



HAL
open science

Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Acyclic Multi-State Models: Application to LTC Insurance

Quentin Guibert, Frédéric Planchet

► **To cite this version:**

Quentin Guibert, Frédéric Planchet. Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Acyclic Multi-State Models: Application to LTC Insurance. 2015. hal-01183542v1

HAL Id: hal-01183542

<https://hal.science/hal-01183542v1>

Preprint submitted on 9 Aug 2015 (v1), last revised 11 Jun 2018 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Acyclic Multi-State Models: 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*^{||}

July 6, 2015

Abstract. Studying Long Term Care (LTC) insurance requires to model the lifetime of individuals in presence of both terminal and non-terminal events which are concurrent. In this paper, we analyze this situation with a multi-state approach and we exhibit non-parametric estimators of transition probabilities considering the Markov assumption does not hold. 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 issue, we derive asymptotic results for this type of estimator under non-dependent random right-censorship in presence of covariates and discuss their possible outlooks. We also develop a methodology to investigate time dependence association measures between cause-specific failure times. For key transition probabilities, we conduct simulations to analyze the performance of our estimators versus the classical Aalen-Johansen estimators. Finally, we propose a numerical application with LTC insurance data, which is traditionally analyzed with semi-Markov model.

Keywords. Multi-state model, Aalen-Johansen integral, non-parametric estimator, non-Markov process, 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

Multi-state models offer a sound modeling framework for the random pattern of states experienced by an individual along time. These stochastic models are very flexible and can be adapted to many applications. In biostatistics (Hougaard, 1999, 2001; Andersen and Keiding, 2002), this specification is generally used to model the transitions between states, defined as the occurrence of a disease or a serious event affecting the survival of an individual. For credit risk and reliability areas, this framework is transposed to account for the lifetime history of a firm or an item (see e.g. Lando and Skødeberg, 2002; Janssen and Manca, 2007). For fifteen years, multi-state models have provoked a growing interest in the actuarial literature to model the random pattern of states experienced by a policyholder during the contract period. In this context, transitions between states occur when an event triggers the payment of premiums and benefits. For health and life insurance modeling purposes, many papers develop comprehensive frameworks for pricing and reserving both with Markov or semi-Markov assumptions (Haberman and Pitacco, 1998; Demuit and Robert, 2007). Recently, Christiansen (2012) gives a wide overview of the use of multi-state models in health insurance, including Long Term Care (LTC) insurance, from an academic perspective. For that purpose, actuaries need to estimate the transition probabilities between states and additionally the transition intensities if a continuous model is used. In practice, these probabilities may be adjusted to account for complex policy conditions (e.g. waiting periods and deferred period).

In this paper, we propose a non-parametric estimation framework based on a multi-state approach with longitudinal data to account for situations with both multiple terminal (e.g. multiple causes of death) and non-terminal (e.g. competing diseases or degrees of disability) events without possibility of recovery. This situation is adapted to the study of some disability, LTC and critical illness¹ insurance guarantees but we believe it may be also applicable for multimorbidity purposes in biostatistics (Varadhan *et al.*, 2014). If all failure causes can occur in any order, this corresponds to the well-known case of competing risks data. Yet, we consider in our setting that death can be caused by different failure events, i.e. the individual can previously experience a set of possible non-terminal events. When there is only one terminal event which dependently censors a non-terminal event, it is natural to refer to the framework of semi-competing risks data (Fine *et al.*, 2001). Many works in biostatistics focus on estimating the survival function of the latent failure time to a non-terminal event, using for instance the so-called copula-graphic estimator defined by Zheng and Klein (1995). However, these approaches only give a very specific information on the latent failure time and do not allow to characterize the lifetime evolution process. Alternatively, multi-state models offer better modeling properties than the classical techniques developed for semi-competing risks data where the states are grouped (Xu *et al.*, 2010).

Fitting multi-state models related to disability and LTC insurance with the available data is generally done with what could be considered as strong assumptions. To keep calculation simple, practitioners generally resort to the 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). This implies that the process ignores the effects of the previous lifetime-path. However, this assumption is inappropriate when modeling e.g. LTC claimants mortality, as the transition probabilities depend on the occurring age and the duration (or sojourn time) of each disease (Czado and Rudolph, 2002) and a semi-Markov model is more relevant. In the actuarial literature, research about fitting non-Markov models is relatively scarce and focused mainly on disability data which are generally fitted with parametric models, e.g. the so-called Poisson model (Haberman and Pitacco, 1998). Semi-parametric models are sometimes used, e.g. the Cox semi-Markov model (Czado and Rudolph, 2002). In this context, there is a need to develop non-

¹For example, where death due to dread disease should be distinguished to other causes.

parametric approaches to check how realistic these assumptions are, especially for the construction of biometric tables in insurance. With the development of Solvency II and the IFRS frameworks, this last need for realistic tables is becoming increasingly important (Tomas and Planchet, 2013; Guibert and Planchet, 2014).

Our main contribution consists in exhibiting a non-parametric framework for the development of accuracy tests² and a more realistic representation of the lifetime process. When the Markov assumption holds, non-parametric inference can be easily performed with the so-called Aalen-Johansen estimator (Aalen and Johansen, 1978; Andersen *et al.*, 1993) used to compute transition probabilities. Yet, to the best of our knowledge, there is no general non-parametric estimator for transition probabilities, in continuous time and with censoring. This field is under development but the complexity of inferring multi-state models without this assumption greatly depends on the number of states defined, the relationship between them and the nature of the available data (censorship and truncation processes). Research on non-Markov models in biostatistics mainly focuses on non-parametric estimation techniques for particular probabilities. Among them, Datta and Satten (2002) defined estimators for the state occupation probabilities and Meira-Machado *et al.* (2006) have considered 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 the Aalen-Johansen estimator can be used to infer the state occupation probabilities without the Markov assumption. An overview of the inference methods used in medical studies when the Markov assumptions does not hold is available in Meira-Machado *et al.* (2009).

Our approach extends the work of Meira-Machado *et al.* (2006). For that, we apply a model featuring two competing risks blocks which are nested to account for the progressive form of the process with right-censored data. With such a structure, our model can be viewed as a particular case of a bivariate competing risks data problem with only one censoring process. Independently from our research, Allignol *et al.* (2013) recently emphasized the interest of competing risks approaches to explain the Meira-Machado *et al.* (2006) estimators for an illness-death model. With non-parametric techniques, Cheng *et al.* (2007) have exhibited estimators for bivariate cause-specific hazards and bivariate cumulative incidence functions under independent censoring. Our framework can be linked with these last estimators but it comprises only one independent censoring variable for both competing risk processes. Motivated by applications in insurance, our main aim in this context is to develop a non-parametric methodology based on Aalen-Johansen integrals for competing risks data (Suzukawa, 2002) and investigate the consistency and weak convergence properties of such estimators in presence of independent right-censoring and covariates. This also contributes to the study of the expectation of a random function depending to two cause-specific lifetimes with right-censoring which is a new topic close to the estimation for the joint distribution of two lifetimes (see e.g. Lopez, 2012). In the statistical literature, this type of integrals with covariates generally appears to build goodness of fit tests in the context of regression analysis (see e.g. Sellero *et al.*, 2005) but in our case, we use it rather as a way to derive complex transition probabilities' properties useful for actuarial applications. Note that this approach encompasses in particular the case of classic illness-death model and progressive model.

Apart from transition probabilities, we present another application, namely the association measures among the cause-specific failure times intervening in our multi-state model. Such an application is quite natural in this context as our estimators involve a bivariate cumulative incidence function for two lifetimes intervening in the model. As there are recent advances in the literature on association measures for bivariate competing risks data (Bandein-Roche and Liang, 2002; Bandein-Roche and Ning, 2008; Cheng *et al.*, 2007, 2010; Scheike and Sun, 2012), another contribution of this paper is to exhibit non-parametric estimators for a cross-odds ratio measuring association. We

²For example, testing the performance for different competing parametric models.

think this can be used practically as a preliminary estimator characterizing the lifetime process, before implementing a parametric model to describe the joint distribution of the two competing risks processes.

This paper is organized as follows. Section 2 introduces our non-Markov multi-state model and defines the quantities used to assess transition probabilities and association measures. An acyclic multi-state model with a set of intermediary and terminal states is re-parametrized as a system of two competing risks blocks which are nested. We build upon this central idea in Section 3 to provide non-parametric estimators of the quantities of interest. We derive the convergence properties of these estimators under independent right-censoring. Section 4 is devoted to a simple simulation analysis to assess the performance of our non-parametric transition probabilities estimator against the 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 an acyclic multi-state structure with two competing risks blocks. This formulation enables to model the individual lifetime, which can be affected by both terminal and non-terminal events. In subsection 2.2, we describe the transition probabilities, used in actuarial science for pricing and reserving purposes. In 2.3, we present a version of the cross-odd ratio that accounts for the association between failures times introduced in the model.

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 $\mathcal{S} = \{a_0, e_1, \dots, e_{m_1}, d_1, \dots, d_{m_2}\}$ 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 $\{e_1, \dots, e_{m_1}\}$ represents m_1 intermediary states or non-terminal events, e.g. disability competing causes, and the set $\{d_1, \dots, d_{m_2}\}$ 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. We define, with the same notations as Rotolo *et al.* (2013) the set of states to which a direct transition from a_0 is possible: a_0 's children, noted $\mathcal{C}(a_0) \subset \mathcal{S}$. For each state $e \in \{e_1, \dots, e_{m_1}\}$, we also introduce the set of its children $\mathcal{C}(e) \subset \{d_1, \dots, d_{m_2}\}$. Of course, the set of children for one terminal event $d \in \{d_1, \dots, d_{m_2}\}$ is $\mathcal{C}(d) = \emptyset$. Hence, an individual can take two types of lifetime paths depending on whether an intermediate event occurs or not. Figure 1 depicts an example of such an acyclical multi-state structure.

[Figure 1 about here.]

This multi-state structure can be expressed with latent failure times. We denote by $T_{0e}, e \in \{e_1, \dots, e_{m_1}\}$ the latent times in healthy state corresponding to each non-terminal event and similarly $T_{0d}, d \in \{d_1, \dots, d_{m_2}\}$, for terminal events. We also set T_{ed} the latent time from disability state e to death due to cause d . No particular assumption is introduced regarding the dependence between the latent failure times. With this notation, we distinguish two steps for the overall lifetime-path including in two competing risks schemes. In a first step, the individual lifetime history can be affected by $Card(\mathcal{C}(a_0))$ -competing exit-causes from the healthy state, i.e. to the non-terminal ($a_0 \rightarrow e$) and to the terminal ($a_0 \rightarrow d$) states. 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 takes its values in $\mathcal{C}(a_0)$. We call with H the distribution function of the sojourn time in healthy state S .

In a second step, we consider that the life is exposed to m_2 -causes of exit, may they be direct ($a \rightarrow d$) or indirect through one of the m_1 intermediary states ($a \rightarrow e \rightarrow d$). This lifetime variable 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, taken is value in a set denoted \mathcal{V} , and V_2 depends on the value taken by V_1 such as

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

With this notation, we have

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

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 function³ for $t \geq 0$ and $v \in \mathcal{V}$.

As noted by [Meira-Machado *et al.* \(2006\)](#) with an acyclic illness-death model, the variables T_{ed} , representing the residual lifetime after the occurrence of a non-terminal event, are not observed in case of direct transition, i.e. where $S = T$. Otherwise, we have $S < T$. Similarly, the variables (T_{0e}) , $e \in \mathcal{C}_E(a_0) \setminus \{e'\}$, and (T_{0d}) , $d \in \mathcal{C}_D(a_0)$, are automatically censored if the individual experiences the intermediary state e' .

As is it classical for the analysis of lifetime data, we now consider a right censoring variable C with a distribution function G

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

Assumption 1 is widely used for simplicity in practice and is generally verified by insurance data as observations are censored by administrative events. It is important to note that the censoring variable is unique for all the latent times and the two lifetime variables (S, T) are not directly observed due to this right-censoring mechanism. 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 $\mathbf{U} = (U_i)_{i=1, \dots, p}$ of p -covariates. Following [Stute \(1993\)](#), we only assume⁴ that these covariates do not provide any further information as to whether censoring will take place or not, i.e. we have

Assumption 2.

³Also called cumulative incidence distribution.

⁴No assumption is made about the dependence structure between (C, S, T, \mathbf{U}, V) .

- i. $\mathbb{P}(S \leq C \mid S, \mathbf{U}, V_1) = \mathbb{P}(S \leq C \mid S, V_1)$,
- ii. $\mathbb{P}(T \leq C \mid S, T, \mathbf{U}, V) = \mathbb{P}(T \leq C \mid T, V)$.

As pointed out above, equality **ii.** of Assumption 2 is explained by the fact that the pair (S, T) is, by construction, subject to censoring and S is uncensored whenever T is. For estimation purposes, we introduce the distribution function of (S, \mathbf{U}) , noted H_0 , and $F_0^{(v)}$ the sub-distribution function of (S, T, \mathbf{U}) where the cause is $V = v$, $v \in \mathcal{V}$. Moreover, we have $F_0 = \sum_v F_0^{(v)}$.

In this paper, the theoretical results presented below can be applied both for discrete (e.g. gender, geographical location, social status) and continuous (e.g. biomedical measures) covariates. However, our practical applications presented in sub-sections 2.2 and 2.3 are limited to discrete covariates and continuous covariates should be beforehand transformed into categorical variables.

2.2 Transition probabilities

With the notation of subsection 2.1, a classical multi-state process $(X_t)_{t \geq 0}$ would be specified in terms of transition probabilities with $0 \leq s \leq t$, $i, j \in \mathcal{S}$ and \mathbf{u} a discrete covariates vectors

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

However in actuarial science, quantities of interest for risk assessment purposes relate to the probabilities of paying or receiving cash-flows which are more complex (Christiansen, 2012) to account for complex policy conditions (e.g. waiting periods and deferred period). In particular, one needs to distinguish each intermediary state since the related payment function may be different depending upon the trajectory. Let us present some situations illustrated these needs.

- i. $p_{00}(s, t \mid \mathbf{u})$, the survival probability in healthy state with $0 \leq s \leq t$,
- ii. $p_{0e}(s, t, \eta \mid \mathbf{u})$, the probability to enter in a non-terminal state due to cause $e \in \mathcal{C}_E(a_0)$ with $0 \leq s \leq t$ and a sojourn time longer than $\eta \geq 0$ at time t ,
- iii. $p_{ee}(s, t, \mid \mathbf{u})$, the probability to stay in the non-terminal event $e \in \mathcal{C}_E(a_0)$ between times s and t with $0 \leq s \leq t$,
- iv. $p_{ed}(s, t, \eta, \zeta \mid \mathbf{u})$, the probability of exit from the non-terminal event $e \in \mathcal{C}_E(a_0)$ due to cause $d \in \mathcal{C}(e)$ when the sojourn time is between η and ζ with $0 \leq \eta \leq \zeta$ and the time of entry in the non-terminal event is between s and t with $0 \leq s \leq t$,
- v. $p_{0d}(s, t \mid \mathbf{u})$, the direct transition probability from the healthy state to the terminal state $d \in \mathcal{C}_D(a_0)$ without suffering from any disease between times s and t with $0 \leq s \leq t$.

This list can be easily extended or modified considering some particular paths or if one needs to exhibit exit probabilities from a non-terminal state to some specific sets of terminal events. Anyways, these quantities can be expressed in terms of the joint distributions of (S, \mathbf{U}) and (S, T, \mathbf{U}, V) as follows

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

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

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

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

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

Without covariates, probability (2.1) can be simply estimated using the Kaplan-Meier estimator even if the Markov assumption is not verified. In Equations (2.2), (2.3) and (2.4), note that $\{V_1 = e\} = \{V_1 = e, V_2 \in \mathcal{C}(e)\}$. When looking at the numerator in Equation (2.5), we remark that considering $\mathbb{P}(T \leq t, \mathbf{U} = \mathbf{u}, V = (0, d))$ and $\mathbb{P}(S \leq t, \mathbf{U} = \mathbf{u}, V_1 = d)$ leads to different results as the former assumes competition between all the terminal states whereas the latter is equivalent to apply right-truncation on the observations which experience non-terminal events.

2.3 Association measures

Considering the general model introduced above, we try to measure association between (S, T) by studying the sub-distribution function $F_0^{(v)}$, $v \in \mathcal{V}$. This function is a particular case of the general bivariate cumulative function for a competing risks model with covariates and a 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 to measure association between several causes, whereas it can be computed non-parametrically for bivariate survival data (see e.g. Lopez, 2012). Recently, this issue has received some attention (Bandein-Roche and Liang, 2002; Bandein-Roche and Ning, 2008; Cheng and Fine, 2008; Cheng *et al.*, 2010) with an independent censoring scheme. In particular, this literature focused on estimating a cause-specific version of the localized Kendall's tau (Bandein-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 use a piecewise constant assumption on disjoint rectangular regions. Scheike *et al.* (2010) and Scheike and Sun (2012) propose alternatively local association measures based on the cross-odds ratios which have the advantage of accounting for covariates. We follow this approach and consider non-parametric estimators based on cross-odds ratio. Contrary to the standard bivariate competing risks model, there may be a strong dependence between the both indicators V_1 and V_2 due to the structure of the model.

Let the cross-odds ratio for $(e, d) \in \mathcal{V}$, $0 \leq s \leq t$ and \mathbf{u} a discrete covariates vector be given by

$$\pi_0^{(e,d)}(s, t | \mathbf{u}) = \frac{\text{odds}(T \leq t, V_2 = d | S \leq s, V_1 = e, \mathbf{U} = \mathbf{u})}{\text{odds}(T \leq t, V_2 = d | V_1 = e, \mathbf{U} = \mathbf{u})}, \quad (2.6)$$

where $\text{odds}(A) = \frac{\mathbb{P}(A)}{1 - \mathbb{P}(A)}$. This measure is based on the comparison between on the one hand the conditional odds of the event $\{T \leq t, V_2 = d\}$ for an individual exiting from the healthy state before time s due to a specific cause e , and on the other hand the odds unconditionally to the event $\{S \leq s\}$. By definition of S and T , it is not possible to reverse the role of s and t in $\pi_0^{(e,d)}(s, t | \mathbf{u})$, thus this function is defined on the upper wedge $\{0 \leq s \leq t\}$. If $\{T \leq t, V_2 = d\}$ and $\{S \leq s, V_1 = e\}$ are independent conditionally to $\{V_1 = e, \mathbf{U} = \mathbf{u}\}$, i.e. the date and the cause of entry in an illness state e have no effect on the overall lifetime ending due to cause d , then $\pi_0^{(e,d)}(s, t | \mathbf{u}) = 1$. There is positive (respectively negative) association if $\pi_0^{(e,d)}(s, t | \mathbf{u}) > (<) 1$ and this measure takes its values in $[0, \infty[$. This quantity is easy to handle and is a valuable tool to analyze duration dependence compared to the cumulative incidence function $F_0^{(v)}$ which is relatively complicated.

Alternatively, remark that [Cheng et al. \(2007\)](#) provide simple association measures without covariates defined, as the ratio of the bivariate cumulative intensity function over the associated univariate cumulative intensity functions. Note also that a test for the Markov assumption is proposed by [Rodríguez-Girondo and de Uña-Álvarez \(2012\)](#).

3 Non-parametric estimation, asymptotic properties and applications

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

3.1 General setup

The problem that we consider entails a unique right-censoring process C (Assumption 1). Thus, the observation of the i -th individual of a sample of length $n \geq 1$ is characterized by

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

which are assumed to be i.i.d. replications of the variable $(Y, \gamma, \gamma V_1, Z, \delta, \delta V_2, \mathbf{U})$. If $\delta = 1$, then obviously $\gamma = 1$. Consider first the ordered Y -values $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ and $(\gamma_{[i:n]}, \mathbf{U}_{[i:n]})$ the concomitant of the i -th order statistic (i.e. the value of $(\gamma_j, \mathbf{U}_j)_{1 \leq j \leq n}$ paired with $Y_{i:n}$). An estimator for H_0 is simply obtained from the multivariate Kaplan-Meier estimator considered by [Stute \(1993\)](#)

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

where the Kaplan-Meier weight for 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 integrals taking the form $S(\varphi) = \int \varphi dH_0$ with some generic function φ are estimated with

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

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 the information that S is necessarily observed when T is not censored. For this purpose, it is convenient to introduce $Z_{1:n} \leq Z_{2:n} \leq \dots \leq Z_{n:n}$ the ordered Z -values and $(Y_{[i:n]}, \delta_{[i:n]}, J_{[i:n]}^{(v)}, \mathbf{U}_{[i:n]})$ the concomitant of the i -th order statistic with $J_i^{(v)} = \mathbb{1}_{\{V=v\}}$ and $v \in \mathcal{V}$. Based on the idea of [Meira-Machado et al. \(2006\)](#), we consider S as

a covariate and estimate $F_0^{(v)}$ using the so-called Aalen-Johansen estimator (Aalen and Johansen, 1978) adapted to competing risk data. This estimators can be rewritten as

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

where \widetilde{W}_{in} denotes the Kaplan-Meier weight of the i -th ordered observation, related to the estimated survival function of T .

The Aalen-Johansen weights for state v , defined as

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

are very close to the Kaplan-Meier weights related to the estimated survival function of T . They can be interpreted as the mass associated to one observation. Remark than the Inverse Probability of Censoring Weighting (IPCW) representation can be easily derived from this expression by writing

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

where \widehat{G}_n represents the Kaplan-Meier estimator of the distribution function of C . Although the results presented below are theoretically true for both discrete and continuous covariates, we recommend in practice to use our estimators for transition probabilities and association measures with discrete covariates only. A possible way to use these estimators with continuous covariates is to consider regression techniques. The IPCW theory is largely used for survival models with dependent censoring and was recently applied to state occupation, exit and waiting times probabilities for acyclic multi-state models (Mostajabi and Datta, 2013) and to transition probabilities for the illness-death model (Meira-Machado *et al.*, 2014). The extension of our estimators with these techniques would be clearly feasible but is out of the scope of this paper.

Based on the representation as a sum of (3.2), we are now interested in obtained estimators of general quantities $S^{(v)}(\varphi) = \int \varphi dF_0^{(v)}$ with φ a generic function. In absence of censoring process, non-parametric estimation is straightforward, resulting to integrals under the empirical multivariate distribution function of (S, T, \mathbf{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

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

These estimators are easily to handle and are similar to those exhibited by Suzukawa (2002) but we refine this approach by the introduction of covariates. In this context, we need to derive asymptotic properties as we consider Y in (3.3) to be an uncensored covariate. Note that these estimators admit also a representation in terms of counting processes through the Kaplan-Meier product-limit estimator (Allignol *et al.*, 2013, Equation (9)).

3.2 Asymptotic properties

Let τ_Y and τ_Z be the least upper bounds of the distribution functions of Y and Z . Under Assumptions 1 and 2, the consistency and weak convergence of estimator (3.1) on $[0, \tau_Y]$ can be easily demonstrated since Stute (1993) and Stute (1996) conditions are satisfied. Note that this result is also verified if H and G have no jump in common which is a less restrictive condition than continuity. Thus, for the rest of this subsection we focus on the consistency and weak convergence properties of estimator (3.3).

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

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

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

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

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 result. These conditions are given below.

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

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

We introduce

$$\begin{aligned} M(z) &= \mathbb{P}(Z \leq z), \quad M_0(z) = \mathbb{P}(Z \leq z, \delta = 0), \\ M^{(v)}(y, z, \mathbf{u}) &= \mathbb{P}(Y \leq y, Z \leq z, \mathbf{U} \leq \mathbf{u}, \delta = 1, V = v), \end{aligned}$$

and

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

We also introduce the functions

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

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 these assumptions. Assumption 3 corresponds to a variance type assumption on φ , guaranteeing the existence of finite second moments. The second condition comes from Stute (1995) and guarantees the distribution convergence results on $[0, \tau_Z]$. Notice that this last assumption is relatively weak.

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

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

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

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

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

The proof of this theorem is postponed in B. From the Equation (3.5), we could 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 in the expression of $\boldsymbol{\Sigma}(\varphi)$ by their empirical counterparts. However, this calculation may be laborious due to the expression of $\mathbf{a}(\varphi)$. Thus, implementing a non-parametric bootstrap procedure is the most appropriate way to obtain asymptotic variance-covariance estimators.

3.3 Possible improvement for handling left-truncation

In this subsection, we discuss a possible outlook to handle left-truncated data but the formal proof of the asymptotic results is not discussed in this paper.

So far, for the sake of simplicity, we ignored the case of a left-truncation process L . For LTC insurance data, which is studied in Section 5, left-truncation is not really an issue as it always occurs when the individual is still in the healthy state. Hence, Y and Z are observed if $Y \geq L$. By revisiting Assumption 1 such that (C, L) are independent from (S, T, V) , the issue can be addressed by removing the data for which Y_i is smaller than L_i , $1 \leq i \leq n$ (Andersen *et al.*, 1993, Chapter III.3). This would lead to a small bias in our application.

However, more complicated situations may arise in a more general framework if the left-truncation events can occur after S . Indeed, simply removing truncated observations may induce considerable loss of information for the estimation of probability $\mathbb{P}(S \leq s, V_1 = e)$ with $s \geq 0$ and $e \in \mathcal{C}(a_0)$ (Peng and Fine, 2006). In such a 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 Sello *et al.* (2005) and assuming particularly that (C, L) is independent of (S, T, V) and C is independent of L . For that, one could adapt the Aalen-Johansen weights such that

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

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 not necessarily satisfied. In case the process is Markovian, transition probabilities can be estimated non-parametrically with the so-called Aalen-Johansen estimator (Aalen and Johansen, 1978; Andersen *et al.*, 1993). However, this methodology fails when the Markov assumption is wrong, especially when the transition probabilities depend on both time and duration (semi-Markov models).

To address this issue, we express (2.1), (2.2), (2.3), (2.4) and (2.5) as integrals of the form $\int \varphi dH_0$ and $\int \varphi dF_0^{(v)}$

$$(2.1) = \frac{\mathbb{E} \left[\varphi_{t,\mathbf{u}}^{(0)}(S, \mathbf{U}) \right]}{\mathbb{E} \left[\varphi_{s,\mathbf{u}}^{(0)}(S, \mathbf{U}) \right]}, \quad (3.6)$$

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

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

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

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

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

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, \mathbf{U}) and estimating these expectations with Kaplan-Meier integrals and Aalen-Johansen integrals. Therefore, natural estimators for these quantities are

$$\hat{p}_{00}(s, t | \mathbf{u}) = \frac{\hat{S}_n \left(\varphi_{t,\mathbf{u}}^{(0)} \right)}{\hat{S}_n \left(\varphi_{s,\mathbf{u}}^{(0)} \right)}, \quad (3.11)$$

$$\hat{p}_{0e}(s, t, \eta | \mathbf{u}) = \frac{\hat{S}_n^{(e,\mathcal{C}_e)} \left(\varphi_{s,t,\eta,\mathbf{u}}^{(1)} \right)}{\hat{S}_n \left(\varphi_{s,\mathbf{u}}^{(0)} \right)}, \quad (3.12)$$

$$\hat{p}_{ee}(s, t | \mathbf{u}) = \frac{\hat{S}_n^{(e,\mathcal{C}_e)} \left(\varphi_{s,t,\mathbf{u}}^{(2)} \right)}{\hat{S}_n^{(e,\mathcal{C}_e)} \left(\varphi_{s,s,\mathbf{u}}^{(2)} \right)}, \quad (3.13)$$

$$\hat{p}_{ed}(s, t, \eta, \zeta | \mathbf{u}) = \frac{\hat{S}_n^{(e,d)} \left(\varphi_{s,t,\eta,\zeta,\mathbf{u}}^{(3)} \right)}{\hat{S}_n^{(e,\mathcal{C}_e)} \left(\varphi_{s,t,\eta,\mathbf{u}}^{(4)} \right)}, \quad (3.14)$$

$$\hat{p}_{0d}(s, t | \mathbf{u}) = \frac{\hat{S}_n^{(0,d)} \left(\varphi_{s,t,\mathbf{u}}^{(5)} \right)}{\hat{S}_n \left(\varphi_{s,\mathbf{u}}^{(0)} \right)}, \quad (3.15)$$

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 $\widehat{p}_{00}(s, t | \mathbf{u})$ can be proved easily with 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 that of C . Further, we deduce weak convergence properties for these estimators as functions $\varphi_{s,t,u,\eta}^{(1)}$, $\varphi_{s,t,u}^{(2)}$, $\varphi_{s,t,\eta,\zeta,u}^{(3)}$, $\varphi_{s,t,\eta,u}^{(4)}$ and $\varphi_{s,t,u}^{(5)}$ satisfy Assumptions 3 and 4. Thus, we have:*

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

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

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

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 a simple bootstrap resampling procedure (Efron, 1979). Recently, Beyersmann *et al.* (2013) provide 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 that followed by Efron. To the best of our knowledge, no further tentative has been proposed to obtain more consistent bootstrap methodologies for cumulative intensity function or other transition probabilities.

3.5 Application to the estimation of association

We are now interested in applying our results to measure nonparametrically associations between the cause-specific failure times. We exhibit the non-parametric estimator of (2.6) and give its asymptotic properties. For the sake of simplicity, we ignore here the discrete covariate vector \mathbf{U} . For $(e, d) \in \mathcal{V}$ and $0 \leq s \leq t$, the cross-odds ratio is a function of

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

and

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

The former admits the following non-parametric estimator

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

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

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

The latter can be estimated considering

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

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

$$\widehat{\pi}_{0n}^{(e,d)}(s, t) = \frac{\frac{\widehat{F}_{0n}^{(e,d)}(s, t)}{\widehat{H}_{0n}^{(e)}(s) - \widehat{F}_{0n}^{(e,d)}(s, t)}}{\frac{\widehat{F}_n^{(e,d)}(t)}{\widehat{H}_{0n}^{(e)}(\infty) - \widehat{F}_n^{(e,d)}(t)}}, \quad (3.16)$$

Proposition 4. *Under assumptions of Theorem 1, the time dependent association estimator (3.16) is consistent w.p.1 if the support of Z is included in that 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 satisfied for the estimator of the sub-distribution function $\widehat{F}_{0n}^{(e,d)}$. The results follow easily using the convergences properties of the Aalen-Johansen estimators (Andersen *et al.*, 1993) and the delta-method for weak convergence. \square

By asymptotic normality of (3.16), we can measure locally the association for particular times s and t . For variance calculation purposes, the bootstrap procedure is appropriate and we suggest a simple resampling procedure as discuss above. To avoid regarding only some arbitrary time points, we develop a global measure on the upper wedge $\{0 \leq s \leq t\}$ with an integrated weighted average version of the cross odds-ratio. A similar approach is followed by Cheng *et al.* (2007) for bivariate competing risks data to setup independence tests 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 a different strategy than ours, using counting processes convergence properties and proving that both the bivariate cause-specific hazard and the bivariate cumulative incidence functions are Hadamard differentiable maps on appropriate domains.

Let $\mu^{(e)} = [\mu_1^{(e)}, \mu_2^{(e)}]$ and $\mu^{(d)} = [\mu_1^{(d)}, \mu_2^{(d)}]$ be two time intervals associated with causes e and d , corresponding to the study period, and let's introduce the following test statistic

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

where

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

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

Proposition 5. *Under the 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 that $\widehat{\pi}_{0n}^{(e,d)*}$ is Hadamard differentiable at $\widehat{\pi}_{0n}^{(e,d)}$. By the results of Proposition 4 and applying the functional delta-method (van der Vaart and Wellner, 2000, Theorem 3.9.4), the weak consistency result follows. \square

The weight function may be chosen constant over the integration region but selecting an appropriate function can give more flexibility, and increase the power of the test. As remarked by Cheng *et al.* (2007), this requires a preliminary analysis to identify time points with large associations or lack of data. In addition, regions with opposite sign association should be investigated separately to avoid reducing the significance of the test.

4 Simulation results for transition probabilities

In this section, we deploy a simulation approach to assess the performance of our new estimators (3.12) and (3.13) relative to the Aalen-Johansen estimators. In the following, the computations are carried out with the software R (R Core Team, 2015). 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 that the model provided by Meira-Machado *et al.* (2006) is also implemented in R (Meira-Machado and Roca-Pardinas, 2011). Our script is available upon request from the first author.

For simulation purposes without the Markov assumption, we need to specify dependence assumption between each latent failure times. 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 bivariate distribution of gap times $(S, T - S)$. With our more general model, we consider the simulation approach set up by Rotolo *et al.* (2013) applied to an acyclic 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 a Clayton copula \mathfrak{C}_{θ_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 with notations defined in sub-section 2.1

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

As a second step, we define another Clayton copula \mathfrak{C}_{θ_e} for each non-terminal event e_1, e_2 to put together its latent times T_{0e} and its children's T_{ed} . This gives for $e \in \{e_1, e_2\}$

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

With this setting, we assume the dependent parameters θ_0 and $\theta_e, e \in \{e_1, e_2\}$, have the same value $\theta = 0.5$ for each copula model. We consider that the latent failure times $T_{0e}, T_{ed}, e \in \{e_1, e_2\}$, and T_{0d} follow Weibull distributions $\text{Wei}(\lambda, \rho)$ with λ and ρ being respectively the scale and shape parameters. To assess the performance of our estimators, we introduce three independent variables C_0, C_1 and C_2 which follow Exponential distributions $\text{Wei}(\lambda, 1)$ and right-censor failure times originated from a_0, e_1 and e_2 . We compare the results under two censoring scenarios: (1) a moderate censoring scenario and (2) a medium censoring scenario. Table 1 shows the selected values for the parameters.

[Table 1 about here.]

The proportion of censoring for the both moderate scenario is 21% for S and 34% for $T - S$. Under the medium scenario, they are respectively 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 $\hat{p}_{0e}(s, t, 0)$ and $\hat{p}_{ee}(s, t)$ to measure the performance of our methodology for some fixed values of time $s = 5, s = 10, s = 15$. We choose these estimators as they can also be computed with the Aalen-Johansen estimators. To do this, we follow [Meira-Machado *et al.* \(2006\)](#) and compute the bias, variance and mean square error as well as their integrated versions on the interval $[s, s + 50]$ ⁵. Tables 2 and 3 report the results for both censoring scenarios.

[Table 2 about here.]

[Table 3 about here.]

The simulations show for $\hat{p}_{e_1e_1}(s, t)$ and $\hat{p}_{e_2e_2}(s, t)$ that our non-parametric estimators (NP) are more relevant than the Aalen-Johansen estimators (AJ), both for the moderate and the medium censoring scenario. Tables 2 and 3 clearly illustrate that the AJ estimator is highly biased compared to our estimator and this bias seems to increase with the size of the sample. This leads to higher MSE for AJ estimators. The good performances of NP estimators in terms of bias and MSE are offset by a higher variance. Note that the bias for the NP estimator is weaker for small values of s in line with the copula specification (lower tail dependence). In average, the time spent in non-terminal state is longer for state e_1 than for state e_2 , leading to higher bias for state e_2 with the NP estimators. For the medium scenario, the situation is different between e_1 and e_2 at time $s = 5$ as individuals in state e_1 are more prone to be censored. For both the AJ and NP estimators, we remark that the bias and the variance increase with the value of s .

In a second step, we compare our estimators $\hat{p}_{0e_1}(s, t, 0)$ and $\hat{p}_{0e_2}(s, t, 0)$ to their Aalen-Johansen equivalent. These results are more complex to investigate as in such a context the bias depends on the choice of marginal distributions, the censoring scenarios, the interval $[s, s + 50]$. The choice of the Copula function also has an important effect as it defines the regions where the Markov assumption is not satisfied. Indeed, the observation of local biases⁶ leads to identifying some particular time points where the NP estimators are clearly relevant and others where these estimators are seriously

⁵We approximate the integral by a sum on time intervals of width 0.2

⁶Not shown here but available on demand.

biased. This feature is illustrated in Tables 2 and 3 and contrasts with the results observed by Meira-Machado *et al.* (2006), as their simulation process does not allow to extract this behavior. A similar contradiction is also observed by Allignol *et al.* (2013) with comparable estimators but with different simulation approach.

As the average time spent in state a_0 before moving to state e_1 is longer than e_2 , the dependance with sojourn time in e_1 has a more pronounced effect on the transition probability $p_{0e_1}(s, t, 0)$ than sojourn time in e_2 on $p_{0e_2}(s, t, 0)$. This explains why the Markovian estimator has better performance for state e_2 . For $p_{0e_1}(s, t, 0)$, the Markovian estimator is highly biased for small values of s and $t - s$ ($s \in \{5, 10\}$ and $t \leq 20$). This bias remains stable with the sample size. For this reason, NP estimators have better performances in terms of integrated bias at time $s = 5$ for e_1 but this gain decreases significantly when the censoring effect increases and is relatively low for e_2 ⁷. Conversely, NP estimators are highly biased when $t > 20$ and this bias increases with the values of s . This effect is particularly pronounced for T_{0e_2} due to the selected parameters. As we choose Clayton copula, this results was expected. With these two divergent effects, the integrated bias is sensitive to the selected period: the first effect dominates for low values of s and t but the second is stronger elsewhere. For $n = 400$ and $n = 800$, this induces a lower integrated bias for Markovian estimators applied to state e_2 and both estimators are equivalent, except for $s = 5$, for state e_1 . As for the integrated variance estimators, the simulations indicate that the NP and AJ estimators are similar.

5 Application to LTC insurance data

Long Term Care (LTC) insurance is a mix of social care 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 activities of daily life. In France, this guarantee is dedicated to elderly people who are partially or totally dependent and benefits are mainly paid as an annuity. Their amounts depend on the policyholders' lifetime-paths and possibly on their degree of dependency (see e.g. Plisson, 2009; Courbage and Roudaut, 2011). This section describes our LTC insurance dataset and discusses the results obtained with our non-parametric estimators. Finally we measure the uncertainty on the estimated transition probabilities with a non-parametric bootstrap process.

5.1 Data description

The dataset that we analyze here is drawn from the database of a French LTC insurer. It is almost the same dataset that is used to estimate the probabilities of entry into dependency in Guibert and Planchet (2014). This data is also studied by Tomas and Planchet (2013) with an adaptive non-parametric approach for smoothing the survival law of LTC claimants but without distinguishing the effect of each disease. In this last study, the authors compute the monthly death rates for the population of LTC heavy claimants and exhibit significant differences for duration times indicating that the lifetime after entry into dependency clearly depends upon the age and the time elapsed since the entry⁸.

The data is longitudinal with independent right-censoring (administrative censoring) and left-truncation. Comparing to the previous studies, we have improved the data by adding the individual cause-specific living path after entry into dependency. The period of observation stretches from the beginning of 1998 to the end of 2010 and the range of ages is 65-90. We observe around 210,000 contracts and 68% of insured are censored before entry in dependency. We refer to Guibert and

⁷We obtain better performance for e_2 when $s < 5$.

⁸In addition, we have fitted a Cox semi-Markov model as a preliminary test to assess the Markov assumption, which returns the duration time as a significant factor. This means that the Markov assumption is not satisfied.

Planchet (2014) and Tomas and Planchet (2013) for more detailed description of this dataset. The data is split into 4 groups of pathologies: a neurological pathologies e_1 (censoring rates=40%), various pathologies e_2 (censoring rates=36%), terminal cancers e_3 (censoring rates=6%) and dementia e_4 (censoring rates=45%). Considering death d_1 and cancellation d_2 as the other exit causes from the initial state, the multi-state structure of our dataset is shown in Figure 3. We do not consider any covariate in this application.

[Figure 3 about here.]

5.2 Estimation results for transition probabilities and association measures

In this section, we perform the 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 the disability state.

[Figure 4 about here.]

We also present for comparison the annual incidence rates

$$q_{0e}(s, s + 1) = \mathbb{P}(S \leq s + 1, V_1 = e \mid S > s),$$

for each disease e computed with simple Aalen-Johansen estimators for competing risks data (see Guibert and Planchet, 2014, Section 4.2). The corresponding pointwise 95% confidence intervals are obtained from 500 bootstrap resamples employing the asymptotic normality and bootstrap standard errors estimates. For each disease, transition probabilities (3.12) globally grow over time and we observe that the gap between incidence rates and transition probabilities increases rapidly after age 75. For terminal cancers, the incidence rates by age 80 is 3.7 times higher than transition probabilities due to a very high mortality of insured just after entry in dependency.

We give in Figure 5 the crude estimated 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 occurrence⁹ 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 higher than 0.5 which are clearly not reliable.

[Figure 5 about here.]

The death rates are not observed for some regions due to lack of data, in particular for duration longer than twelve months. We find that the residual lifetime (after an individual enters dependency) is very different from one the type of disease to another and this clearly means that a 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 intervals 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.

[Table 4 about here.]

For each disease, the association decreases over age s and decreases slightly with duration for a given age s . This confirms the strong effect of 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 of data.

⁹We have considered integer ages.

6 Discussion

This paper focuses on the non-parametric analysis of particular acyclic multi-state models relaxing the Markov assumption in presence of independent right-censoring. Building on a competing risks set-up distinguishing two different lifetimes, we propose estimators for the treatment of such model based on Aalen-Johansen integrals for competing risks data. The asymptotic properties can be derived by adapting the results obtained for Kaplan-Meier integrals with covariates to competing risks data. This allows to exhibit estimators for transition probabilities which can be used to check any assumption usually made in applications with observed data. Due to the bivariate structure of our model, we attempt to measure the association between the failure times for each joint causes. The measure chosen is inspired by those recently developed for bivariate competing risks data, and adapted to account for the mechanical association between failure causes.

The simulations demonstrate the relevance of our approach to estimate particular transition probabilities. 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 clearly rejected. As an illustration, we apply estimators adapted to actuarial needs on LTC insurance data where the Markov assumption is not verified. In addition, we introduce association estimators that can be used to pinpoint the region where the Markov assumption is not satisfied.

Besides, a interesting extension would be to develop methods to reduce the bias of our Aalen-Johansen integrals estimators with finite samples. The work of Amorim *et al.* (2011) is a natural direction to follow when covariates are not present. Note also that recent improvements (better performance) of these initial estimators of Meira-Machado *et al.* (2006) are given by de Uña-Álvarez and Meira-Machado (2015). Our non-parametric estimators are applicable with observed discrete covariates. However, it may be difficult to apply directly these estimators with lots of covariates due to the dimension of the problem. An outlook for future investigations consist in analysis non-parametric and semi-parametric regressions with the framework introduced in this paper. Developing a semi-parametric approach for the cross odds ratio following Scheike and Sun (2012) may be also interesting to exploit the bivariate structure of our model. Finally, there is another issue in constructing some goodness-of-fit tests which requires to compute appropriate critical values. We have considered a classic Efron's bootstrap technique (Efron, 1981) with resampling approach for our applications. Nevertheless with survival data in presence of censoring, this methodology is generally chosen since it is more consistent. This subject is left for future works.

Acknowledgements

We are grateful to Olivier Lopez, Christian Robert and anonymous referees for the valuable advice and suggestions which they provided in relation to the first draft of this paper. The authors are also very thankful to Pierre-Emmanuel Darpeix for his helpful comments. This work is partially funded by the ANR project *LoLitA : Dynamic models for human Longevity with Lifestyle Adjustments*.

A Proof of Theorem 1

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

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

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 converges towards $\mathcal{F}_\infty^{(v)} = \bigcap_{n \geq 1} \mathcal{F}_n^{(v)}$. Our strategy, following [Stute and Wang \(1993\)](#), is to demonstrate that $(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geq 0)$ is a reverse-time supermartingale and then apply 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 remain applicable by decomposing φ into positive and negative parts.

Lemma 1. *For $\varphi \geq 0$ and assuming that the distribution function of Z is continuous, $(\widehat{S}_n^{(v)}(\varphi), \mathcal{F}_n^{(v)}, n \geq 0)$ 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

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

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

$$\begin{aligned} \sum_{i=1}^{k-1} \varphi(Y_{[i:n]}, Z_{i:n}, \mathbf{U}_{[i:n]}) \widehat{F}_n^{(v)}\{Z_{i:n}\} &= \sum_{i=1}^{k-1} \varphi(Y_{[i:n+1]}, Z_{i:n+1}, \mathbf{U}_{[i:n+1]}) \widehat{F}_n^{(v)}\{Z_{i:n+1}\}, \\ \sum_{i=k}^n \varphi(Y_{[i:n]}, Z_{i:n}, \mathbf{U}_{[i:n]}) \widehat{F}_n^{(v)}\{Z_{i:n}\} &= \sum_{i=k+1}^{n+1} \varphi(Y_{[i:n+1]}, Z_{i:n+1}, \mathbf{U}_{[i:n+1]}) \widehat{F}_n^{(v)}\{Z_{i:n+1}\}, \end{aligned}$$

and

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

Hence, we obtain that

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^{n+1} \varphi(Y_{[i:n+1]}, Z_{i:n+1}, \mathbf{U}_{[i:n+1]}) \widehat{F}_n^{(v)}\{Z_{i:n+1}\}. \quad (\text{A.1})$$

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

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

and

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

Since $\varphi \geq 0$, the result follows immediately by writing the conditionnal expectation of (A.1). \square

From 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 limit \mathbb{P} -almost surely. Due to the Hewitt-Savage zero-one law, $\mathcal{F}_\infty^{(v)}$ is trivial and then

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} \mathbb{E} \left[\widehat{S}_n^{(v)}(\varphi) \right] &= \mathbb{E} \left[\sum_{i=1}^n \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n - i + 1} \prod_{j=1}^{i-1} \mathbb{E} \left[\left(\frac{n - j}{n - j + 1} \right)^{\delta_{[j:n]}} \mid Z_{j:n} \right) \right] \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n - i + 1} \prod_{j=1}^{i-1} \left(1 - \frac{m(Z_{j:n})}{n - j + 1} \right) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \frac{\widetilde{\varphi}^{(v)}(Z_{i:n})}{n} \prod_{j=1}^{i-1} \left(1 + \frac{1 - m(Z_{j:n})}{n - j} \right) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \frac{\widetilde{\varphi}^{(v)}(Z_i)}{n} \prod_{j=1}^n \left(1 + \frac{1 - m(Z_j)}{n - R_{jn}} \right)^{\mathbb{1}_{\{Z_j < Z_i\}}} \right] \\ &= \mathbb{E} \left[\widetilde{\varphi}^{(v)}(Z_1) \prod_{j=1}^n \left(1 + \frac{1 - m(Z_j)}{n - R_{jn}} \right)^{\mathbb{1}_{\{Z_j < Z_1\}}} \right]. \end{aligned} \tag{A.2}$$

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

A 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

$$\mathbb{E}[\Psi_n(z)] \uparrow \frac{1}{1-G(z)}. \quad (\text{A.3})$$

Hence, under the Assumption 2 and $\varphi \geq 0$, we obtain by applying Lemma 2, Equation (A.3) and the monotone convergence theorem that

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

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

$$S_\infty^{(v)}(\varphi) = \int \mathbb{1}_{\{t < \tau_Z\}} \varphi(s, t, \mathbf{u}) F_0^{(v)}(ds, dt, d\mathbf{u}). \quad (\text{A.4})$$

As indicated earlier for a continuous $F^{(v)}$, the desired proof follows from Lemma 1, Equation (A.4) and proposition V-3-11 of [Neveu \(1975\)](#). \square

B Proof of Theorem 2

Here, we denote

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

the empirical distribution functions of M , M_0 and $M_0^{(v)}$. Directly based on [Stute \(1995\)](#)'s 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 that [Suzukawa \(2002\)](#) also follows the same strategy.

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

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

Proof. From the same rationale used to obtain (A.2), we find

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

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). \square

The exponential term in (B.1) is expanded in Stute (1995) such as

$$\exp \{ \dots \} = \frac{1}{1 - G(Z_i)} (1 + B_{in} + C_{in}) + \frac{1}{2} \exp \{ \Delta_i \} (B_{in} + C_{in})^2, \quad (\text{B.2})$$

where

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

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

and Δ is between the 2 terms

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

Considering (B.1) and (B.2), we write

$$\begin{aligned} \widehat{S}_n^{(v)}(\varphi) &= \frac{1}{n} \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \frac{1 + B_{in} + C_{in}}{1 - G(Z_i)} \\ &\quad + \frac{1}{2n} \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \exp \{ \Delta_i \} (B_{in} + C_{in})^2. \end{aligned} \quad (\text{B.3})$$

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 that $\int \varphi^2 dF_0 < \infty$ and $\varphi(s, t, \mathbf{u}) = 0$ for $\nu < t$ where $\nu < \tau_Z$.

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

Lemma 4. Under Assumption 5, we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \frac{C_{in}}{1 - G(Z_i)} \\ &= - \iiint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t, \tau < \omega\}}}{(1 - G(t))(1 - M(\tau))^2} \widehat{M}_n(d\omega) M_0(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) \\ &\quad + \iint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} \widehat{M}_{0n}(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) + R_n^{(v)}, \end{aligned} \quad (\text{B.4})$$

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} ,

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

we can write

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \frac{C_{in}}{1 - G(Z_i)} \\ &= - \iiint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t, \tau < \omega\}}}{(1 - G(t))(1 - M(\tau))^2} \widehat{M}_n(d\omega) \widehat{M}_{0n}(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \\ &+ 2 \iint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} \widehat{M}_{0n}(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \\ &- \iint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} M_0(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) + R_{n1}^{(v)}, \end{aligned} \tag{B.5}$$

where

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

Under Assumption 5 and with same argument as 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 (B.5) as a U-statistic plus a negligible remainder. Formally, we have

$$\begin{aligned} & \iiint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t, \tau < \omega\}}}{(1 - G(t))(1 - M(\tau))^2} \widehat{M}_n(d\omega) \widehat{M}_{0n}(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \\ &= \iiint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t, \tau < \omega\}}}{(1 - G(t))(1 - M(\tau))^2} \\ &\quad \times \left[\widehat{M}_n(d\omega) M_0(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) + M(d\omega) \widehat{M}_{0n}(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) \right. \\ &\quad \left. - 2M(d\omega) M_0(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) + M(d\omega) M_0(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \right] \\ &+ R_{n2}^{(v)}, \end{aligned} \tag{B.6}$$

and

$$\begin{aligned} & \iint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} \widehat{M}_{0n}(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \\ &= \iint \frac{\varphi(s, t, \mathbf{u}) \mathbb{1}_{\{\tau < t\}}}{(1 - G(t))(1 - M(\tau))} \\ &\quad \times \left[\widehat{M}_{0n}(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) - M_0(d\tau) M^{(v)}(ds, dt, d\mathbf{u}) + M_0(d\tau) \widehat{M}_n^{(v)}(ds, dt, d\mathbf{u}) \right] \\ &+ R_{n3}^{(v)}, \end{aligned} \tag{B.7}$$

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 as for Lemmas 2.3 and 2.4 of Stute (1995) to obtain the two representations based on the Hajek projection

of a V-statistic of the multivariate data $(Y_i, Z_i, \mathbf{U}_i, \delta_i, J_i^{(v)})$, $1 \leq i \leq n$. Finally, the proof of (B.4) follows by substituting (B.6) and (B.7) into (B.5). \square

Now, we regard the other terms in (B.3) in the following Lemma.

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

$$\frac{1}{n} \left| \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \frac{B_{in}}{1 - G(Z_i)} \right| = O(n^{-1}), \quad (\text{B.8})$$

and

$$\frac{1}{2n} \left| \sum_{i=1}^n \varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)} \exp\{\Delta_i\} (B_{in} + C_{in})^2 \right| = O(n^{-1} \ln n). \quad (\text{B.9})$$

Proof. The proof follows immediately from the proofs of Lemmas 2.6 and 2.7 of Stute (1995). \square

Proof of Theorem 2. With Lemmas 4 and 5, Equation (B.3) yields

$$\widehat{S}_n^{(v)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \frac{\varphi(Y_i, Z_i, \mathbf{U}_i) \delta_i J_i^{(v)}}{1 - G(Z_i)} + \frac{1}{n} \sum_{i=1}^n \left[\lambda_1^{(v)}(Z_i) (1 - \delta_i) - \lambda_2^{(v)}(Z_i) \right] + R_{n4}^{(v)} \quad (\text{B.10})$$

where $|R_{n4}^{(v)}| = O(n^{-1} \ln n)$ w.p.1. As a consequence, we have CLT results for $\widehat{S}_n^{(v)}(\varphi)$, $v \in \mathcal{V}$, 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 that of the proof of Theorem 1.1 in Stute (1995) under Assumptions 3 and 4. \square

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.
- Bandein-Roche, K. and Liang, K. Y. (2002). Modelling multivariate failure time associations in the presence of a competing risk. *Biometrika*, **89**(2), 299–314.
- Bandein-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.
- 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 Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics*, pages n/a–n/a.
- 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.
- 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.
- 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.
- Meira-Machado, L., de Uña-Álvarez, J., and Datta, S. (2014). Nonparametric estimation of conditional transition probabilities in a non-Markov illness-death model. *Computational Statistics*, pages 1–21.
- Mostajabi, F. and Datta, S. (2013). Nonparametric regression of state occupation, entry, exit, and waiting times with multistate right-censored data. *Statistics in Medicine*, **32**(17), 3006–3019.
- 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.
- R Core Team (2015). *R: A Language and Environment 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.

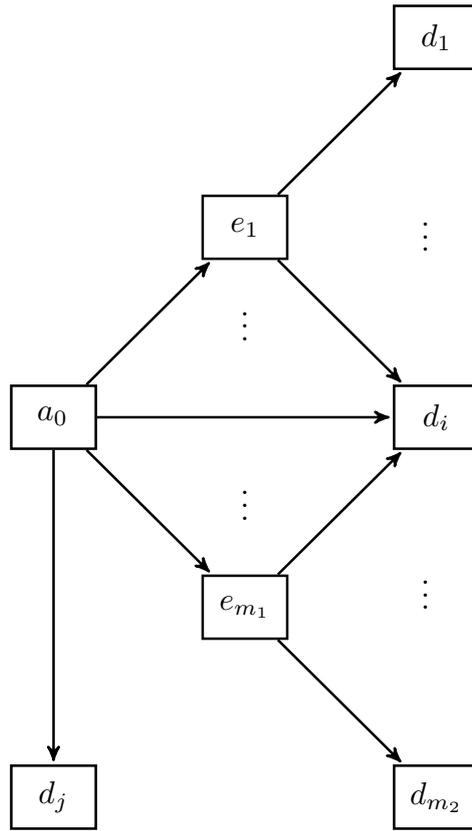


Figure 1: Example of an acyclic multi-state model with 2-level of competing risks.

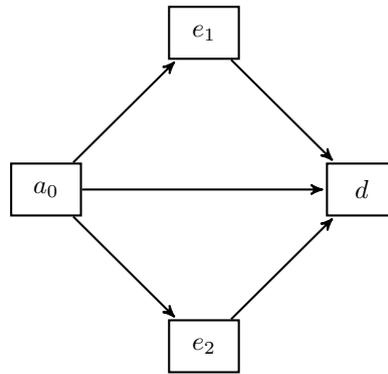


Figure 2: Multi-state structure for the simulation study.

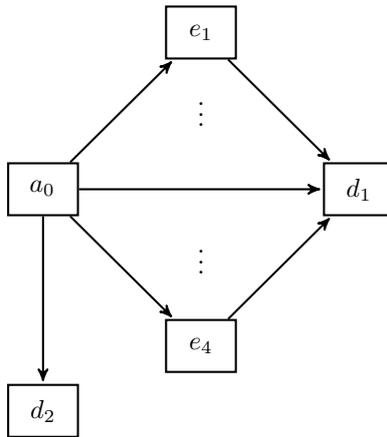


Figure 3: Multi-state structure of LTC data.

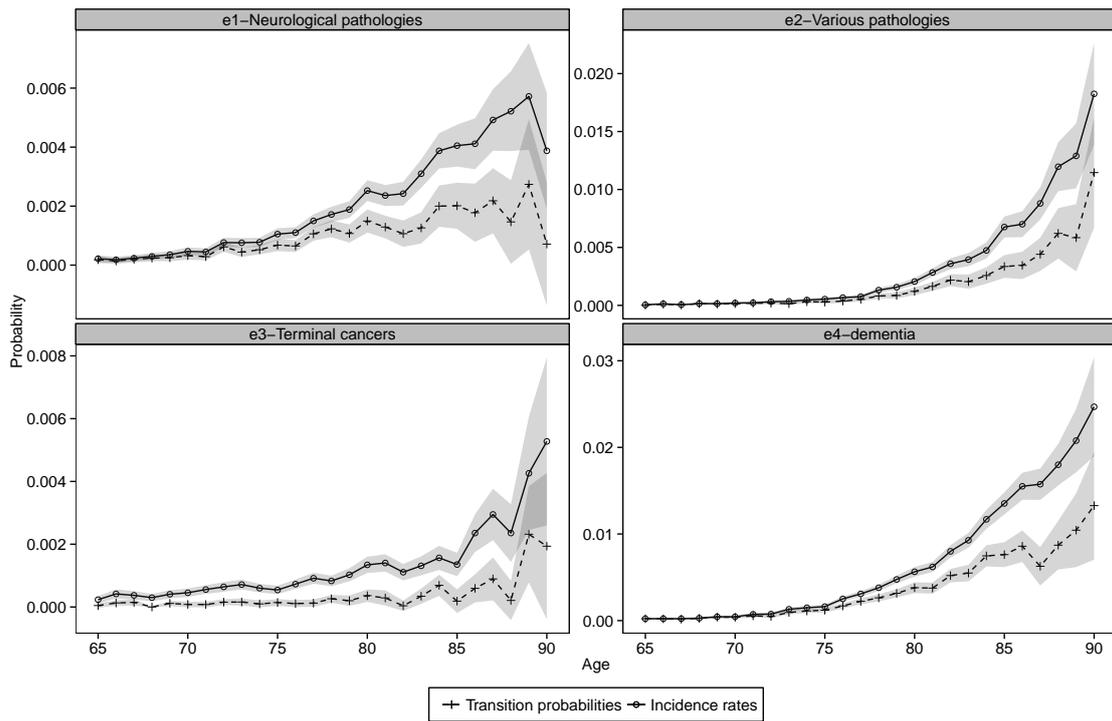
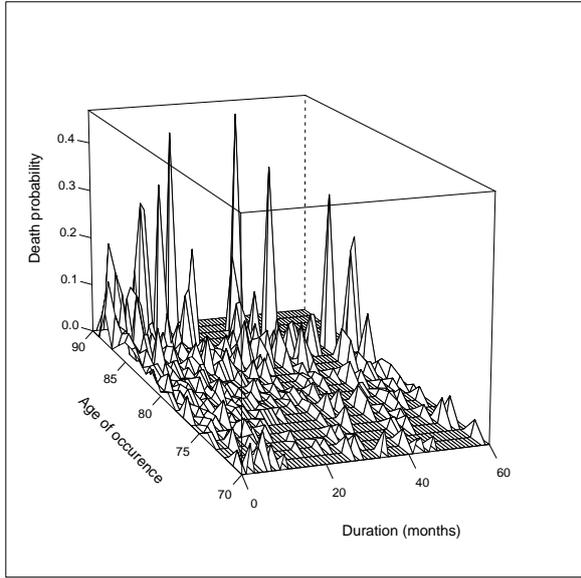
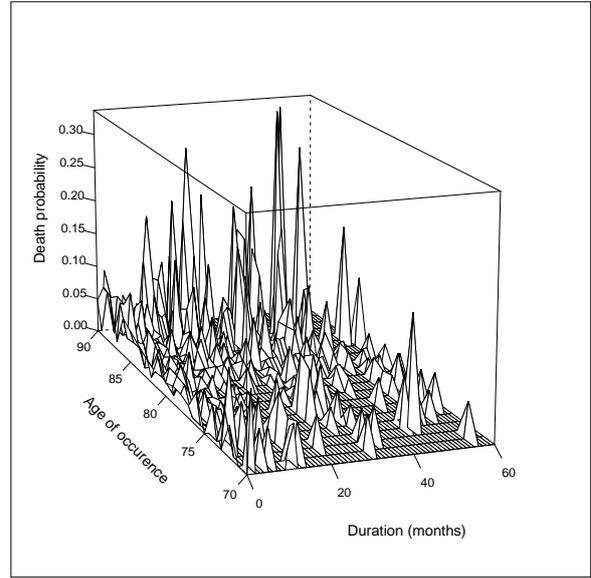


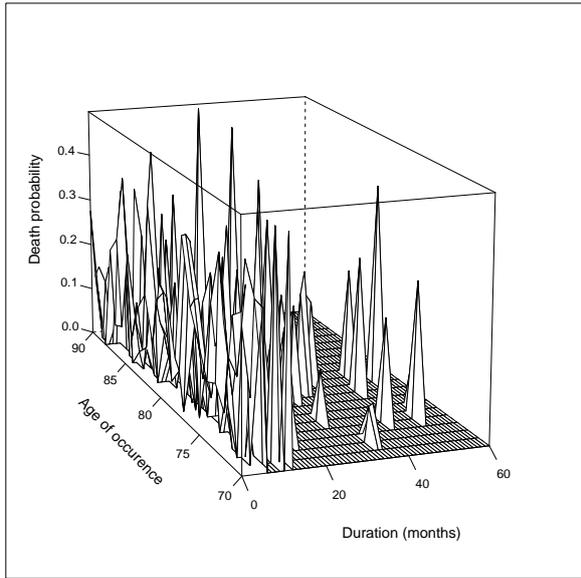
Figure 4: Estimated transition probabilities (dashed line) from the healthy state a_0 to the dependency states e_1, \dots, e_4 and the related incidence rates (solid line). The corresponding pointwise 95% confidence interval are obtained from 500 bootstrap samples.



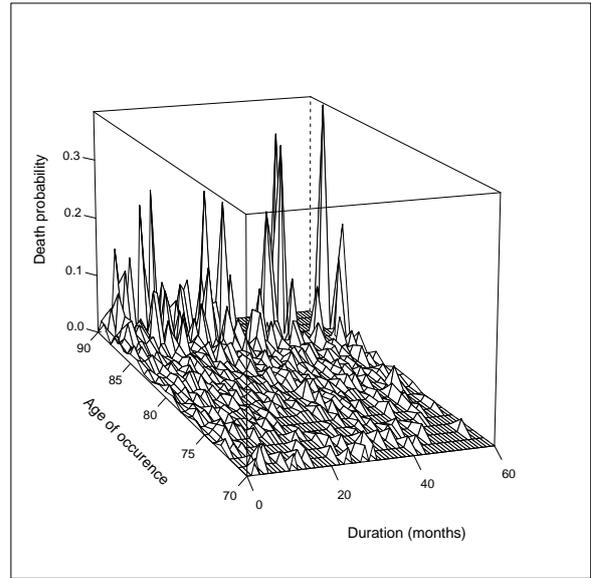
(a) e_1 -Neurologic pathologies.



(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
<i>Moderate censoring scenario</i>								
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
<i>Medium censoring scenario</i>								
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 $\text{Wei}(\lambda, \rho)$.

Table 2: Performance analysis for the moderate censoring scenario.

n	s	Estimator	$p_{0e_1}(s, t, 0)$			$p_{e_1e_1}(s, t)$			$p_{0e_2}(s, t, 0)$			$p_{e_2e_2}(s, t)$		
			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 estimates 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.

n	s	Estimator	$p_{0e_1}(s, t, 0)$			$p_{e_1e_1}(s, t)$			$p_{0e_2}(s, t, 0)$			$p_{e_2e_2}(s, t)$		
			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 estimates 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 measures.

s	1 month	6 months	12 months	18 months	24 months
<i>e₁-Neurologic pathologies</i>					
70	3.302 (3.034, 3.569)	3.253 (2.98, 3.525)	3.201 (2.921, 3.481)	3.138 (2.853, 3.423)	3.06 (2.768, 3.352)
75	2.604 (2.417, 2.79)	2.543 (2.352, 2.733)	2.48 (2.287, 2.672)	2.403 (2.209, 2.596)	2.329 (2.13, 2.528)
80	1.892 (1.725, 2.059)	1.844 (1.676, 2.013)	1.761 (1.586, 1.935)	1.679 (1.501, 1.857)	1.61 (1.429, 1.791)
85	1.373 (1.18, 1.567)	1.348 (1.151, 1.544)	1.277 (1.07, 1.484)	1.223 (1.007, 1.439)	1.182 (0.961, 1.404)
90	0.959 (0.67, 1.248)	0.925 (0.617, 1.234)	0.944 (0.627, 1.26)	0.958 (0.575, 1.34)	0.889 (0.45, 1.328)
<i>e₂-Various pathologies</i>					
70	5.15 (4.706, 5.594)	5.124 (4.686, 5.562)	5.06 (4.612, 5.509)	5.044 (4.589, 5.498)	4.974 (4.52, 5.428)
75	4.184 (3.939, 4.429)	4.09 (3.835, 4.345)	4.012 (3.752, 4.273)	3.898 (3.635, 4.161)	3.841 (3.574, 4.107)
80	3.27 (3.066, 3.474)	3.229 (3.023, 3.436)	3.135 (2.92, 3.349)	3.007 (2.787, 3.226)	2.914 (2.694, 3.134)
85	2.263 (2.057, 2.47)	2.217 (2.002, 2.431)	2.169 (1.944, 2.394)	2.149 (1.915, 2.383)	2.067 (1.823, 2.311)
90	1.621 (1.333, 1.909)	1.661 (1.343, 1.979)	1.694 (1.339, 2.048)	1.736 (1.328, 2.144)	1.706 (1.24, 2.172)
<i>e₃-Terminal cancers</i>					
70	3.846 (3.436, 4.257)	3.802 (3.384, 4.221)	3.761 (3.332, 4.189)	3.692 (3.266, 4.119)	3.637 (3.212, 4.062)
75	3.366 (3.065, 3.666)	3.388 (3.073, 3.703)	3.353 (3.035, 3.671)	3.303 (2.982, 3.625)	3.254 (2.931, 3.576)
80	3.004 (2.708, 3.3)	3.036 (2.722, 3.35)	3.026 (2.701, 3.352)	2.984 (2.642, 3.325)	2.915 (2.566, 3.264)
85	2.371 (2.059, 2.683)	2.39 (2.055, 2.725)	2.39 (2.029, 2.751)	2.366 (1.977, 2.755)	2.298 (1.894, 2.701)
90	1.652 (1.026, 2.277)	– –	– –	– –	– –
<i>e₄-Dementia</i>					
70	4.45 (4.226, 4.673)	4.469 (4.241, 4.698)	4.403 (4.165, 4.641)	4.282 (4.027, 4.537)	4.174 (3.907, 4.442)
75	3.396 (3.27, 3.522)	3.333 (3.199, 3.467)	3.258 (3.117, 3.4)	3.115 (2.969, 3.26)	2.994 (2.845, 3.142)
80	2.333 (2.233, 2.433)	2.272 (2.168, 2.376)	2.193 (2.082, 2.304)	2.102 (1.989, 2.214)	2.007 (1.889, 2.125)
85	1.487 (1.38, 1.594)	1.434 (1.323, 1.545)	1.379 (1.264, 1.493)	1.313 (1.192, 1.434)	1.234 (1.107, 1.36)
90	0.885 (0.737, 1.033)	0.876 (0.718, 1.034)	0.826 (0.651, 1.002)	0.775 (0.583, 0.968)	0.751 (0.53, 0.973)

This table gives the estimated association measures for each dependency state e_1, \dots, 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.