ which is

relatively complicated.

Alternatively, remark that Cheng *et al.* (2007) gives simple association measures without covariates defined as the ratio of bivariate cumulative intensity function divided by the associated univariate cumulative intensity functions.

3 Non-parametric estimation, asymptotic properties and applications

Our aim here is to provide non-parametric estimation and asymptotic results for the sub-distribution functions $F_0^{(v)}$, $v \in \mathcal{V}$, the transition probabilities defined in subsection 2.2 and the association measures introduced in subsection 2.3. To do this, we introduce a general estimation framework for Aalen-Johansen integrals in presence of covariates in subsection 3.1 and demonstrate asymptotic properties 3.2. Possible improvement to taken account properly of left-truncation is discussed in Subsections 3.3. Subsections 3.4 and 3.5 are devoted to application of our results for transition probabilities and association measure.

3.1 General setup

The problem that we consider is in presence of unique right-censoring process C in respect of Assumption 1. Further, we ignore left-truncation for the ease of the presentation. Thus, we define the observation of i-th individual of a sample of length $n \geq 1$ is either of the form

$$(Y_i, \gamma_i, \gamma_i V_{1,i}, Z_i, \delta_i, \delta_i V_{2,i}, \boldsymbol{U}_i) \;\; 1 \leq i \leq n \;,$$

which are assumed to be i.i.d. replications of the variable $(Y, \gamma, \gamma V_1, Z, \delta, \delta V_2, \boldsymbol{U})$. In case of $\delta = 1$, we have of course $\gamma = 1$.

Consider first the ordered Y-values $Y_{1:n} \leq Y_{2:n} \leq \ldots \leq Y_{n:n}$ and $\left(\gamma_{[i:n]}, \boldsymbol{U}_{[i:n]}\right)$ the concomitant of the i-th order statistic (i.e. the value of $(\gamma_j, \boldsymbol{U}_j)_{1\leq j\leq n}$ paired with $Y_{i:n}$), an estimator for H_0 is simply obtained from the multivariate Kaplan-Meier estimator considered by Stute (1993)

$$\widehat{H}_{0n}(s, \boldsymbol{u}) = \sum_{i=1}^{n} W_{in} \mathbb{1}_{\left\{Y_{i:n} \leq s, \boldsymbol{U}_{[i:n]} \leq \boldsymbol{u}\right\}}, \tag{3.1}$$

where the Kaplan-Meier weight for the the i-th ordered observation is

$$W_{in} = \frac{\gamma_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\gamma_{[j:n]}}.$$

Kaplan-Meier integral taking the form $S(\varphi) = \int \varphi \, dH_0$ with some φ generic function are well-know is the literature and are estimated with

$$\widehat{S}_n(\varphi) = \int \varphi(s, \boldsymbol{u}) \; \widehat{H}_{0n}(ds, d\boldsymbol{u}) = \sum_{i=1}^{n} W_{in} \varphi\left(Y_{i:n}, \boldsymbol{U}_{[i:n]}\right).$$

Considering the bivariate setup of our model, the bivariate cumulative incidence function for (S, T, V_1, V_2) is defined and estimated nonparametrically by Cheng *et al.* (2007) under independent right-censoring. However, their representation is devoted to general bivariate competing risks data and we aim to provide estimators which exploit information that S is necessary observed when T is not censored. For this purpose, it is convenient to introduce $Z_{1:n} \leq Z_{2:n} \leq \ldots \leq Z_{n:n}$ the ordered Z-values and $\left(Y_{[i:n]}, \delta_{[i:n]}, J^{(v)}_{[i:n]}, \boldsymbol{U}_{[i:n]}\right)$ the concomitant of the i-th order statistic with $J^{(v)}_i = \mathbb{1}_{\{V=v,\}}$ with $v \in \mathcal{V}$. Based on the idea of Meira-Machado *et al.* (2006), we consider S as a covariate and propose estimator of $F^{(v)}_0$ which are termed as Aalen-Johansen estimator (Aalen and Johansen, 1978) and has the following form

$$\begin{aligned} \widehat{F}^{(v)}_{0n}(y, z, \boldsymbol{u}) &= \sum_{i=1}^{n} \widetilde{W}^{(v)}_{in} \mathbb{1}_{\left\{Y_{[i:n]} \leq y, Z_{i:n} \leq z, \boldsymbol{U}_{[i:n]} \leq \boldsymbol{u}\right\}} \\ &= \sum_{i=1}^{n} \widetilde{W}_{in} J^{(v)}_{[i:n]} \mathbb{1}_{\left\{Y_{[i:n]} \leq y, Z_{i:n} \leq z, \boldsymbol{U}_{[i:n]} \leq \boldsymbol{u}\right\}}, \end{aligned} \tag{3.2}$$

where $\widetilde{W}_{in}$ denotes the Kaplan-Meier weight of the i-th ordered observation related to estimated survival function of T and the Aalen-Johansen weight are

$$\widetilde{W}^{(v)}_{in} = \frac{\delta_{[i:n]} J^{(v)}_{[i:n]}}{n - i + 1} \prod_{j=1}^{i-1} \left(\frac{n - j}{n - j + 1}\right)^{\delta_{[j:n]}}.$$

The Aalen-Johansen weights $\widetilde{W}^{(v)}_{in}$, $1 \leq i \leq n$, for state v are very close the Kaplan-Meier weights related to the estimated survival function of T and can be interpreted as the mass associated to one observation. Further, we remark than IPCW representation can be easily derived from the expression of Aalen-Johansen weights as

$$\widetilde{W}^{(v)}_{in} = \frac{\delta_{[i:n]} J^{(v)}_{[i:n]}}{n\left(1 - \widehat{G}_n\left(Y_{i:n}\right)\right)},$$

where $\widehat{G}_n$ represents the Kaplan-Meier estimator of the distribution function of C. That opens a possible way to extend the results of IPCW theory and relax the assumptions concerning the censoring process, see e.g. Datta and Satten (2002).

Based on the representation as a sum of (3.2), we are now interesting in pertaining estimators of general quantities $S^{(v)}(\varphi) = \int \varphi \, dF^{(v)}_0$ with φ a generic function. In absence

of censoring process, non-parametric estimation is straightforward using integrals under the empirical multivariate distribution function of $(S, T, \boldsymbol{U})$. In this context, we have complete information and each observation has the same weight into the empirical process. Since the joint distribution of (T, V) has the aspect of a competing risks model, we estimate $S^{(v)}(\varphi)$ by computing the Aalen-Johansen integral of the form

$$\widehat{S}_n^{(v)}(\varphi) = \int \varphi(s,t,\boldsymbol{u})\ \widehat{F}_{0n}^{(v)}(ds,dt,d\boldsymbol{u}) = \sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right). \tag{3.3}$$

These estimators are similar to those exhibit by Suzukawa (2002) but the difference with this author is the presence of covariates in our approach. It is needed to derive asymptotic properties as we consider Y in (3.3) as an uncensored covariate. The representation of (3.3) as a sum is justify as easily to handle. Remark it admits also representation in terms of counting process by using the Kaplan-Meier product-limit estimator, see e.g. Allignol *et al.* (2013, Equation (9)).

3.2 Asymptotic properties

Let τ_Y and τ_Z the least upper bounds of the distribution functions of Y and Z. Under the Assumptions 1 and 2, the consistency and weak convergence of the estimator (3.1) on $[0, \tau_Y]$ can be easily demonstrated since the Stute (1993) and the Stute (1996) conditions are verified, see also Meira-Machado *et al.* (2006) without additional covariates. Note that this result is also verified if H and G have no jumps in common which is a more less restrictive condition than continuity assumption of distribution functions. Thus, we focus for the rest of this subsection on the consistency and weak convergence of the estimator (3.3).

Theorem 1. *Under the Assumptions 1 and 2 and assuming that φ is an F_0-integrable function. Then, we have with probability 1*

$$\widehat{S}_n^{(v)}(\varphi) \longrightarrow S_\infty^{(v)}(\varphi) = \int \mathbb{1}_{\{t<\tau_Z\}} \varphi(s,t,\boldsymbol{u})\ F_0^{(v)}(ds,dt,d\boldsymbol{u})\ ,\ v \in \mathcal{V}. \tag{3.4}$$

In addition, if the support of Z is included in those of C, we have obviously $\widehat{S}_n^{(j)}(\varphi) \to S^{(j)}(\varphi)$ w.p.1.

The present theorem constitutes an extension of the results demonstrated by Suzukawa (2002, Theorem 1) which are directly based on the proof of the Stute and Wang (1993) theorem's. More details about the proof is given in Appendix 1.

To obtain weak convergence properties, we adapt the approach followed by Stute (1995) for Kaplan-Meier integrals and Stute (1996) for the version with covariates. We define similar integrability conditions for any function φ F_0-integrable to prove a general convergence results. These conditions are given below.

Assumption 3. $\int \frac{\varphi(S,T,\boldsymbol{U})^2 \delta}{(1-G(T))^2} d\mathbb{P} = \int \frac{\varphi(S,T,\boldsymbol{U})^2}{1-G(T)} d\mathbb{P} < \infty.$

Assumption 4. $\int |\varphi(S,T,\boldsymbol{U})| \sqrt{C_0(T)} \mathbb{1}_{\{T<\tau_Z\}} d\mathbb{P} < \infty.$

Here, we denote

$$M(z) = \mathbb{P}(Z \leq z), M_0(z) = \mathbb{P}(Z \leq z, \delta = 0),$$
$$M^{(v)}(y,z,\boldsymbol{u}) = \mathbb{P}(Y \leq y, Z \leq z, \boldsymbol{U} \leq \boldsymbol{u}, \delta = 1, V = v),$$

and

$$C_0(x) = \int_0^{x-} \frac{G(dy)}{(1-M(y))(1-G(y))}.$$

We also introduce the functions

$$\lambda_1^{(v)}(x) = \frac{1}{1-M(x)} \int \frac{\varphi(s,t,\boldsymbol{u}) \mathbb{1}_{\{x<t<\tau_Z\}}}{(1-G(t))} M^{(v)}(ds,dt,d\boldsymbol{u}),$$

and

$$\lambda_2^{(v)}(x) = \int \frac{\lambda_1^{(v)}(\tau) \mathbb{1}_{\{\tau<x\}}}{1-M(\tau)} M_0(d\tau).$$

Let us discuss about these assumptions. Assumption 3 corresponds to a variance type assumption on φ guaranteeing the existence of finite second moments. The second condition is nothing but it originates from Stute (1995) and guarantees the distribution convergence results on $[0, \tau_Z]$. Notice that this last assumption is relatively low.

Now, we put $\widehat{\boldsymbol{S}}_n(\varphi) = \left(\widehat{S}_n^{(v)}(\varphi)\right)_{v\in\mathcal{V}}^\top$ and $\boldsymbol{S}(\varphi) = \left(S^{(v)}(\varphi)\right)_{v\in\mathcal{V}}^\top$. The following theorem gives asymptotic property for $\widehat{\boldsymbol{S}}_n(\varphi)$.

Theorem 2. *Suppose that the assumptions of Theorem 1 are verify and the support of Z is included in those of C. Under the Assumptions 3 and 4, we have*

$$\sqrt{n}\left\{\widehat{\boldsymbol{S}}_n(\varphi) - \boldsymbol{S}(\varphi)\right\} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\varphi)), \tag{3.5}$$

where $\boldsymbol{\Sigma}(\varphi)$ is a $m \times m$ symmetric matrix associated to the covariance matrix of the vector $\boldsymbol{a}(\varphi) = (a_v(\varphi))_{v\in\mathcal{V}}$ where

$$a_v(\varphi) = \frac{\varphi(Y,Z,\boldsymbol{U})\delta J^{(v)}}{1-G(Z)} + \lambda_1^{(v)}(Z)(1-\delta) - \lambda_2^{(v)}(Z), \ v \in \mathcal{V}.$$

The proof of this theorem is postponed in Appendix 2. From the Equation (3.5), we can obtain asymptotic confidence intervals if functions $\frac{1}{1-G}$, $\lambda_1^{(v)}$ and $\lambda_2^{(v)}$ were known.

This can be done by just replacing the distribution functions H, M, M_0 and M^v involved in the expression of $\boldsymbol{\Sigma}(\varphi)$ by their empirical counterparts. However, this calculation may be laborious due to the expression of $\boldsymbol{a}(\varphi)$ and implementing non-parametric bootstrap procedure is the most appropriate way to obtain asymptotic variance-covariance estimator.

3.3 Possible improvement for handing left-truncation

In this subsection, we discuss possible outlook to handle left-truncated data issue but formal proof of asymptotic results is out of the scope of this paper.

For the sake of simplicity, we ignore in the present paper the case of left-truncation process L. In the context of LTC insurance data that we study in Section 5, the left-truncation process is not really an issue as left-truncation times occur always when insured is still in the healthy state and when the amount of transition toward a non-terminal event is very low. Hence, Y and Z are observed if $Y \geq L$. By revisiting Assumptions 1 such as (C, L) are independent to (S, T, V), this can be taken into account by removing data such as Y_i, $1 \leq i \leq n$, is smaller than L_i (Andersen *et al.*, 1993, Chapter III.3). That would conduct to a small bias in our application.

However, more complicated situations may be envisaged in a more general framework if the left-truncation events can occur after S. Indeed, this situation implies issues for estimation $\mathbb{P}(S \leq s, V_1 = e)$ with $s \geq 0$ and $e \in \mathcal{C}(a_0)$ and simply remove truncated observations may incur considerable loss of information, see Peng and Fine (2006). In such framework, we suggest to refine our proposed product-limit integral estimators (3.3) for left-truncated and right-censored data drawing e.g. the representation defined by Sellero *et al.* (2005) and assuming particularly that (C, L) is independent of (S, T, V) and C is independent of L. The studying of this approach is out the scope of this paper but one could adapt the Aalen-Johansen weights such as

$$\widetilde{W}_{in}^{(v)} = \frac{\delta_{[i:n]} J_{[i:n]}^{(v)}}{nC_n(Z_{i:n})} \prod_{j=1}^{i-1} \left(1 - \frac{1}{nC_n(Z_{i:n})}\right)^{\delta_{[j:n]}},$$

where $C_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{L_i \leq x \leq Z_i}$.

3.4 Application for transition probabilities estimation

In this subsection, we consider the problem of estimating non-parametrically the transition probabilities introduced in 2.2 when the Markov assumption is non necessary fulfilled. In case the process is Markovian as it is often assumed for multi-state model, transition probabilities can be estimated non-parametrically with the so-called Aalen-Johansen estimator (see Aalen and Johansen, 1978; Andersen *et al.*, 1993). However, this methodology falls when the

Markov assumption is wrong, especially when the transition probabilities depend both the time and the duration (semi-Markov process models). Therefore, our estimation approach should not be limited by the Markov restriction.

To do this,we express (2.1), (2.2), (2.3) , (2.4) and (2.5) as integral of the form $\int \varphi \, dH_0$ and $\int \varphi \, dF_0^{(v)}$. So, we have

$$(2.1) = \frac{\mathbb{E}\left[\varphi^{(0)}_{t,\boldsymbol{u}}(S,\boldsymbol{U})\right]}{\mathbb{E}\left[\varphi^{(0)}_{s,\boldsymbol{u}}(S,\boldsymbol{U})\right]}, \tag{3.6}$$

$$(2.2) = \frac{\mathbb{E}\left[\varphi^{(1)}_{s,t,\eta,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(e,\mathcal{C}(e))\}}\right]}{\mathbb{E}\left[\varphi^{(0)}_{s,\boldsymbol{u}}(S,\boldsymbol{U})\right]}, \tag{3.7}$$

$$(2.3) = \frac{\mathbb{E}\left[\varphi^{(2)}_{s,t,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(e,\mathcal{C}(e))\}}\right]}{\mathbb{E}\left[\varphi^{(2)}_{s,s,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(e,\mathcal{C}(e))\}}\right]}, \tag{3.8}$$

$$(2.4) = \frac{\mathbb{E}\left[\varphi^{(3)}_{s,t,\eta,\zeta,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(e,d)\}}\right]}{\mathbb{E}\left[\varphi^{(4)}_{s,t,\eta,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(e,\mathcal{C}(e))\}}\right]}, \tag{3.9}$$

$$(2.5) = \frac{\mathbb{E}\left[\varphi^{(5)}_{s,t,\boldsymbol{u}}(S,T,\boldsymbol{U})\,\mathbb{1}_{\{V=(0,d)\}}\right]}{\mathbb{E}\left[\varphi^{(0)}_{s,\boldsymbol{u}}(S,\boldsymbol{U})\right]}, \tag{3.10}$$

with $\varphi^{(0)}_{s,u}(x,z) = \mathbb{1}_{\{x>s,z=u\}}$, $\varphi^{(1)}_{s,t,\eta,u}(x,y,z) = \mathbb{1}_{\{s<x\leq\min(t,t-\eta),y>t,z=u\}}$, $\varphi^{(2)}_{s,t,u}(x,y,z) = \mathbb{1}_{\{s<x,y>t,z=u\}}$, $\varphi^{(3)}_{s,t\eta,\zeta,u}(x,y,z) = \mathbb{1}_{\{s<x\leq t,\eta<y-x\leq\zeta,z=u\}}$, $\varphi^{(4)}_{s,t\eta,u}(x,y,z) = \mathbb{1}_{\{s<x\leq t,\eta<y-x,z=u\}}$ and $\varphi^{(5)}_{s,t,u}(x,y,z) = \mathbb{1}_{\{x>s,y\leq t,z=u\}}$.

Based on the idea developed by Meira-Machado *et al.* (2006), our approach consists in regarding the numerators and denominators in equations (3.6), (3.7), (3.8), (3.9) and (3.10) as expectation of a simple function of $(S,T,\boldsymbol{U})$ and estimating these expectations from Kaplan-Meier integrals and Aalen-Johansen integrals. Therefore, natural estimators for these quantities are

$$\widehat{p}_{00}(s,t \mid \boldsymbol{u}) = \frac{\widehat{S}_n\left(\varphi^{(0)}_{t,\boldsymbol{u}}\right)}{\widehat{S}_n\left(\varphi^{(0)}_{s,\boldsymbol{u}}\right)}, \tag{3.11}$$

$$\widehat{p}_{0e}(s,t,\eta \mid \boldsymbol{u}) = \frac{\widehat{S}^{(e,\mathcal{C}_e)}_n\left(\varphi^{(1)}_{s,t,\eta,\boldsymbol{u}}\right)}{\widehat{S}_n\left(\varphi^{(0)}_{s,\boldsymbol{u}}\right)}, \tag{3.12}$$

$$\widehat{p}_{ee}\left(s,t \mid \boldsymbol{u}\right) = \frac{\widehat{S}_n^{(e,\mathcal{C}_e)}\left(\varphi^{(2)}_{s,t,\boldsymbol{u}}\right)}{\widehat{S}_n^{(e,\mathcal{C}_e)}\left(\varphi^{(2)}_{s,s,\boldsymbol{u}}\right)}, \tag{3.13}$$

$$\widehat{p}_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right) = \frac{\widehat{S}_n^{(e,d)}\left(\varphi^{(3)}_{s,t,\eta,\zeta,\boldsymbol{u}}\right)}{\widehat{S}_n^{(e,\mathcal{C}_e)}\left(\varphi^{(4)}_{s,t,\eta,\boldsymbol{u}}\right)}, \tag{3.14}$$

$$\widehat{p}_{0d}\left(s,t \mid \boldsymbol{u}\right) = \frac{\widehat{S}_n^{(0,d)}\left(\varphi^{(5)}_{s,t,\boldsymbol{u}}\right)}{\widehat{S}_n\left(\varphi^{(0)}_{s,\boldsymbol{u}}\right)}, \tag{3.15}$$

Theorems 1 and 2 can be applied to derive asymptotic properties for (3.12), (3.13) , (3.14) and (3.15). The consistency and weak convergence of $\widehat{p}_{00}\left(s,t \mid \boldsymbol{u}\right)$ can be proved easily applying the results of Stute (1996) and the delta-method.

Proposition 3. *Under assumptions of Theorem 1, estimators for (3.7), (3.8) , (3.9) and (3.9) are consistent w.p.1 if the support of Z is included in those of C. Further, we deduce weak convergence properties for these estimators as functions $\varphi^{(1)}_{s,t,u,\eta}$, $\varphi^{(2)}_{s,t,u}$, $\varphi^{(3)}_{s,t,\eta,\zeta,u}$, $\varphi^{(4)}_{s,t,\eta,u}$ and $\varphi^{(5)}_{s,t,u}$ satisfy Assumptions 3 and 4. Thus, we have:*

- ***i.*** $\sqrt{n}\left\{\widehat{p}_{0e}\left(s,t,\eta \mid \boldsymbol{u}\right) - p_{0e}\left(s,t,\eta \mid \boldsymbol{u}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{0e}\left(s,t,\eta \mid \boldsymbol{u}\right)\right)$, *with* $0 \leq s \leq t$, $\eta \in [0,\infty[$ *and* $e \in \mathcal{C}_E\left(a_0\right)$,
- ***ii.*** $\sqrt{n}\left\{\widehat{p}_{ee}\left(s,t \mid \boldsymbol{u}\right) - p_{ee}\left(s,t \mid \boldsymbol{u}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{ee}\left(s,t \mid \boldsymbol{u}\right)\right)$, *with* $0 \leq s \leq t$ *and* $e \in \mathcal{C}_E\left(a_0\right)$,
- ***iii.*** $\sqrt{n}\left\{\widehat{p}_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right) - p_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right)\right)$, *with* $0 \leq s \leq t$, $0 \leq \eta \leq \zeta$, $e \in \mathcal{C}_E\left(a_0\right)$ *and* $d \in \mathcal{C}\left(e\right)$,
- ***iv.*** $\sqrt{n}\left\{\widehat{p}_{0d}\left(s,t \mid \boldsymbol{u}\right) - p_{0d}\left(s,t \mid \boldsymbol{u}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{0d}\left(s,t \mid \boldsymbol{u}\right)\right)$, *with* $0 \leq s \leq t$ *and* $d \in \mathcal{C}_D\left(a_0\right)$,

where $\sigma_{0e}\left(s,t,\eta \mid \boldsymbol{u}\right)$, $\sigma_{ee}\left(s,t \mid \boldsymbol{u}\right)$, $\sigma_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right)$ and $\sigma_{0d}\left(s,t \mid \boldsymbol{u}\right)$ are some limit variance functions to be precised.

Proof of the Proposition 3. First, the simple function $\varphi^{(0)}_{s,u}$ satisfy conditions of Theorem 1.1 in Stute (1996) and therefore $\widehat{S}_n\left(\varphi^{(0)}_{s,\boldsymbol{u}}\right)$ admit consistent and weak convergence properties. Second, by applying the result of Theorem 1 to the functions $\varphi^{(1)}_{s,t,u,\eta}$, $\varphi^{(2)}_{s,t,u}$, $\varphi^{(3)}_{s,t,\eta,\zeta,u}$, $\varphi^{(4)}_{s,t,\eta,u}$, and $\varphi^{(5)}_{s,t,u}$ which are clearly F_0-integrable, we obtain respectively the consistency result. The proof for weak convergence is obviously obtained by application of Theorem 2 on our particular functions. The form of variances $\sigma_{0e}\left(s,t,\eta \mid \boldsymbol{u}\right)$, $\sigma_{ee}\left(s,t \mid \boldsymbol{u}\right)$, $\sigma_{ed}\left(s,t,\eta,\zeta \mid \boldsymbol{u}\right)$ and $\sigma_{0d}\left(s,t \mid \boldsymbol{u}\right)$ follows with delta method but are not easy. □

This proposition enlarges the results of Meira-Machado *et al.* (2006, Corollary 1,2). As seen above, the variance functions of the limiting Gaussian process is tricky to estimate and can be computed by means of bootstrap techniques. In section 5, we construct non-parametric bootstrap pointwise confidence bands for our estimators. This is done with simple bootstrap resampling procedure (Efron, 1979). Recently, Beyersmann *et al.* (2013) provides wild bootstrap approach for the Aalen-Johansen estimator for competing risks data but, as it is remarked in their paper, this approach is quite close to those given by Efron. To our knowledge, no further tentative has been proposed to more consistent bootstrap methodology for cumulative intensity function or other transition probabilities.

3.5 Application for estimation of association and testing dependence assumption

We are now interesting in applying our previous results for measuring nonparametrically cause-specific associations between the failure times S and T. For this, we exhibit non-parametric estimator of (2.6) and given its asymptotic properties. For the sake of simplicity, we ignore here the covariate vector $\boldsymbol{U}$. For $(e,d) \in \mathcal{V}$ and $0 \leq s \leq t$, the cross-odds ratio is function of

$$\mathbb{P}\left(T \leq t, V_2 = d \mid S \leq s, V_1 = e\right) = \frac{\mathbb{P}\left(T \leq t, V_2 = d, S \leq s, V_1 = e\right)}{\mathbb{P}\left(S \leq s, V_1 = e\right)},$$

and

$$\mathbb{P}\left(T \leq t, V_2 = d \mid V_1 = e\right) = \frac{\mathbb{P}\left(T \leq t, V_2 = d, V_1 = e\right)}{\mathbb{P}\left(V_1 = e\right)}.$$

The former admit the following non-parametric estimator

$$\frac{\widehat{F}_{0n}^{(e,d)}\left(s,t\right)}{\widehat{H}_{0n}^{(e)}\left(s\right)},$$

where $\widehat{H}_{0n}^{(e)}$ is the estimator of the sub-distribution function related to S i.e.

$$\widehat{H}_{0n}^{(e)}\left(s\right) = \sum_{i=1}^{n} W_{in} \mathbb{1}_{\left\{V_{1,[i:n]}=e\right\}} \mathbb{1}_{\{Y_{i:n} \leq s\}}.$$

The latter can be estimated considering

$$\frac{\widehat{F}_{n}^{(e,d)}\left(t\right)}{\widehat{H}_{0n}^{(e)}\left(\infty\right)}.$$

Hence, we have the following non-parametric estimator for the cross-odds ratio

$$\widehat{\pi}_{0n}^{(e,d)}\left(s,t\right) = \frac{\widehat{F}_{0n}^{(e,d)}\left(s,t\right)}{\frac{\widehat{H}_{0n}^{(e)}\left(s\right) - \widehat{F}_{0n}^{(e,d)}\left(s,t\right)}{\widehat{F}_{n}^{(e,d)}\left(t\right)}}, \tag{3.16}$$

Proposition 4. *Under assumptions of Theorem 1, time dependent association estimator (3.16) is consistent w.p.1 if the support of Z is included in those of C. Then, we have*

$$\sqrt{n}\left\{\widehat{\pi}_{0n}^{(e,d)}(s,t)-\pi_{0}^{(e,d)}(s,t)\right\}\xrightarrow{d}\mathcal{N}\left(0,\sigma_{\pi}^{(e,d)}(s,t)\right),$$

with $0\leq s\leq t$, and $(e,d)\in\mathcal{V}$ and where $\sigma_{\pi}^{(e,d)}(s,t)$ is some variance function to be precised.

Proof of the Proposition 4. Theorems 1 and 2 are verify for estimator of the sub-distribution function $\widehat{F}_{0n}^{(e,d)}$. The results follows easily using the convergences properties of the Aalen-Johansen estimators (Andersen *et al.*, 1993) and the delta-method for weak convergence. □

By asymptotic normality of (3.16), we can measure locally the association for particular times s and t. For variance calculation purpose, bootstrap procedure is appropriate way and we suggest using simple resampling procedure as discuss above. To avoid regarding only some arbitrary time points, we develop global measure on the upper wedge $\{0\leq s\leq t\}$ with integrated weighted average version of the cross odds-ratio. Analogously approach is provided by Cheng *et al.* (2007) for bivariate competing risks data to set-up independence test based on the ratio of the conditional cumulative incidence function divided by the unconditional cumulative incidence function. They have obtained convergence results for their estimators with different road than ours using counting processes convergence properties and proving both the bivariate cause-specific hazard and the bivariate cumulative incidence functions are Hadamard differentiable maps on appropriate domains.

Let $\mu^{(e)}=\left[\mu_1^{(e)},\mu_2^{(e)}\right]$ and $\mu^{(d)}=\left[\mu_1^{(d)},\mu_2^{(d)}\right]$ two times intervals associated to causes e and d corresponding to the study period and introduce the following test statistic as

$$\widehat{\pi}_{0n}^{(e,d)*}=\iint_{\mu^{(e)}\times\mu^{(d)}}\mathbb{1}_{\{s\leq t\}}\widetilde{w}_n(s,t)\,\widehat{\pi}_{0n}^{(e,d)}(s,t)\;dsdt, \tag{3.17}$$

where

$$\widetilde{w}_n(s,t)=\frac{\widehat{w}_n(s,t)}{\iint_{\mu^{(e)}\times\mu^{(d)}}\mathbb{1}_{\{s\leq t\}}\widehat{w}_n(s,t)\;dsdt},$$

representing weight function that we assume to converge uniformly to $w(s,t)$, a bounded deterministic function, over $\mu^{(e)}\times\mu^{(d)}$. We next derive asymptotic properties for $\widehat{\pi}_{0n}^{(e,d)*}$.

Proposition 5. *Under assumptions of Proposition 4, $\widehat{\pi}_{0n}^{(e,d)*}$ is consistent and $\sqrt{n}\left(\widehat{\pi}_{0n}^{(e,d)*}-\pi^{(e,d)*}\right)$ is asymptotically normal with mean* 1.

Proof of the Proposition 5. First, we note than $\widehat{\pi}_{0n}^{(e,d)*}$ is Hadamard differentiable at $\widehat{\pi}_{0n}^{(e,d)}$. By the results of Proposition 4 and applying the functionnal delta-method (van der Vaart and Wellner, 2000, Theorem 3.9.4), the weak consistency result follow. □

The choice of the weight function is important to have meaningfull global association measure and appropriate choice consist in accentuate dependencies times points. Practical approach requires dedicated analysis to investigated times points where there is large associations. As remarked by Cheng *et al.* (2007), this may become complex when $\pi^{(e,d)}(s,t)$ is positive on some regions of $\mu^{(e)} \times \mu^{(d)}$ and negative on others as this can reduce the significance of the measure. A suitable choice for the weight function is needed but is out the scope of this paper.

4 Simulation results for transition probabilities

In this section, we deploy simulation approach in order to assess the performance of our new estimators (3.12) and (3.13) versus the Aalen-Johansen estimators. In the following, the computation are carried out with the help of the software `R` (R Core Team, 2014). As our model is based on competing risks models, the `R`-package *mstate* designed by De Wreede *et al.* (2011) and the book of Beyersmann *et al.* (2011) give useful initial toolkits for the development of our code. Note also that the model provides by Meira-Machado *et al.* (2006) is also implemented in `R`, see Meira-Machado and Roca-Pardinas (2011). Our script is available on request from the first author.

For simulation purpose, we need in our procedure accounting for dependence between each latent survival times and not fulfilling the Markov assumption. In presence of only one disease state, i.e. simple illness-death model, Amorim *et al.* (2011) use Farlie-Gumbel-Morgenstern copula to model the bivariation distribution of gap times $(S, T-S)$. But as our model is a quite general, we consider the simulation approach set up by Rotolo *et al.* (2013) applying to a multi-state model as depicted in Figure 2. For the sake of simplicity and without loss of generality, we consider only one terminal event and two non-terminal events.

[Figure 2 about here.]

First, we set Clayton copula $\mathfrak{C}_{\theta_0}$ with dependent parameter θ_0 to combine the failure times T_{0e}, $e \in \{e_1, e_2\}$, and T_{0d} from the starting state a_0

$$\begin{aligned}\mathbb{P}\left(T_{0e_1} > t_{e_1}, T_{0e_1} > t_{e_2}, T_{0d} > t_d\right) &= \mathfrak{C}_{\theta_0}\left(\mathbb{P}\left(T_{0e} > t_{e_1}\right), \mathbb{P}\left(T_{0e} > t_{e_2}\right), \mathbb{P}\left(T_{0d} > t_d\right)\right) \\ &= \left(1 + \mathbb{P}\left(T_{0d} > t_d\right)^{-\theta_0} + \sum_{e \in \{e_1, e_2\}} \left[\mathbb{P}\left(T_{0e} > t_e\right)^{-\theta_0} - 1\right]\right)^{-1/\theta_0},\end{aligned}$$

At a second step, we define another Clayton copula $\mathfrak{C}_{\theta_e}$ for each non-terminal event e_1,

e_2 to put together the latent times T_{0e} and its children T_{ed}. That gives for $e \in \{e_1, e_2\}$

$$\begin{aligned}\mathbb{P}\left(T_{0e} > t_e, T_{ed} > t_d\right) &= \mathfrak{C}_{\theta_e}\left(\mathbb{P}\left(T_{0e} > t_e\right), \mathbb{P}\left(T_{ed} > t_d\right)\right) \\ &= \left(1 + \mathbb{P}\left(T_{0e} > t_e\right)^{-\theta_e} + \mathbb{P}\left(T_{ed} > t_d\right)^{-\theta_e}\right)^{-1/\theta_e}.\end{aligned}$$

With this setting, we assume the dependent parameters θ_0 and θ_e, $e \in \{e_1, e_2\}$, have the same value $\theta = 0.5$ for each copula models. We consider the latent failure times T_{0e}, T_{ed}, $e \in \{e_1, e_2\}$, and T_{0d} follows Weibull distributions Wei (λ, ρ) with the respective scale λ and shape ρ parameters. Associated to each competing risks blocks, i.e. we have three independent censoring process C_0, C_1 and C_3, we simulate two different level of independent censoring from Exponential distributions Wei $(\lambda, 1)$ to assess the performance of our estimator under moderate and medium censoring. Table 1 shows the chosen values for the parameters.

[Table 1 about here.]

The proportion of censoring related to the both moderate scenario results in 21% of censoring for S and in 34% of censoring for the time $T - S$. Under the medium scenario, there are respectively of 24% and 47%. For each scenario, we consider three samples with size $n = 200$, $n = 400$ and $n = 800$ based on $K = 1000$ replicated datasets.

Then, we focus especially on $\widehat{p}_{0e}(s, t, 0)$ and $\widehat{p}_{ee}(s, t)$ to measure the performance of our methodology for some fixed values of time $s = 5$, $s = 10$, $s = 15$. To do this, we follow Meira-Machado *et al.* (2006) and compute the bias, variance and mean square error and their integrated version on the interval $[s, s + 50]$[2]. Tables 2 and 3 report the results for the both censoring scenarios.

[Table 2 about here.]

[Table 3 about here.]

The simulations show for $\widehat{p}_{e_1e_1}(s, t)$ and $\widehat{p}_{e_2e_2}(s, t)$ with our non-parametric estimators (NP) is more relevant than the Aalen-Johansen estimators (AJ) both for moderate and medium censoring scenario. Tables 2 and 3 clearly illustrate the AJ estimator is highly biased compared to our estimators and this bias seems to increase with the size of the sample for AJ estimators. This conducts to higher MSE for AJ estimators. The good performance of NP estimators in terms of bias and MSE is offset by higher variance. In average, time spent in non-terminal state is higher for state e_2 than e_1, leading to higher bias for state e_2 regarding the AJ estimators. The remark is not true for the NP estimators as the bias mainly depends

[2]We approximate the integral by a sum on time intervals of width 0.2

here how censoring is important. Indeed, individuals in state e_1 for the medium scenario are more prone to censor while the situation is reversed for the moderate scenario. For both the AJ and NP estimators, we remark that the bias and the variance increase with the value of s.

At a second step, we compare by simulation our estimators $\widehat{p}_{0e_1}(s,t,0)$ and $\widehat{p}_{0e_2}(s,t,0)$ to their Aalen-Johansen equivalent. These results are more complex as in a such context the bias depends to the choice of marginal distributions, the censoring scenarios, the interval $[s, s+50]$ and copula functions used. Indeed, observations of local bias[3] lead to identify some particular time regions where the NP estimators are clearly relevant and other regions where it appears that these estimators are seriously biased. This picture is illustrated in Tables 2 and 3 and are in contrast to the results observed by Meira-Machado *et al.* (2006) as their simulation process does not allow to extract this behavior. Similar contradiction is also observed by Allignol *et al.* (2013) with comparable estimators but different simulation approach.

For $p_{0e_1}(s,t,0)$, the Markovian estimator has a lower bias for $s = 5$, except for the moderate censoring scenario when $n = 800$, and it becomes less relevant in terms of bias for $s = 10$ and $= 15$ in the both censoring scenario. However, this better behavior observed for the NP estimators compared to the AJ estimators seems to decrease with the size of the sample. The reason is the bias of NP estimators decrease with the size of the observed sample for time regions where s and t are sufficiently close, i.e. local dependence between S and $T-S$ is high. Regarding to the integrated variance estimators, the simulations indicate the NP and AJ estimators have similar figures contrary to the NP estimators exhibit for $p_{e_1e_1}(s,t)$ and $p_{e_2e_2}(s,t)$. Simulations for $p_{0e_2}(s,t,0)$ allows exhibiting similar phenomena but with some magnification due to parameters chosen for T_{0e_2}.

5 Application to LTC insurance data

This section described our data set and discuss the results obtained with our non-parametric estimators. Finally we measure the uncertainty on the estimated transition probabilities with non-parametric bootstrap process.

5.1 Data description

The data set that we analyze here is an extract from the database of a french LTC insurer. It is almost the same dataset used to estimate the probabilities of entry into dependency in Guibert and Planchet (2014). These data are also studying by Tomas and Planchet (2013) with adaptive non-parametric approach for smoothing the survival law of LTC claimants but without distinguished the effect of each disease.

[3]Not shown here but available on demand.

The data are longitudinal with independent right-censoring and left-truncation and comparing to the previous study, we have improved the data by adding the individual living path after entry into dependency distinguish per disease. The period of observation stretches from the beginning of 1998 to the end of 2010 and the range of ages is 65-90. We refer to Guibert and Planchet (2014) and Tomas and Planchet (2013) for more detailed description of this dataset. Regarding pathologies, the data are grouped within 4 groups: neurological pathologies e_1, various pathologies e_2, terminal cancers e_3 and dementia e_4. Considering death d_1 and cancellation d_2 as other exit causes from the initial state, the multi-state structure related to our dataset is shown in Figure 3. In this model, the lifetime after entry into dependency clearly depends to the age and the duration from the entry.

[Figure 3 about here.]

5.2 Estimation results for transition probabilities and association measures

In this section, we perform estimation of transition probabilities (3.12) and (3.14). Figure 4 displays the annual probabilities of becoming dependent between s and $s+1$ year and staying at least one month in disability state.

[Figure 4 about here.]

We also present for comparison the annual incidence rates for each disease computed with simple Aalen-Johansen estimators for competing risks data (see Guibert and Planchet, 2014, Section 4.2). The corresponding pointwise 95% confidence interval are obtained from 500 bootstrap resamples employing the asymptotic normality and bootstrap standard errors estimates. For each diseases, transition probabilities (3.12) globally grow over times and we observe the gap between incidence rates and transition probabilities increases rapidly after age 75. For terminal cancers, the incidence probabilities by age 80 is 3.7 times than transition probabilities due to high death rates after entry in dependency.

We give in Figure 5 the crude surface of monthly death rates $p_{ed}(s, s+1, \eta, \eta+1/12)$ with our model for each entry cause. For the sake clarity, these rates are presented in terms of age of occurrence[4] and duration times in dependency state. The range of ages of occurrence is $70-90$ and the maximum duration is 60 months. Beforehand, we have removed rates upper than 0.5 which are clearly not reliable.

[Figure 5 about here.]

[4]We have considered integer ages.

The death rates are not observed for some regions due to lack in data in particular for duration above the first twelve months. However, we obtain that causes are very different and this clearly means that specific association structure is needed to model this phenomena. In particular, the model shows extreme death rates for terminal cancers for the first six months. We now use the cross-odds ratio to measure association. We compute $\log\left(\widehat{\pi}_{0n}^{(e,d)}(s, s+t)\right)$ along with 95% confidence interval at age $s = 70, 75, 80, 85$, and 90 and duration $t = 1, 6, 12$ and 24 months obtained with 500 bootstrap samples. The results are given in Table 4. For each disease, the association decreases over age s and decreases slightly with duration for a given age s. This confirms the strong effect entry in non-terminal events on survival time. Association for terminal cancers is no longer available beyond 6 months at age 90 due to lack in data.

[Table 4 about here.]

6 Discussion

This paper focus on the non-parametric analysis of particular acyclic multi-state models relaxing the Markov assumption in presence of independent right-censoring. Based on competing risks set-up distinghishing two different lifetimes, we have proposed generalized Aalen-Johansen estimator for the treatment of such data. The asymptotic properties can be derived by adapting results obtained for Kaplan-Meier integrals with covariates to competing risks data. This allows exhibiting estimators for transition probabilities which can be used to check any assumptions usually done in applications with observed data. Due to the bivariate structure of our model, we attempt to measure association between the both introduced failure times for each joint causes. The measure chosen is inspired by those recently developed for bivariate competing risks data with adaptation to consider mechanical association between failure causes.

The simulation study demonstrate the relevance of our approach for estimation of particular transition probabilities. However, our estimators become locally superior to the so-called Aalen-Johansen estimator for transition probabilities of the form (2.2) for time points where the Markov assumption is cleary rejected. This seems to be well adapted for actuarial needs with LTC insurance data as transition probabilities between two time points materializing a short period of time. Besides, a interesting room for future researches is to develop method to reduce the biais of our Aalen-Johansen integrals estimators with finite sample. The work of Amorim *et al.* (2011) is a natural direction to follow when covariates are not present.

Our non-parametric estimators are applicable with observed covariates. However, it may be difficult to apply directly these estimators due to the dimension of the problem. An outlook for future investigations consist in analysis non-parametric and semi-parametric regression with the framework introduced in this paper. For application, it seems semi-parametric approach for the cross odds ration developed by Scheike and Sun (2012) may be

interesting to consider. There is an another issue in constructing some goodness-of-fit tests which require to compute appropriate critical values. We have considered a classic Efron's bootstrap with resampling approach for our applications. Nevertheless with survival data in presence of censoring, the Efron's methodology (Efron, 1981) is generally choose since this is more consistent. Such a procedure is left for future works.

Acknowledgements

We are grateful to Olivier Lopez and Christian Robert for the valuable advice and suggestions which they provided in relation to the first draft of this paper. This work is partially funded by the *ANR project LoLitA : Dynamic models for human Longevity with Lifestyle Adjustments*.

Appendix 1 Proof of Theorem 1

Let, for $i = 1, \ldots, n$ and $v \in \mathcal{V}$, $D_i^{(v)} = \left(Y_i, \delta_i, J_i^{(v)}, \boldsymbol{U}_i\right)$ and for each $n \geq 0$, the σ-algebra

$$\mathcal{F}_n^{(v)} = \sigma\left(Z_{i:n}, D_{[i:n]}^{(v)}, 1 \leq i \leq n, Z_{n+1}, D_{n+1}^{(v)}, \ldots\right),$$

where $D_{[i:n]}^{(v)}$ are the value paired with $Z_{i:n}$.

Clearly for $v \in \mathcal{V}$, $\widehat{S}_n^{(v)}(\varphi)$ is adapted to $\mathcal{F}_n^{(v)}$ and $\mathcal{F}_n^{(v)}$ is decreasing and converge towards $\mathcal{F}_\infty^{(v)} = \bigcap_{n\geq 1} \mathcal{F}_n^{(v)}$. Our strategy, following Stute and Wang (1993), is to demonstrate that $\left(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geq 0\right)$ is a reverse-time supermartingale and then applying convergence result given by Neveu (1975, Proposition V-3-11, p. 116) to obtain consistency. For the following lemma, we consider that φ is a nonnegative fonction. Otherwise, the results remains applicable by decomposing φ into positive and negative parts.

Lemma 1. *For $\varphi \geq 0$ and assume that the distribution function of Z is continuous, $\left(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geq 0\right)$ is a reverse-time supermartingale for $v \in \mathcal{V}$.*

Proof. Denote by $\widehat{F}_n^{(v)}(z) = \sum_{i=1}^n \widetilde{W}_{in}^{(v)} \mathbb{1}_{\{Z_{i:n} \leq z\}}$ and let $\widehat{F}_n^{(v)}\{z\} = \widehat{F}_n^{(v)}(z) - \widehat{F}_n^{(v)}(z-)$, we can remark that

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^n \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\}.$$

If Z_{n+1} has rank k with $1 \leq k \leq n+1$, then $Z_{i:n} = Z_{i:n+1}$ for all $i < k$ and thus we have

$$\sum_{i=1}^{k-1} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\} = \sum_{i=1}^{k-1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{U}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\},$$

$$\sum_{i=k}^{n} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right) \widehat{F}_n^{(v)}\{Z_{i:n}\} = \sum_{i=k+1}^{n+1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{U}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\},$$

and

$$\varphi\left(Y_{[k:n+1]}, Z_{k:n+1}, \boldsymbol{U}_{[k:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{k:n+1}\} = 0.$$

Hence, we obtain that

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^{n+1} \varphi\left(Y_{[i:n+1]}, Z_{i:n+1}, \boldsymbol{U}_{[i:n+1]}\right) \widehat{F}_n^{(v)}\{Z_{i:n+1}\}. \qquad \text{(Appendix 1.1)}$$

Following the same lines of the proof of Lemma 2.2 in Stute and Wang (1993) (see also Stute (1993, Lemma 2.2)), we show with Lemma 2.1 of Stute and Wang (1993) applied to $D_{[i:n]}$ that

$$\mathbb{E}\left[\widehat{F}_n^{(v)}\{Z_{i:n+1}\} \mid \mathcal{F}_{n+1}^{(v)}\right] = \widetilde{W}_{i,n+1}^{(v)}, \ 1 \leq i \leq n$$

and

$$\mathbb{E}\left[\widehat{F}_n^{(v)}\left\{Z_{n+1:n+1}\right\} \mid \mathcal{F}_{n+1}^{(v)}\right] \leq \widetilde{W}_{n+1,n+1}^{(v)} .$$

Since $\varphi \geq 0$, the result follows immediately by writing the conditionnal expectation of (Appendix 1.1). □

From the Lemma 1, we have by applying the Proposition V-3-11 of Neveu (1975) that $\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi) \mid \mathcal{F}_\infty^{(v)}\right]$ admits limits $\mathbb{P}$-almost surely. Due to the Hewitt-Savage zero-one, we have $\mathcal{F}_\infty^{(v)}$ is trivial and then

$$\lim_{n\to\infty} \mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi) \mid \mathcal{F}_\infty^{(v)}\right] = \lim_{n\to\infty} \mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] = S_\infty^{(v)}(\varphi) .$$

Now, we aim to determinate the value of $S_\infty^{(v)}(\varphi)$. To do this, we let

$$m(z) = \mathbb{P}(\delta = 1 \mid Z = z) ,$$

$$\Psi_n(z) = \prod_{i=1}^{n} \left(1 + \frac{1 - m(Z_{i:n})}{n - i + 1}\right)^{\mathbb{1}_{\{Z_{i:n} < z\}}} ,$$

and for $v \in \mathcal{V}$

$$\widetilde{\varphi}^{(v)}(z) = \mathbb{E}\left[\varphi(Y, Z, \boldsymbol{U})\, \delta J^{(v)} \mid Z = z\right] .$$

Lemma 2. *Under the assumptions of the Lemma 1, we have for $v \in \mathcal{V}$*

$$\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] = \mathbb{E}\left[\widetilde{\varphi}^{(v)}(Z)\, \mathbb{E}\left[\Psi_{n-1}(Z)\right]\right] .$$

Proof. Let R_{jn} denote the rank of Z_j among $Z_1, \ldots, Z_n$, we can write

$$\begin{aligned}
\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \widetilde{W}_{in}^{(v)} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{n-i+1} \mathbb{E}\left[\begin{array}{l} \varphi\left(Y_{[i:n]}, Z_{i:n}, \boldsymbol{U}_{[i:n]}\right) \delta_{[i:n]} J_{[i:n]}^{(v)} \\ \times \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\delta_{[j:n]}} \mid Z_{1:n}, \ldots, Z_{n:n} \end{array}\right]\right] .
\end{aligned}$$

From Lemma 2.1 of Stute and Wang (1993) applied to $D_i^{(v)}$ for $i = 1, \ldots, n$, we have, conditionally on $Z_{1:n} < \ldots < Z_{n:n}$, the concomitants among the D's are independent. Hence,

$$
\begin{aligned}
\mathbb{E}\left[\widehat{S}_n^{(v)}(\varphi)\right] &= \mathbb{E}\left[\sum_{i=1}^{n}\frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n-i+1}\prod_{j=1}^{i-1}\mathbb{E}\left[\left(\frac{n-j}{n-j+1}\right)^{\delta_{[j:n]}} \mid Z_{j:n}\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n-i+1}\prod_{j=1}^{i-1}\left(1-\frac{m(Z_{j:n})}{n-j+1}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n}\prod_{j=1}^{i-1}\left(1+\frac{1-m(Z_{j:n})}{n-j}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\frac{\widetilde{\varphi}^{(v)}(Z_i)}{n}\prod_{j=1}^{n}\left(1+\frac{1-m(Z_j)}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j<Z_i\}}\right] \\
&= \mathbb{E}\left[\widetilde{\varphi}^{(v)}(Z_1)\prod_{j=1}^{n}\left(1+\frac{1-m(Z_j)}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j<Z_1\}}\right]. \qquad \text{(Appendix 1.2)}
\end{aligned}
$$

If $Z_j < Z_1$ then $R_{jn} = R_{j,n-1}$. Conditioning on Z_1, the result follows easily. □

Similar proof is established in Stute and Wang (1993)[Lemma 2.4] and reused in Stute (1994). Now, we are in position to prove the Theorem 1 by studying the process $\Psi_n(z)$.

Proof of the Theorem 1. From Stute and Wang (1993)[Lemma 2.5 and Lemma 2.6] and assuming that G and the distribution function of F are continuous, for each $z < \tau_Z$, we have

$$
\mathbb{E}\left[\Psi_n(z)\right] \uparrow \frac{1}{1-G(z)}. \qquad \text{(Appendix 1.3)}
$$

Hence, under the Assumption 2 and $\varphi \geq 0$, we obtain by applying Lemma 2, Equation (Appendix 1.3) and the monotone convergence theorem that

$$
\begin{aligned}
S_\infty^{(v)}(\varphi) &= \int \mathbb{1}_{\{Z<\tau_Z\}}\frac{\widetilde{\varphi}^{(v)}(Z)}{1-G(Z)}\,d\mathbb{P} \\
&= \int \mathbb{1}_{\{Z<\tau_Z\}}\mathbb{E}\left[\varphi(Y,Z,\boldsymbol{U})\,\delta J^{(v)} \mid Z\right]\frac{1}{1-G(Z)}\,d\mathbb{P} \\
&= \int \varphi(S,T,\boldsymbol{U})\,\frac{\mathbb{1}_{\{T<\tau_Z\}}\delta J^{(v)}}{1-G(T)}\,d\mathbb{P} \\
&= \int \varphi(S,T,\boldsymbol{U})\,\frac{\mathbb{1}_{\{T<\tau_Z\}}J^{(v)}}{1-G(T)}\mathbb{P}(T\leq C \mid S,V,T,\boldsymbol{U})\,d\mathbb{P} \\
&= \int \varphi(S,T,\boldsymbol{U})\,\frac{\mathbb{1}_{\{T<\tau_Z\}}J^{(v)}}{1-G(T)}\mathbb{P}(T\leq C \mid V,T)\,d\mathbb{P}.
\end{aligned}
$$

Since C and (V,T) are independent (see Assumption 1), we remark that $\mathbb{P}(T \leq C \mid V,T) = 1 - G(T)$. Hence, we obtain

$$S_{\infty}^{(v)}(\varphi) = \int \mathbb{1}_{\{t<\tau_Z\}} \varphi(s,t,\boldsymbol{u}) \; F_0^{(v)}(ds,dt,d\boldsymbol{u}). \qquad \text{(Appendix 1.4)}$$

As indicated earlier, from the Lemma 1, the Equation (Appendix 1.4), the proposition V-3-11 of Neveu (1975) and for a continuous $F^{(v)}$, the desired proof follows. □

Appendix 2 Proof of Theorem 2

Here, we denote

$$\widehat{M}_n(z) = \sum_{i=1}^{n} \mathbb{1}_{\{Z_i \leq z\}},$$
$$\widehat{M}_{0n}(z) = \sum_{i=1}^{n} \mathbb{1}_{\{Z_i \leq z, \delta_i = 0\}},$$
$$\widehat{M}_n^{(v)}(y,z,\boldsymbol{u}) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leq y, Z_i \leq z, \boldsymbol{U}_i \leq \boldsymbol{u}, \delta_i = 1, V_i = v\}},$$

the empirical distribution functions of M, M_0 and $M_0^{(v)}$. Directly based on Stute (1995) proof, our strategy is in 2 steps: prove CLT when φ vanishes to the right of some $\nu < \tau_Z$ and then extend it on $[0,\tau_Z]$. Note Suzukawa (2002) follows also the same way.

Lemma 3. *We have for* $v \in \mathcal{V}$

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i,Z_i,\boldsymbol{U}_i)\,\delta_i J_i^{(v)} \exp\left\{ n \int_0^{Z_i-} \ln\left\{ 1 + \frac{1}{n\left(1-\widehat{M}_n(\tau)\right)} \right\} \widehat{M}_{0n}(d\tau) \right\} \qquad \text{(Appendix 2.1)}$$

Proof. From the same rationale used to obtain (Appendix 1.2), we find

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i,Z_i,\boldsymbol{U}_i)\,\delta_i J_i^{(v)} \prod_{j=1}^{n} \left(1 + \frac{1-\delta_j}{n-R_{jn}}\right)^{\mathbb{1}\{Z_j<Z_i\}}.$$

The result follows immediately by definition of $\widehat{M}_n(z)$ and $\widehat{M}_{0n}(z)$, see proof of Lemma 2.1 in Stute (1995). □

The exponential term in (Appendix 2.1) is expanded in Stute (1995) such as

$$\exp\{\ldots\} = \frac{1}{1-G(Z_i)}(1 + B_{in} + C_{in}) + \frac{1}{2}\exp\{\Delta_i\}(B_{in}+C_{in})^2, \qquad \text{(Appendix 2.2)}$$

where

$$B_{in} = n \int_0^{Z_i-} \ln \left\{ 1 + \frac{1}{n\left(1 - \widehat{M}_n(\tau)\right)} \right\} \widehat{M}_{0n}(d\tau) - \int_0^{Z_i-} \frac{\widehat{M}_{0n}(d\tau)}{1 - \widehat{M}_n(\tau)},$$

$$C_{in} = \int_0^{Z_i-} \frac{\widehat{M}_{0n}(d\tau)}{1 - \widehat{M}_n(\tau)} - \int_0^{Z_i-} \frac{M_0(d\tau)}{1 - M(\tau)},$$

and Δ is between the 2 terms

$$n \int_0^{Z_i-} \ln \left\{ 1 + \frac{1}{n\left(1 - \widehat{M}_n(\tau)\right)} \right\} \widehat{M}_{0n}(d\tau) \text{ and } \int_0^{Z_i-} \frac{M_0(d\tau)}{1 - M(\tau)}.$$

Considering (Appendix 2.1) and (Appendix 2.2), we write

$$\begin{aligned} \widehat{S}_n^{(v)}(\varphi) = & \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i, Z_i, \boldsymbol{U}_i) \delta_i J_i^{(v)} \frac{1 + B_{in} + C_{in}}{1 - G(Z_i)} \\ & + \frac{1}{2n} \sum_{i=1}^{n} \varphi(Y_i, Z_i, \boldsymbol{U}_i) \delta_i J_i^{(v)} \exp\{\Delta_i\} (B_{in} + C_{in})^2. \end{aligned} \tag{Appendix 2.3}$$

Now, we decompose the last equation and study approximations for each component. To do this, we make for φ the following assumption

Assumption 5. φ is an F_0-integrable function such as $\int \varphi^2 dF_0 < \infty$ and $\varphi(s, t, \boldsymbol{u}) = 0$ for $\nu < t$ where $\nu < \tau_Z$.

This assumption aims to bound the denominators of the terms obtaining in the following lemmas.

Lemma 4. *Under Assumption 5, we have*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i, Z_i, \boldsymbol{U}_i) \delta_i J_i^{(v)} \frac{C_{in}}{1 - G(Z_i)} \\ & = - \iiint \frac{\varphi(s, t, \boldsymbol{u}) \mathbb{1}_{\{\tau < t, \tau < \omega\}}}{(1 - G(t))(1 - M(\tau))^2} \widehat{M}_n(d\omega) M_0(d\tau) M^{(v)}(ds, dt, d\boldsymbol{u}) \\ & + \iint \frac{\varphi(s, t, \boldsymbol{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} \widehat{M}_{0n}(d\tau) M^{(v)}(ds, dt, d\boldsymbol{u}) + R_n^{(v)}, \end{aligned} \tag{Appendix 2.4}$$

where $|R_n^{(v)}| = O(n^{-1} \ln n)$ *w.p.1.*

Proof. Using the following decomposition for $z < Z_{n:n}$ in C_{in},

$$\frac{1}{1 - \widehat{M}_n(z)} = -\frac{1 - \widehat{M}_n(z)}{(1 - M(z))^2} + \frac{2}{1 - M(z)} + \frac{\left(\widehat{M}_n(z) - M(z)\right)^2}{(1 - M(z))^2 \left(1 - \widehat{M}_n(z)\right)},$$

we can write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\varphi\left(Y_i,Z_i,\boldsymbol{U}_i\right)\delta_i J_i^{(v)}\frac{C_{in}}{1-G\left(Z_i\right)}\\
&=-\iiint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)^2}\widehat{M}_n\left(d\omega\right)\widehat{M}_{0n}\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\\
&+2\iint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)}\widehat{M}_{0n}\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\\
&-\iint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)}M_0\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)+R_{n1}^{(v)},
\end{aligned}
\tag{Appendix 2.5}
$$

where

$$
R_{n1}^{(v)}=\iint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t\}}}{\left(1-G\left(t\right)\right)}\frac{\left(\widehat{M}_n\left(t\right)-M\left(t\right)\right)^2}{\left(1-M\left(t\right)\right)^2\left(1-\widehat{M}_n\left(t\right)\right)}\widehat{M}_{0n}\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right).
$$

Under Assumption 5 and with same argument to Stute (1995, Lemma 2.5), i.e. using iterated logarithm for empirical measures and strong law of large numbers (SLLN), we obtain $|R_{n1}^{(v)}|=O(n^{-1}\ln n)$ w.p.1. For the rest of the proof, we shall decompose the other terms in the previous equation (Appendix 2.5) as a U-statistic plus a negligent remainder. Formally, we have

$$
\begin{aligned}
&\iiint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)^2}\widehat{M}_n\left(d\omega\right)\widehat{M}_{0n}\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\\
&=\iiint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t,\tau<\omega\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)^2}\\
&\qquad\times\left[\begin{array}{l}\widehat{M}_n\left(d\omega\right)M_0\left(d\tau\right)M^{(v)}\left(ds,dt,d\boldsymbol{u}\right)+M\left(d\omega\right)\widehat{M}_{0n}\left(d\tau\right)M^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\\-2M\left(d\omega\right)M_0\left(d\tau\right)M^{(v)}\left(ds,dt,d\boldsymbol{u}\right)+M\left(d\omega\right)M_0\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\end{array}\right]\\
&+R_{n2}^{(v)},
\end{aligned}
\tag{Appendix 2.6}
$$

and

$$
\begin{aligned}
&\iint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)}\widehat{M}_{0n}\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\\
&=\iint\frac{\varphi\left(s,t,\boldsymbol{u}\right)\mathbb{1}_{\{\tau<t\}}}{\left(1-G\left(t\right)\right)\left(1-M\left(\tau\right)\right)}\\
&\qquad\times\left[\widehat{M}_{0n}\left(d\tau\right)M^{(v)}\left(ds,dt,d\boldsymbol{u}\right)-M_0\left(d\tau\right)M^{(v)}\left(ds,dt,d\boldsymbol{u}\right)+M_0\left(d\tau\right)\widehat{M}_n^{(v)}\left(ds,dt,d\boldsymbol{u}\right)\right]\\
&+R_{n3}^{(v)},
\end{aligned}
\tag{Appendix 2.7}
$$

where $|R_{n2}^{(v)}| = O(n^{-1} \ln n)$ and $|R_{n3}^{(v)}| = O(n^{-1} \ln n)$ w.p.1. We refer to similar arguments of Lemmas 2.3 and 2.4 of Stute (1995) to obtain the both representations based on the Hajek projection of a V-statistic of the multivariate data $\left(Y_i, Z_i, \boldsymbol{U}_i, \delta_i, J_i^{(v)}\right)$, $1 \leq i \leq n$. Finally, the proof of (Appendix 2.4) follows by substitute (Appendix 2.6) and (Appendix 2.7) into (Appendix 2.5). □

Now, we regard the other terms in (Appendix 2.3) in the following Lemma.

Lemma 5. *Under Assumption 5, we have with w.p.1*

$$\frac{1}{n}\left|\sum_{i=1}^{n} \varphi\left(Y_i, Z_i, \boldsymbol{U}_i\right) \delta_i J_i^{(v)} \frac{B_{in}}{1 - G\left(Z_i\right)}\right| = O\left(n^{-1}\right), \tag{Appendix 2.8}$$

and

$$\frac{1}{2n}\left|\sum_{i=1}^{n} \varphi\left(Y_i, Z_i, \boldsymbol{U}_i\right) \delta_i J_i^{(v)} \exp\left\{\Delta_i\right\} \left(B_{in} + C_{in}\right)^2\right| = O\left(n^{-1} \ln n\right). \tag{Appendix 2.9}$$

Proof. The proof follows immediately with the proofs of Lemmas 2.6 and 2.7 of Stute (1995). □

Proof of the Theorem 2. With Lemmas 4 and 5, Equation (Appendix 2.3) yields

$$\widehat{S}_n^{(v)}\left(\varphi\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{\varphi\left(Y_i, Z_i, \boldsymbol{U}_i\right) \delta_i J_i^{(v)}}{1 - G\left(Z_i\right)} + \frac{1}{n}\sum_{i=1}^{n}\left[\lambda_1^{(v)}\left(Z_i\right)\left(1 - \delta_i\right) - \lambda_2^{(v)}\left(Z_i\right)\right] + R_{n4}^{(v)} \tag{Appendix 2.10}$$

where $|R_{n4}^{(v)}| = O(n^{-1} \ln n)$ w.p.1. As a consequence, we have CLT results for $\widehat{S}_n^{(v)}\left(\varphi\right)$, $v \in \mathcal{V}$, and Theorem 2 follows under Assumption 5.

Finally, the results of Theorem 2 can be extended on $]\nu, \tau_Z]$ by an argument similar to the proof of Theorem 1.1 in Stute (1995) under Assumptions 3 and 4. □

References

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**(3), 141–150.

Allignol, A., Beyersmann, J., Gerds, T., and Latouche, A. (2013). A competing risks approach for nonparametric estimation of transition probabilities in a non-markov illness-death model. *Lifetime Data Analysis*, pages 1–19.

Amorim, A. P., de Uña-Álvarez, J., and Meira-Machado, L. (2011). Presmoothing the transition probabilities in the illness–death model. *Statistics & Probability Letters*, **81**(7), 797–806.

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, **11**(2), 91–115.

Andersen, P. K., Borgan, r., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag New York Inc.

Bandeen-Roche, K. and Liang, K. Y. (2002). Modelling multivariate failure time associations in the presence of a competing risk. *Biometrika*, **89**(2), 299–314.

Bandeen-Roche, K. and Ning, J. (2008). Nonparametric estimation of bivariate failure time associations in the presence of a competing risk. *Biometrika*, **95**(1), 221–232.

Beyersmann, J., Schumacher, M., and Allignol, A. (2011). *Competing Risks and Multistate Models with R*. Use R ! Springer.

Beyersmann, J., Termini, S. D., and Pauly, M. (2013). Weak convergence of the wild bootstrap for the aalen–johansen estimator of the cumulative incidence function of a competing risk. *Scandinavian Journal of Statistics*, **40**(3), 387–402.

Cheng, Y. and Fine, J. P. (2008). Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data. *Biometrika*, **95**(1), 233–240.

Cheng, Y., Fine, J. P., and Kosorok, M. R. (2007). Nonparametric association analysis of bivariate competing-risks data. *Journal of the American Statistical Association*, **102**(480), 1407–1415.

Cheng, Y., Fine, J. P., and Bandeen-Roche, K. (2010). Association analyses of clustered competing risks data via cross hazard ratio. *Biostatistics*, **11**(1), 82–92.

Christiansen, M. C. (2012). Multistate models in health insurance. *Advances in Statistical Analysis*, **96**(2), 155–186.

Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research*, **11**(2), 167–182.

Courbage, C. and Roudaut, N. (2011). Long-term care insurance: The french example. *European Geriatric Medicine*, **2**(1), 22–25.

Czado, C. and Rudolph, F. (2002). Application of survival analysis methods to long-term care insurance. *Insurance: Mathematics and Economics*, **31**(3), 395–413.

Datta, S. and Satten, G. A. (2001). Validity of the aalen–johansen estimators of stage occupation probabilities and nelson–aalen estimators of integrated transition hazards for non-markov models. *Statistics & Probability Letters*, **55**(4), 403–411.

Datta, S. and Satten, G. A. (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-markov systems under dependent censoring. *Biometrics*, **58**(4), 792–802.

De Wreede, L. C., Fiocco, M., and Putter, H. (2011). mstate: An r package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, **38**(7).

Deléglise, M. P., Hess, C., and Nouet, S. (2009). Tarification, provisionnement et pilotage d'un portefeuille dépendance. *Bulletin Français d'Actuariat*, **9**(17), 70–108.

Denuit, M. and Robert, C. (2007). *Actuariat des assurances de Personnes - Modélisation, tarification et provisionnement*. Assurance Audit Actuariat. Economica, Paris.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**(1), 1–26.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, **76**(374), 312–319.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, **88**(4), 907–919.

Gauzère, F., Commenges, D., Barberger-Gateau, P., Letenneur, L., and Dartigues, J.-F. (1999). Maladie et dépendance - description des évolutions par des modèles multi-états. *Population*, **54**(2), 205–222.

Glidden, D. V. (2002). Robust inference for event probabilities with non-markov event data. *Biometrics*, **58**(2), 361–368.

Guibert, Q. and Planchet, F. (2014). Construction de lois d'expérience en présence d'évènements concurrents – application à l'estimation des lois d'incidence d'un contrat dépendance. *Bulletin Français d'Actuariat*, **13**(27), 5–28.

Haberman, S. and Pitacco, E. (1998). *Actuarial Models for Disability Insurance*. Chapman and Hall/CRC, 1 edition.

Helwich, M. (2008). *Durational effects and non-smooth semi-Markov models in life insurance*. Ph.d. thesis, University of Rostock.

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis*, **5**(3), 239–264.

Hougaard, P. (2001). *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. Springer-Verlag New York Inc.

Janssen, J. and Manca, R. (2007). *Semi-Markov Risk Models for Finance, Insurance and Reliability*. Springer-Verlag New York Inc.

Lakhal, L., Rivest, L.-P., and Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics*, **64**(1), 180–188.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, **26**(2–3), 423–444.

Levantesi, S. and Menzietti, M. (2012). Managing longevity and disability risks in life annuities with long term care. *Insurance: Mathematics and Economics*, **50**(3), 391–401.

Lopez, O. (2012). A generalization of kaplan-meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models. *Insurance: Mathematics and Economics*, **51**(3), 505–516.

Meira-Machado, L. and Roca-Pardinas, J. (2011). p3state.msm: Analyzing survival data from an illness-death model. *Journal of Statistical Software*, **38**(3), 1–18.

Meira-Machado, L., de Uña-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-markov illness–death model. *Lifetime Data Analysis*, **12**(3), 325–344.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., and Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, **18**(2), 195–222.

Neveu, J. (1975). *Discrete Parameter Martingales*. North-Holland / American Elsevier, Amsterdam and New-York.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**(406), 487–493.

Peng, L. and Fine, J. P. (2006). Nonparametric estimation with left-truncated semicompeting risks data. *Biometrika*, **93**(2), 367–383.

Plisson, M. (2009). *Assurabilité et développement de l'assurance dépendance*. Ph.d. thesis, Université Paris Dauphine, Paris.

Pritchard, D. J. (2006). Modeling disability in long-term care insurance. *North American Actuarial Journal*, **10**(4), 48–75.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**(11), 2389–2430.

R Core Team (2014). R: A language and environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria.

Rodríguez-Girondo, M. and de Uña-Álvarez, J. (2012). A nonparametric test for markovianity in the illness-death model. *Statistics in Medicine*, **31**(30), 4416–4427.

Rotolo, F., Legrand, C., and Van Keilegom, I. (2013). A simulation procedure based on copulas to generate clustered multi-state survival data. *Computer Methods and Programs in Biomedicine*, **109**(3), 305–312.

Scheike, T. H. and Sun, Y. (2012). On cross-odds ratio for multivariate competing risks data. *Biostatistics*, **13**(4), 680–694.

Scheike, T. H., Sun, Y., Zhang, M.-J., and Jensen, T. K. (2010). A semiparametric random effects model for multivariate competing risks data. *Biometrika*, **97**(1), 133–145.

Sellero, C. S., Manteiga, W. G., and Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scandinavian Journal of Statistics*, **32**(4), 563–581.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, **45**(1), 89–103.

Stute, W. (1994). The bias of kaplan-meier integrals. *Scandinavian Journal of Statistics*, **21**(4), 475–484.

Stute, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics*, **23**(2), 422–439.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian journal of statistics*, **23**(4), 461–471.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics*, **21**(3), 1591–1607.

Suzukawa, A. (2002). Asymptotic properties of aalen-johansen integrals for competing risks data. *Journal of the Japan Statistical Society*, **32**, 77–93.

Tomas, J. and Planchet, F. (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: Application to long-term care insurance. *Insurance: Mathematics and Economics*, **52**(3), 573–589.

van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag New York Inc., New York.

Varadhan, R., Xue, Q.-L., and Bandeen-Roche, K. (2014). Semicompeting risks in aging research: methods, issues and needs. *Lifetime Data Analysis*, pages 1–25.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, **66**(3), 716–725.

Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**(1), 127–138.

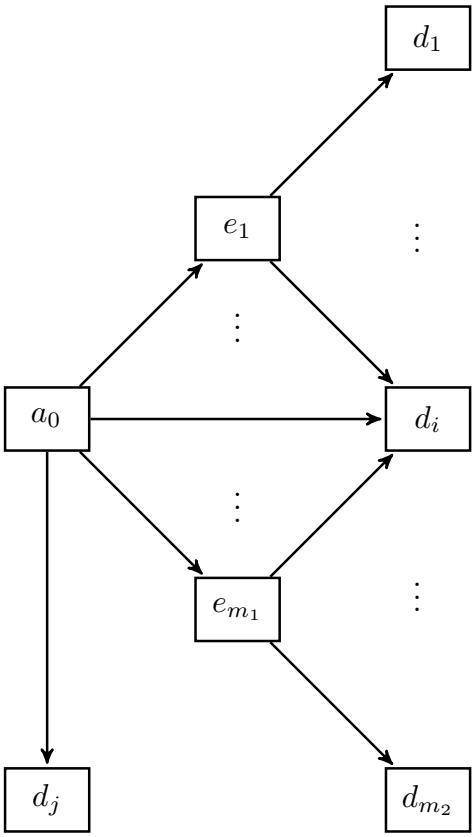


Figure 1: Exemple of acyclic multi-state model with 2-level of competing risks.

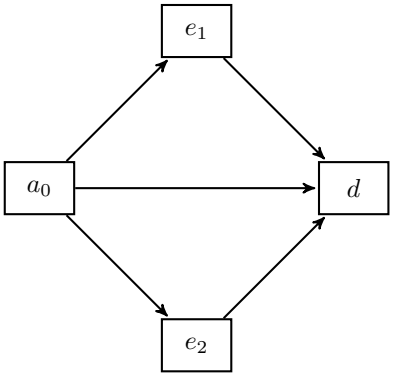


Figure 2: Multi-state structure for the simulation study.

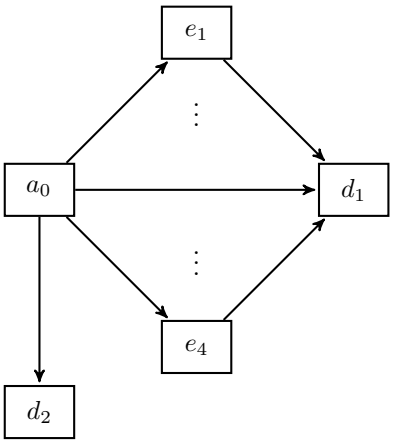


Figure 3: Multi-state structure of LTC data.

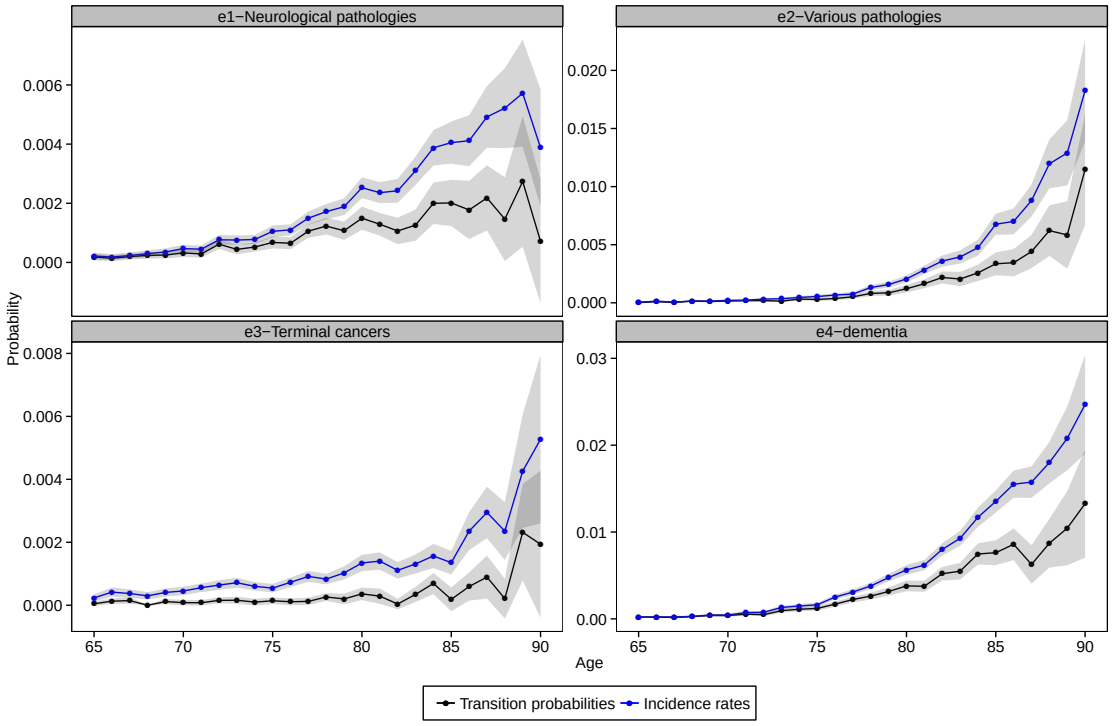
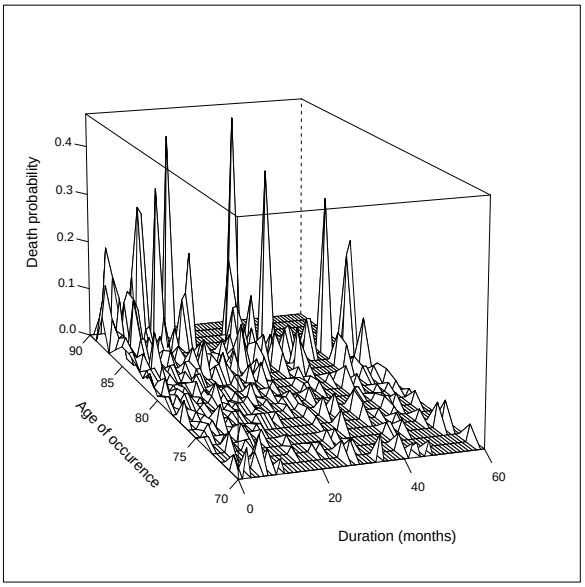
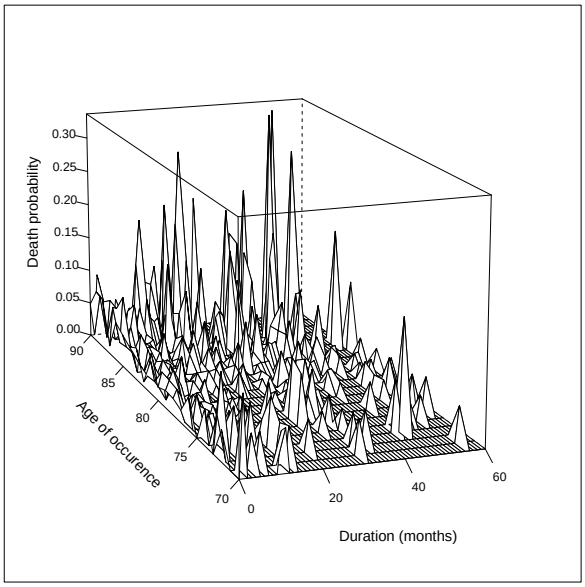


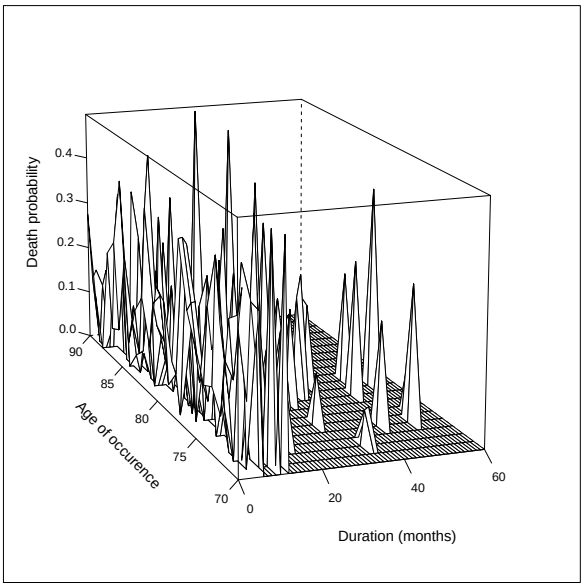
Figure 4: Estimated transition probabilities (in black) from the healthy state a_0 to the dependency states $e_1, \ldots, e_4$ and the related incidence rates (in blue). The corresponding pointwise 95% confidence interval are obtained from 500 bootstrap samples.



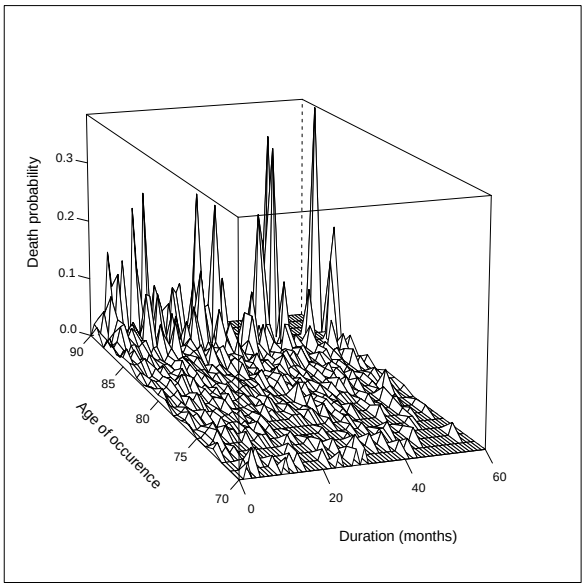
(a) e_1-Neurologic pathologies.



(b) e_2-Various pathologies.



(c) e_3-Terminal cancers.



(d) e_4-Dementia.

Figure 5: Surfaces of estimated death rates for (a) Neurologic pathologies, (b) Various pathologies, (c) Terminal cancers and (d) Dementia entry causes.

Table 1: Simulation parameters for the both censoring scenarios.

Parameter	T_{0e_1}	T_{0e_2}	T_{0d}	C_0	T_{e_1d}	C_1	T_{e_2d}	C_2
Moderate censoring scenario								
Scale λ	25	15	40	35	2.5	10	5	15
Shape ρ	0.9	0.8	1.1	1	0.5	1	0.8	1
Medium censoring scenario								
Scale λ	25	15	40	27	2.5	3	5	7
Shape ρ	0.9	0.8	1.1	1	0.5	1	0.8	1

Note: This table displays the simulation parameters used for the moderate and the medium censoring scenarios. Our simulation approach uses Weibull distributions $\mathrm{Wei}(\lambda, \rho)$ with the respective scale λ and shape ρ parameters.

Table 2: Performance analysis for the moderate censoring scenario.

			$p_{0e_1}(s,t,0)$			$p_{e_1e_1}(s,t)$			$p_{0e_2}(s,t,0)$			$p_{e_2e_2}(s,t)$		
n	s	Estimator	BIAS	VAR	MSE	BIAS	VAR	MSE	BIAS	VAR	MSE	BIAS	VAR	MSE
200	5	NP	0.2677	0.0292	0.0314	0.6066	0.7473	0.7559	0.1974	0.0708	0.0719	1.0558	0.4734	0.4986
	5	AJ	0.2107	0.0238	0.0250	12.5020	0.1686	4.7780	0.2025	0.0545	0.0557	13.0388	0.0823	4.6581
	10	NP	0.2400	0.0479	0.0495	1.6947	1.4419	1.5208	0.3353	0.1322	0.1353	2.4179	0.9449	1.0810
	10	AJ	0.4173	0.0562	0.0600	13.3934	0.4767	5.4007	0.3671	0.1225	0.1263	14.6820	0.1721	5.6497
	15	NP	0.2955	0.0846	0.0871	2.2481	2.7788	2.9194	0.4811	0.2538	0.2597	3.1900	1.8140	2.0565
	15	AJ	0.6290	0.1150	0.1240	13.0411	1.5824	6.0611	0.5685	0.2548	0.2649	15.6106	0.4091	6.3747
400	5	NP	0.1410	0.0150	0.0155	0.4132	0.3997	0.4036	0.0971	0.0391	0.0393	0.7755	0.2467	0.2595
	5	AJ	0.1157	0.0103	0.0110	12.6670	0.0759	4.8312	0.1268	0.0270	0.0275	13.0637	0.0383	4.6253
	10	NP	0.1769	0.0242	0.0250	1.7003	0.7829	0.8509	0.4501	0.0695	0.0747	2.3527	0.5160	0.6397
	10	AJ	0.1962	0.0244	0.0255	13.7911	0.1779	5.3736	0.2102	0.0595	0.0607	14.7317	0.0820	5.5742
	15	NP	0.2338	0.0401	0.0415	2.5634	1.3643	1.5265	0.6775	0.1209	0.1332	3.3957	0.9095	1.1744
	15	AJ	0.3331	0.0510	0.0537	14.5204	0.4781	5.9202	0.3074	0.1201	0.1231	15.8460	0.1699	6.2659
800	5	NP	0.0771	0.0082	0.0084	0.2412	0.2174	0.2188	0.1602	0.0202	0.0208	0.9523	0.1337	0.1536
	5	AJ	0.1250	0.0050	0.0057	12.7165	0.0378	4.8383	0.0907	0.0125	0.0128	13.1200	0.0181	4.6491
	10	NP	0.2097	0.0132	0.0144	1.6509	0.4465	0.5060	0.7095	0.0335	0.0452	2.5794	0.2817	0.4321
	10	AJ	0.1513	0.0114	0.0123	13.9589	0.0839	5.4243	0.1214	0.0273	0.0278	14.7700	0.0375	5.5601
	15	NP	0.3212	0.0207	0.0234	2.6677	0.7873	0.9486	1.0721	0.0545	0.0809	3.8228	0.5062	0.8399
	15	AJ	0.2075	0.0235	0.0249	14.9268	0.1820	5.9346	0.1559	0.0551	0.0558	15.9453	0.0779	6.2465

Note: This table contains the estimate of integrated absolute bias (BIAS), integrated variance (VAR), integrated mean square error (MSE) with our non-parametric estimators (NP) and the so-called Aalen-Johansen estimators (AJ). We compare the results at time $s = 5$, $s = 10$ and $s = 15$ for samples with size $n = 200$, $n = 400$ and $n = 800$. The results are obtained with $K = 1,000$ simulations.

Table 3: Performance analysis for the medium censoring scenario.

			$p_{0e_1}(s,t,0)$			$p_{e_1e_1}(s,t)$			$p_{0e_2}(s,t,0)$			$p_{e_2e_2}(s,t)$		
n	s	Estimator	BIAS	VAR	MSE	BIAS	VAR	MSE	BIAS	VAR	MSE	BIAS	VAR	MSE
200	5	NP	0.9159	0.0668	0.0887	1.914	1.2698	1.3744	0.4886	0.1229	0.1305	0.8676	0.7658	0.7836
	5	AJ	0.4688	0.0451	0.0521	12.9462	0.2773	5.3693	0.5384	0.0979	0.1073	13.0583	0.1052	4.7245
	10	NP	0.7760	0.1135	0.1308	2.1609	2.5397	2.6578	0.6324	0.2168	0.2267	2.9612	1.5380	1.7775
	10	AJ	0.9717	0.1197	0.1462	13.3084	1.1390	6.127	1.0419	0.2340	0.2678	14.3679	0.3248	5.6281
	15	NP	0.9537	0.2150	0.2421	3.1541	4.9698	5.2163	0.8754	0.3978	0.4167	3.9565	3.0569	3.5118
	15	AJ	1.6298	0.2595	0.3284	9.9289	4.5391	7.5448	1.7081	0.4857	0.5766	14.572	1.1424	6.5033
400	5	NP	0.6869	0.0380	0.0503	1.4312	0.7400	0.7953	0.2841	0.0675	0.0699	0.9746	0.4318	0.4546
	5	AJ	0.2855	0.0177	0.0197	13.192	0.0991	5.4109	0.2573	0.0446	0.0472	13.1923	0.0500	4.7607
	10	NP	0.4185	0.0576	0.0624	1.6965	1.4035	1.4806	0.6878	0.1133	0.1274	3.6053	0.8671	1.1742
	10	AJ	0.4284	0.0484	0.0547	14.6438	0.2995	6.3756	0.5511	0.1086	0.1197	14.7660	0.1252	5.6891
	15	NP	0.4635	0.0977	0.1037	2.8219	2.4924	2.7194	0.9460	0.1983	0.2250	5.0151	1.5387	2.1608
	15	AJ	0.8309	0.1140	0.1350	14.7602	1.0902	6.8767	0.9716	0.2321	0.2663	15.6127	0.35102	6.3629
800	5	NP	0.5821	0.0214	0.0296	1.2003	0.4007	0.4364	0.1554	0.0364	0.0371	0.8650	0.2357	0.2512
	5	AJ	0.2491	0.0075	0.0092	13.2184	0.0460	5.3865	0.1506	0.0213	0.0220	13.1564	0.0256	4.6949
	10	NP	0.2420	0.0275	0.0291	2.1227	0.7551	0.8598	0.9525	0.0519	0.0747	3.7492	0.4470	0.7607
	10	AJ	0.2786	0.0202	0.0223	14.7628	0.1348	6.3297	0.2993	0.0516	0.0545	14.8136	0.0595	5.6305
	15	NP	0.3710	0.0410	0.0449	4.0669	1.3465	1.7387	1.3976	0.0854	0.1343	5.6074	0.7752	1.4943
	15	AJ	0.4045	0.0485	0.0542	15.8200	0.3886	7.0898	0.5068	0.1112	0.1204	15.8090	0.1446	6.2374

Note: This table contains the estimate of integrated absolute bias (BIAS), integrated variance (VAR), integrated mean square error (MSE) with our non-parametric estimators (NP) and the so-called Aalen-Johansen estimators (AJ). We compare the results at time $s = 5$, $s = 10$ and $s = 15$ for samples with size $n = 200$, $n = 400$ and $n = 800$. The results are obtained with $K = 1,000$ simulations.

Table 4: Estimated association measure.

s	1 month	6 months	12 months	18 months	24 months
			e_1-*Neurologic pathologies*		
70	3.302	3.253	3.201	3.138	3.06
	(3.034, 3.569)	(2.98, 3.525)	(2.921, 3.481)	(2.853, 3.423)	(2.768, 3.352)
75	2.604	2.543	2.48	2.403	2.329
	(2.417, 2.79)	(2.352, 2.733)	(2.287, 2.672)	(2.209, 2.596)	(2.13, 2.528)
80	1.892	1.844	1.761	1.679	1.61
	(1.725, 2.059)	(1.676, 2.013)	(1.586, 1.935)	(1.501, 1.857)	(1.429, 1.791)
85	1.373	1.348	1.277	1.223	1.182
	(1.18, 1.567)	(1.151, 1.544)	(1.07, 1.484)	(1.007, 1.439)	(0.961, 1.404)
90	0.959	0.925	0.944	0.958	0.889
	(0.67, 1.248)	(0.617, 1.234)	(0.627, 1.26)	(0.575, 1.34)	(0.45, 1.328)
			e_2-*Various pathologies*		
70	5.15	5.124	5.06	5.044	4.974
	(4.706, 5.594)	(4.686, 5.562)	(4.612, 5.509)	(4.589, 5.498)	(4.52, 5.428)
75	4.184	4.09	4.012	3.898	3.841
	(3.939, 4.429)	(3.835, 4.345)	(3.752, 4.273)	(3.635, 4.161)	(3.574, 4.107)
80	3.27	3.229	3.135	3.007	2.914
	(3.066, 3.474)	(3.023, 3.436)	(2.92, 3.349)	(2.787, 3.226)	(2.694, 3.134)
85	2.263	2.217	2.169	2.149	2.067
	(2.057, 2.47)	(2.002, 2.431)	(1.944, 2.394)	(1.915, 2.383)	(1.823, 2.311)
90	1.621	1.661	1.694	1.736	1.706
	(1.333, 1.909)	(1.343, 1.979)	(1.339, 2.048)	(1.328, 2.144)	(1.24, 2.172)
			e_3-*Terminal cancers*		
70	3.846	3.802	3.761	3.692	3.637
	(3.436, 4.257)	(3.384, 4.221)	(3.332, 4.189)	(3.266, 4.119)	(3.212, 4.062)
75	3.366	3.388	3.353	3.303	3.254
	(3.065, 3.666)	(3.073, 3.703)	(3.035, 3.671)	(2.982, 3.625)	(2.931, 3.576)
80	3.004	3.036	3.026	2.984	2.915
	(2.708, 3.3)	(2.722, 3.35)	(2.701, 3.352)	(2.642, 3.325)	(2.566, 3.264)
85	2.371	2.39	2.39	2.366	2.298
	(2.059, 2.683)	(2.055, 2.725)	(2.029, 2.751)	(1.977, 2.755)	(1.894, 2.701)
90	1.652	–	–	–	–
	(1.026, 2.277)	–	–	–	–
			e_4-*Dementia*		
70	4.45	4.469	4.403	4.282	4.174
	(4.226, 4.673)	(4.241, 4.698)	(4.165, 4.641)	(4.027, 4.537)	(3.907, 4.442)
75	3.396	3.333	3.258	3.115	2.994
	(3.27, 3.522)	(3.199, 3.467)	(3.117, 3.4)	(2.969, 3.26)	(2.845, 3.142)
80	2.333	2.272	2.193	2.102	2.007
	(2.233, 2.433)	(2.168, 2.376)	(2.082, 2.304)	(1.989, 2.214)	(1.889, 2.125)
85	1.487	1.434	1.379	1.313	1.234
	(1.38, 1.594)	(1.323, 1.545)	(1.264, 1.493)	(1.192, 1.434)	(1.107, 1.36)
90	0.885	0.876	0.826	0.775	0.751
	(0.737, 1.033)	(0.718, 1.034)	(0.651, 1.002)	(0.583, 0.968)	(0.53, 0.973)

This tables gives the estimated association measure for each dependency states $e_1, \ldots, e_4$ with the 95% confidence interval in parentheses computed with 500 bootstrap replications. The results are calculated at age $s = 70, 75, 80, 85, 90$ and with duration $t = 1, 6, 12, 24$ months.