DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 30, ARTICLE 62, PAGES 1681–1696
PUBLISHED 28 MAY 2014
http://www.demographic-research.org/Volumes/Vol30/62/
DOI: 10.4054/DemRes.2014.30.62

Research Article

A note on computing average state occupation times

Jan Beyersmann

Hein Putter

This publication is part of the Special Collection on "Multistate Event History Analysis," organized by Guest Editors Frans Willekens and Hein Putter.

Table of Contents

Demographic Research: Volume 30, Article 62
Research Article

A note on computing average state occupation times

Jan Beyersmann [1]

Hein Putter [2]

Abstract

OBJECTIVE
This review discusses how biometricians would probably compute or estimate expected waiting times, if they had the data.

METHODS
Our framework is a time-inhomogeneous Markov multistate model, where all transition hazards are allowed to be time-varying. We assume that the cumulative transition hazards are given. That is, they are either known, as in a simulation, determined by expert guesses, or obtained via some method of statistical estimation. Our basic tool is product integration, which transforms the transition hazards into the matrix of transition probabilities. Product integration enjoys a rich mathematical theory, which has successfully been used to study probabilistic and statistical aspects of multistate models. Our emphasis will be on practical implementation of product integration, which allows us to numerically approximate the transition probabilities. Average state occupation times and other quantities of interest may then be derived from the transition probabilities.

[1] Ulm University, Germany. E-Mail: jan.beyersmann@uni-ulm.de.
[2] University of Leiden, the Netherlands. E-Mail: H.Putter@lumc.nl.

1. Introduction

This review is motivated by an interdisciplinary workshop on 'Multistate event history analysis' at the Netherlands Interdisciplinary Demographic Institute, The Hague, in April 2011. The workshop brought together demographers and biometricians, who both use multistate models, but appear to follow somewhat different methodological traditions.

At the workshop, there were discussions on how to compute average state occupation times. The organizer, Frans Willekens, informed us that the most common approaches in demography use either piecewise linear approximations of the survival function or piecewise constant transition hazards (personal communication). See also Gill and Keilman (1990) on these approaches, including a critique of the former method.

We review how biometricians might compute expected waiting times, if they had the data. In biometry, the restriction is that right-censoring typically precludes evaluating waiting time distributions on the whole of their support. In other words, the maximum follow-up in most of the data sets analyzed by biometricians is considerably smaller than the assumed maximum age in the population under study. As a consequence, it is more common to consider median waiting times (e.g., Brookmeyer and Crowley 1982) or expectations restricted to a maximum follow-up (e.g., Andersen et al. 1993, Example IV.3.8). In the context of demography, expected (restricted) waiting times are arguably more relevant, for instance for policy making.

The key idea is that a certain transformation, product integration, allows us to move from the transition hazards of a time-inhomogeneous Markov process to the matrix of transition probabilities. As also noted by Gill and Keilman (1990), combined with the initial distribution of the process, this allows us to derive expected waiting times in a given state and other quantities of interest.

In demography, product integration appears to be rarely used. Gill and Keilman (1990) mention the relation, but then proceed to attack a different problem, namely estimation of constant transition hazards with population registry data. In a recent tutorial on multistate methods, Kuo, Suchindran, and Koo (2008) value product integration as a 'basic tool', but argue that it 'is difficult to implement.' These authors therefore proceed to work under special assumptions such as uniform right-censoring. Another reference is Schoen (2005) who mentions product integration as a tool for numerical evaluation, but then concentrates on special cases with an analytical solution.

2. The relation between transition hazards and transition probabilities

Consider a time-inhomogeneous Markov process $(X_t)_{t\geq 0}$ with state space $\{0, 1, 2, \ldots, J\}$. We assume that $(X_t)_{t\geq 0}$ has right-continuous sample paths, which are constant between transition times. That is, if the process moves from state j to state k, $j \neq k$, at time t_0, $X_{t_0} = k$ and $X_{t_0-} = j$. Here X_{t_0-} refers to the state occupied by the process just before time t_0. We also assume that on any finite interval there are only finitely many transitions. No assumptions are made on irreversibility of the multistate model. The Markov assumption is

$$\mathrm{P}(X_t = k \,|\, X_s = j) = \mathrm{P}(X_t = k \,|\, X_s = j, \text{Past}), s \leq t, \tag{1}$$

where we have written 'Past' for the history generated by the process. More formally, it is a sigma-algebra and, at time s, reflects knowledge of the development of the process in the interval $[0, s]$. The Markov assumption states that the past and future of the process are independent, given the present. We will briefly discuss non-Markov processes in Section 5.

The matrix of transition probabilities is

$$\mathbf{P}(s,t) := \left(\mathrm{P}_{jk}(s,t)\right)_{j,k}, \; j,k \in \{0, 1, 2, \ldots, J\}, \tag{2}$$

with transition probabilities

$$\mathrm{P}_{jk}(s,t) := \mathrm{P}(X_t = k \,|\, X_s = j), s \leq t. \tag{3}$$

Multiplying the initial distribution $(\mathrm{P}(X_0 = 0), \mathrm{P}(X_0 = 1), \ldots, \mathrm{P}(X_0 = J))$ of the multistate process with $\mathbf{P}(0,t)$ yields the state occupation probabilities $(\mathrm{P}(X_t = 0), \mathrm{P}(X_t = 1), \ldots, \mathrm{P}(X_t = J))$ at time t. These can, e.g., be used to compute the expected time spent in state k as

$$E_k = \int_0^\infty \mathrm{P}(X_u = k) \, \mathrm{d}u, \tag{4}$$

see, e.g., Equation (6) of Gill and Keilman (1990). The restricted expected time spent in state k is

$$E_k^\tau = \int_0^\tau \mathrm{P}(X_u = k) \, \mathrm{d}u, \tag{5}$$

where τ is some fixed value, representing end of follow-up or some lower relevant time limit. In the Markov case, one can also define the expected remaining time (residual life expectancy) spent in state k, given that the subject is in state j at time s, as

$$E_{jk}^{s,\tau} = \int_s^\tau \mathrm{P}(X_u = k \,|\, X_s = j) \, \mathrm{d}u, \tag{6}$$

for $\tau \leq \infty$. For $\tau < \infty$, this is a restricted expected remaining time spent in state k. Quantities (4) – (6) may easily be evaluated numerically. Sometimes the transition probabilities may be step functions which allow us to further simplify (4) and (5). We return to this issue later.

The aim is now to find the connection between $\mathbf{P}(s,t)$ and the cumulative transition hazards (or intensities) $A_{jk}(s)$, $j \neq k$, which we assume to be given. That is, the A_{jk}'s may have been pre-specified as in a simulation, determined by expert guesses or statistically estimated. We write $A_{jj}(t) = -\sum_{k=0,k\neq j}^{J} A_{jk}(t)$ and we assume that $A_{jj}(t) - A_{jj}(t-) \geq -1$. For the discrete case, the latter assumption is implied by the fact that the probability of leaving state j at time t, on the condition that one is in state j at the last time point before t, does not exceed 1. In the purely continuous case, we have $A_{jj}(t) - A_{jj}(t-) = 0$. Our presentation below is indebted to Aalen and Johansen (1978) and Gill and Johansen (1990), see also the textbook accounts in Andersen et al. (1993) and Aalen, Borgan, and Gjessing (2008).

Consider times $s < v < t$. Using the Markov property, we get for the (j,k)-th entry of $\mathbf{P}(s,t)$ that

$$\mathrm{P}(X_t = k \,|\, X_s = j) = \sum_{\tilde{\jmath}=0}^{J} \mathrm{P}(X_v = \tilde{\jmath} \,|\, X_s = j) \cdot \mathrm{P}(X_t = k \,|\, X_v = \tilde{\jmath}). \tag{7}$$

If v is close to t, the usual interpretation of the transition hazards is that

$$\mathrm{P}(X_t = k \,|\, X_v = \tilde{\jmath}) \approx \Delta A_{\tilde{\jmath}k}(t),\; \tilde{\jmath} \neq k,$$

and consequently

$$\mathrm{P}(X_t = \tilde{\jmath} \,|\, X_v = \tilde{\jmath}) \approx 1 + \Delta A_{\tilde{\jmath}\tilde{\jmath}}(t),$$

where $\Delta A_{\tilde{\jmath}k}(t)$ is $A_{\tilde{\jmath}k}(t) - A_{\tilde{\jmath}k}(v)$. We summarize this by

$$\mathrm{P}(X_t = k \,|\, X_s = j) \approx \sum_{\tilde{\jmath}=0}^{J} \mathrm{P}(X_v = \tilde{\jmath} \,|\, X_s = j) \cdot (\mathbf{1}(\tilde{\jmath} = k) + \Delta A_{\tilde{\jmath}k}(t)), \tag{8}$$

where we have written $\mathbf{1}(\cdot)$ for the indicator function. The matrix version of (8) is

$$\mathbf{P}(s,t) \approx \mathbf{P}(s,v)\left(\mathbf{I} + \Delta\mathbf{A}(t)\right), \tag{9}$$

where we have written $\mathbf{I}$ for the $(J+1) \times (J+1)$ identity matrix and $\Delta\mathbf{A}(t)$ is the $(J+1) \times (J+1)$ matrix with (j,k) entry $\Delta A_{jk}(t)$.

Going through the approximation of (9) recursively and for a fine partition $s = t_0 < t_1 < t_2 < \ldots < t_{L-1} < t_L = t$ of the time interval $[s,t]$, we get the approximation

$$\mathbf{P}(s,t) \approx \prod_{l=1}^{L} \left(\mathbf{I} + \Delta\mathbf{A}(t_l)\right). \tag{10}$$

We note two important facts about (10): First, the approximation is precisely what we need in order to numerically compute $\mathbf{P}(s,t)$. Secondly, the right hand side of (10) returns a step function which simplifies evaluating (4). (Note that for constant or piecewise constant transition intensities, explicit matrix exponential solutions of the product integral are available (e.g. Andersen and Pohar Perme 2008); this is a well known result for homogeneous Markov processes (e.g. Kijima 1997). For practical implementation, these explicit solutions will also rely on some approximation, typically based on a Taylor expansion.)

The mathematical task is now to show that the right hand side of (10) in fact approaches $\mathbf{P}(s,t)$ for ever finer partitions of $[s,t]$. A complete account of this has been given by Gill and Johansen (1990). These authors suggest the following *product integral* notation for the limit of the finite product in (10),

$$\prod_{u\in(s,t]} \left(\mathbf{I} + \,\mathrm{d}\mathbf{A}(u)\right), \tag{11}$$

and they prove that the product integral (11) equals $\mathbf{P}(s,t)$. In summary,

$$\lim_{\max|t_l - t_{l-1}\to 0|} \prod \left(\mathbf{I} + \Delta\mathbf{A}(t_l)\right) = \prod_{u\in(s,t]} \left(\mathbf{I} + \,\mathrm{d}\mathbf{A}(u)\right) = \mathbf{P}(s,t). \tag{12}$$

3. Estimating expected waiting times

Combining relations (4)–(6) and (11) gives a relation between the transition intensities and the expected time spent in a given state. An estimate of E_k immediately presents itself from these relations, namely

$$\hat{E}_k = \int_0^\infty \hat{\mathrm{P}}(X_u = k)\,\mathrm{d}u, \tag{13}$$

where $\hat{\mathrm{P}}(X_u = k)$ would be obtained from estimates of the initial distribution $(\hat{\mathrm{P}}(X_0 = 0), \hat{\mathrm{P}}(X_0 = 1), \ldots, \hat{\mathrm{P}}(X_0 = J))$ and $\hat{\mathbf{P}}(s,t)$ through estimates $\hat{A}_{jk}(s)$ of the cumulative transition intensities. Estimates of E_k^τ and $E_{jk}^{s,\tau}$ may be obtained in a similar way.

One advantage of the estimator in (13) is that it follows the same concatenation of mappings which move from the hazards to expected waiting times as the theoretical quantities. Provided that the initial estimation 'works', this concatenation can be used to derive asymptotic unbiasedness using the continuous mapping theorem and to show correctness of the bootstrap for quantifying uncertainty using the functional delta method; we will discuss the latter at the end of Section 5. Similar 'plug in' estimates have been used by, e.g., Lièvre, Brouard, and Heathcote (2003) and Touraine, Helmer, and Joly (2013). Dropping the statistical perspective, we note that researchers have also used the trapezoidal rule or the related, but more complex Simpson's rule (also known as *Keplersche Fassregel*) for numerically approximating the integral in question. We do not consider these approaches here, but refer to Gill and Keilman (1990) who prove that using the trapezoidal rule may lead to implicit violations of the Markov assumption.

It is common for the cumulative transition intensities to be estimated as step-functions, for instance, using non-parametric methods such as the Nelson-Aalen estimate or using semi-parametric methods such as the Cox model or additive hazards. Then for a fixed state k, $\hat{\mathrm{P}}(X_u = k)$ will be constant with respect to u over time intervals. Suppose that the values of $\hat{\mathrm{P}}(X_u = k)$ are $\hat{p}_l$ on intervals $[a_{l-1}, a_l)$ for $l = 1, \ldots, L$ with $a_0 = 0$ and $a_L = \infty$. Both L, the series $a_0, \ldots, a_L$ and $\hat{p}_0, \ldots, \hat{p}_L$ may depend on the state k of interest. Then $\hat{E}_k$ can be written as

$$\hat{E}_k = \sum_{l=1}^{L} (a_l - a_{l-1})\hat{p}_l, \tag{14}$$

which is both easy and quick to calculate. If $\hat{p}_L = 0$, this sum is finite and the last element of the sum is zero (and hence could be removed). If $\hat{p}_L > 0$, then $\hat{E}_k$ will be infinite. In that case it is appropriate to consider instead the restricted expected time E_k^τ spent in state k from Equation (5). An estimate of E_k^τ would then also be given by the right-hand side of (14), the only difference with the unrestricted $\hat{E}_k$ being that here $a_L = \tau$.

We also note that estimation of (13) using non- or semi-parametric methods requires that the assumed maximum age in the population is part of the data. If this is not the case, estimation of the restricted quantity E_k^τ would be more appropriate, where τ is some fixed value, representing end of follow-up or some lower relevant time limit, as noted earlier. Alternatively, one might use parametric models such as constant or piecewise constant transition intensities, which would allow for extrapolation beyond the maximum time contained in the data.

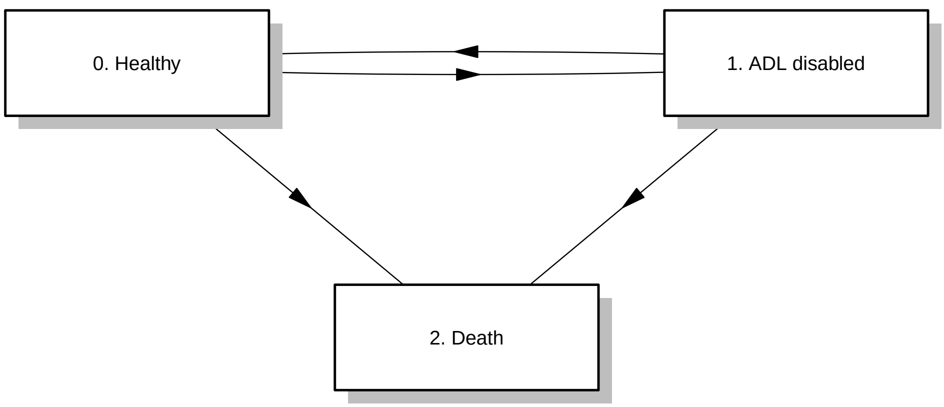
We refer to the closing Section 5 for a discussion on how to estimate uncertainty.

4. Illustration

We illustrate the methods outlined in Sections 2 and 3 using data of the Asset and Health Dynamics Among the Oldest Old (AHEAD), now part of the wider US Health and Retirement Study (HRS, Juster and Suzman 1995). The AHEAD survey includes a nationally representative sample of initially non-institutionalized persons born before 1923, aged 70 and older in 1993. The present analysis uses only the non-Hispanic white subset; the time scale is age. Subjects were interviewed every two years. For the purpose of illustration, the fact that the data are actually panel data, hence interval-censored, is conveniently ignored. Properly taking interval censoring into account will only affect the initial estimation of the transition intensities, not the subsequent calculation of the expected time spent in the different states; it is this last step that we wish to illustrate. We will further comment on interval-censoring in Section 5.

The multistate model we consider is a reversible illness-death model, illustrated in Figure 1.

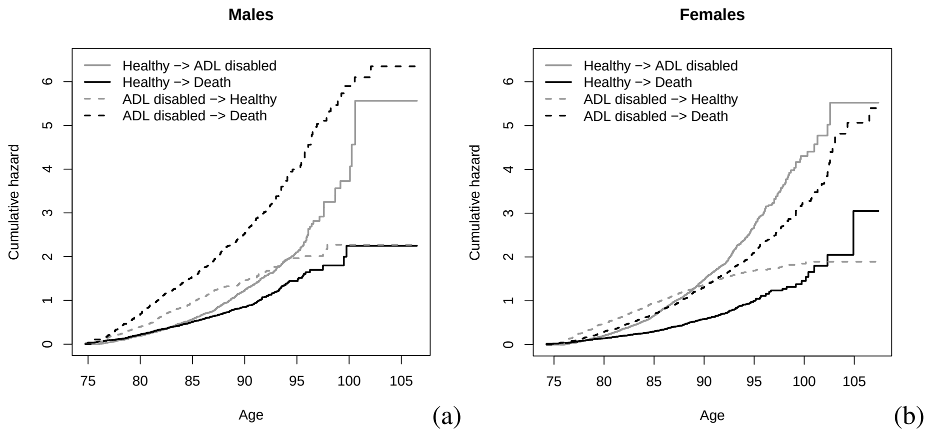
Figure 1: The reversible illness-death model of the AHEAD data



The illness state is state 1, disabled according to the Basic Activities of Daily Living (ADL) scale of Katz et al. (1963). The ADL scale consists of six items: walking across a room, bathing, dressing, getting in and out of bed, using the toilet, and eating. A subject is classified as ADL disabled when he/she responds "with difficulty" when questioned about one or more of these items. The illness-death model of Figure 1 is reversible because recovery from ADL disability is possible (there is a transition from state 1, ADL disabled, back to state 0, healthy). For a total of 4032 subjects, 1929 transitions from healthy to ADL disabled occurred during follow-up, and 679 recoveries (transitions from ADL disability to healthy). A total of 1994 deaths (state 2) were observed, 922 from the healthy state and 1072 from ADL disability. Since this is an older population, the majority of subjects (2468 and 61%) are females.

Figure 2 shows the estimated cumulative transition intensities for each of the four transitions, based on the non-parametric Nelson-Aalen estimates. They are shown for males (a) and females (b) separately.

Figure 2: Non-parametric estimates of the transition intensities in the US Health and Retirement Study, for males (a) and females (b) separately



The estimated transition intensities of disability (Healthy $\rightarrow$ ADL disabled) and recovery (ADL disabled $\rightarrow$ Healthy), in gray, are comparable between males and females. The shape of the cumulative intensity of disability is convex, indicating that with older age the disability rate increases, while the shape of the cumulative intensity of recovery is concave, indicating a decreased recovery rate with older age. Death rates, in black, are considerably higher from the ADL disability state, compared to those from the healthy state. In a Cox proportional hazards model, the hazard ratio between the ADL disabled $\rightarrow$ Death rate and the Healthy $\rightarrow$ Death rate was estimated as 2.89 (95% confidence interval (CI): 2.52 - 3.31) for males and 2.25 (95% CI: 1.99 – 2.55) for females. As expected, death rates are higher for males than for females.

Figure 3 shows estimated transition probabilities $\widehat{P}_{jk}(75, t)$ based on the transition intensities of Figure 2, again for males and females separately.

Figure 3: **Stacked plots of estimated transition probabilities $\widehat{P}_{jk}(75, t)$ for $j = 0$ ((a) and (b)), and for $j = 1$ ((c) and (d)). The lightest gray corresponds to the probability of being healthy, middle gray to the probability of being ADL disabled and the darkest gray to the probability of having died**

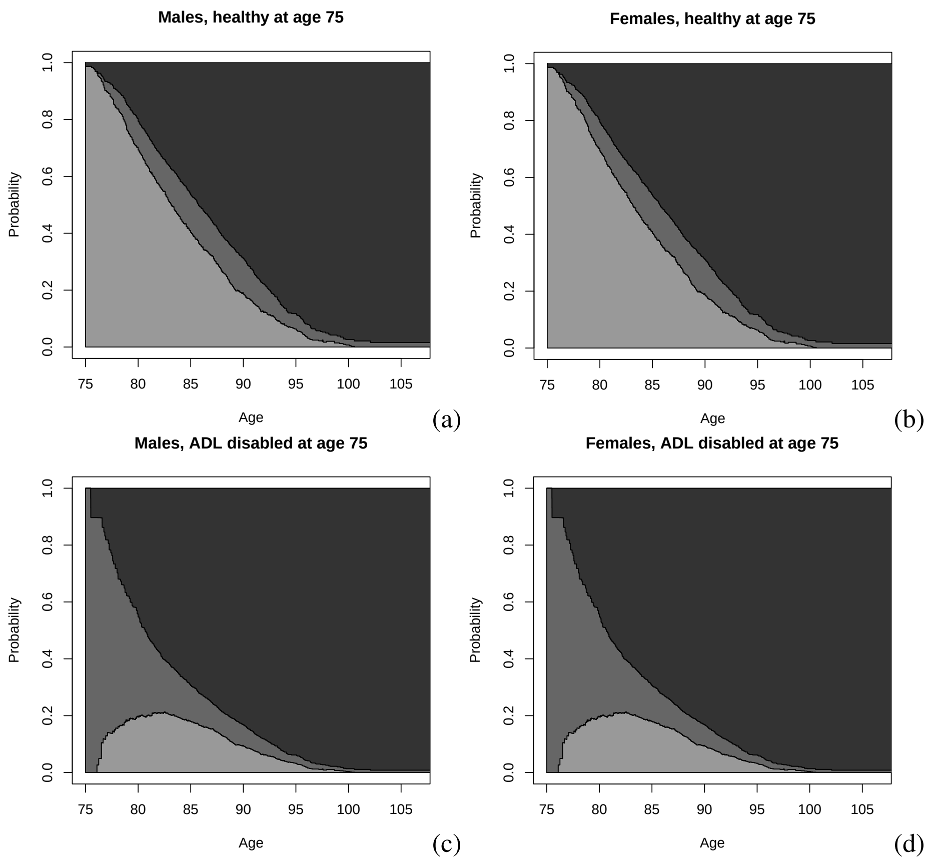


Figure 3(a) and (b) show, for males and females respectively, estimates of $P_{0k}(75, t)$, i.e., conditional probabilities of being in state 0 (healthy), state 1 (ADL disabled) and state 2 (death) ate age t, given that the subject is in state 0 (healthy) at age $s = 75$. The curves are stacked; the lower curve shows $\widehat{P}_{00}(75, t)$, the distance between the lower and upper curve is $\widehat{P}_{01}(75, t)$, the probability of being ADL disabled, and finally the distance between the upper curve and 1 is $\widehat{P}_{02}(75, t)$, the death probability. Figure 3(c) and (d) show estimates of $P_{1k}(75, t)$, probabilities of being in state k at age t, given in state 1 (ADL disabled) at age 75. Clearly, until age 85 at least, the probability of being healthy at age t is much smaller in Figure 3(c) and (d), compared to Figure 3(a) and (b), and the probability of being ADL disabled much larger. Also the probability of having died is considerably larger in Figure 3(c) and (d), compared to Figure 3(a) and (b).

Estimates of the expected remaining time spent in a particular state can be “read off” from Figure 3 as the area between curves. For instance, for males, conditional on being healthy at age 75, an estimate of the expected remaining healthy life, $\widehat{E}_{00}^{75,\tau} = \int_{75}^{\tau} \widehat{P}_{00}(75, u)\, \mathrm{d}u$, is the area under the $\widehat{P}_{00}(75, t)$ curve, i.e., the lightest gray area of Figure 3(a). Similarly, $\widehat{E}_{01}^{75,\tau} = \int_{75}^{\tau} \widehat{P}_{01}(75, u)\, \mathrm{d}u$ is the area between the lower and upper curve, the middle gray area of Figure 3(a). Both are easily calculated from $\widehat{P}_{0k}(75, t)$ using the methods outlined in Section 3. Taking $\tau = 110$, the expected remaining healthy life of males, given healthy at age 75 is estimated to be 9.21 years. For females this number (the area under the lower curve of Figure 3(b)) is estimated as 10.04. Given subjects who are healthy at age 75, the expected remaining life spent in disability equals 2.26 years for males and 4.33 for females. Total remaining life for females (14.37 years) is almost three years longer than males (11.47), but more than two thirds of these additional years are spent in ADL disability.

Given subjects who are ADL disabled at age 75, expected remaining healthy life equals 2.63 years for males and 4.11 years for females; expected remaining life in disability equals 5.21 years for males and 8.59 years for females. The difference between males and females in expected residual life is almost 5 years; again the majority of these additional life years is spent in disability.

At age 75, 10.9% of men and 15.7% of women were disabled. This means that expected remaining healthy life is $0.891 \cdot 9.21 + 0.109 \cdot 2.63 = 8.49$ years for males and $0.843 \cdot 10.04 + 0.157 \cdot 4.11 = 9.11$ years for females. The expected remaining life in disability is $0.891 \cdot 2.26 + 0.109 \cdot 5.21 = 2.58$ years for males and $0.843 \cdot 4.33 + 0.157 \cdot 8.59 = 4.91$ years for females. The former numbers are not unlike those of Crimmins et al. (2009) who report disability-free life expectancy of around 11 years for 70-year-old Americans and of less than six years at age 80. They also report less than 2 years of ADL-disabled life expectancy at both age 70 and age 80, which is less than our numbers.

5. Closing remarks

Product integration is the mapping that switches from the transition hazards of a multistate model towards its matrix of transition probabilities. The product integral may easily be approximated by a finite product which allows us to evaluate expected waiting times and other quantities of interest even in the absence of closed formulae.

We refer to Gill and Johansen (1990) for a comprehensive overview on product integration, including historic remarks. An important statistical paper is Aalen and Johansen (1978), who used product integration on the matrix of the Nelson-Aalen estimators of $A_{jk}(t)$. It is interesting to note that the resulting so-called Aalen-Johansen estimator of $\mathbf{P}(s,t)$, recently implemented in R by Allignol, Schumacher, and Beyersmann (2011) and de Wreede, Fiocco, and Putter (2011), may also be used for numerical approximation. If the transition hazards have been determined for a simulation study, say, we may directly use the approximation of (10). Alternatively, we may simulate a large number of individuals, and subsequently compute the Aalen-Johansen estimator in order to approximate $\mathbf{P}(s,t)$.

Important statistical papers following Aalen and Johansen (1978) include Gill and Johansen (1990), who established compact differentiability of the product integral, which enables use of the function delta method; Andersen, Hansen, and Keiding (1991), who predicted transition probabilities based on Cox-type models for the transition hazards (recently made available in R by de Wreede, Fiocco, and Putter (2011)), and Aalen, Borgan, and Fekjær (2001), who used additive models for prediction.

Both in the example, where we have used the non-parametric Nelson-Aalen and Aalen-Johansen estimators, and in the more technical part of the paper we have relied on a time-inhomogeneous Markov assumption, which may be violated in applications. A major breakthrough in non-parametric inference for non-Markov multistate models was achieved by Datta and Satten (2001), who showed asymptotic unbiasedness of the Aalen-Johansen estimation of the state occupation probabilities (but not transition probabilities) in the presence of random right-censoring; Glidden (2002), also using product integration, subsequently provided weak convergence results. For the special case of the illness-death model without recovery, Meira-Machado, Uña-Álvarez, and Cadarso-Suárez (2006) developed non-parametric estimators of the transition probabilities based on an estimator of a bivariate survival function; see also Meira-Machado and Pardinas (2011) for an implementation in R. More recently, Allignol et al. (2013) showed that simple competing risks-type techniques can be used to compute the Meira-Machado et al. estimator. They also provided a simplified estimator, which allows for random left-truncation and generalizes to arbitrary multistate models. We refer readers to these works for further references on non-Markov models. We also mention Spitoni, Verduijn, and Putter (2012) who studied estimation of transition probabilities in Markov renewal models, also using product integration.

Throughout, our point of view has been that the transition hazards are *given*, i.e., known as in a simulation, determined by expert guesses or statistically estimated, and our aim has been to demonstrate the link from the transition hazards via the transition probabilities towards expected waiting times. In the illustration of Section 4, we have emphasized this link, but have ignored that the data were interval-censored, a common complication, not only with demographic data. One popular approach to account for interval-censoring are embedded Markov chains, see, e.g., Laditka and Wolf (1998), Izmirlian et al. (2000), Van Den Hout, Jagger, and Matthews (2009) and also Lièvre, Brouard, and Heathcote (2003), who have provided the popular IMaCh software (http://euroreves.ined.fr/imach/); see, e.g., Crimmins et al. (2009) and Cambois et al. (2011) for recent applications of IMaCh. A recent review on statistical inference in the illness-death model without recovery in the presence of interval-censoring, has been given by Touraine, Helmer, and Joly (2013) who also pay special attention to life expectancies. In another recent paper, Wolf and Gill (2009) have compared using embedded Markov chains with ignoring interval-censoring; interestingly, these authors found that no method performed uniformly superior with respect to life expectancies.

Finally, we have not considered quantification of uncertainty. If asymptotic distributional properties are available for the initial transition hazard estimation, these may be transferred to estimation of the transition probabilities and expected waiting times using the functional delta method. However, the formulae may become formidable, and we suggest following the advice of Andersen et al. (1993), p.221, who consider the bootstrap as 'an attractive alternative to the calculation of a complicated asymptotic distribution'. For multistate data, different bootstrap variants are available. The most straightforward choice is to repeatedly draw with replacements from the individual units, implemented in R in the `msboot` function of the `mstate` package. Another option, typically computationally faster, is the so-called wild bootstrap (e.g. Martinussen and Scheike 2006; Spitoni, Verduijn, and Putter 2012; Beyersmann, di Termini, and Pauly 2013). Sometimes simply called 'simulation method', the wild bootstrap relies on introducing computer-generated standard normal variates into the estimation procedure in such a way that the asymptotic normal limit is approximated by a Gaussian process with approximately the right covariance structure. Correctness of the bootstrap can, e.g., be shown by first verifying that the bootstrap works for the initial hazard estimation and subsequently applying the functional delta method again. For example, van der Vaart and Wellner (1996), p. 383, prove that the bootstrap works for the non-parametric Nelson-Aalen estimator.

References

Aalen, O., Borgan, Ø., and Fekjær, H. (2001). Covariate Adjustment of Event Histories Estimated from Markov Chains: The Additive Approach. *Biometrics* 57(4): 993–1001. doi:10.1111/j.0006-341X.2001.00993.x.

Aalen, O., Borgan, Ø., and Gjessing, H. (2008). *Survival and Event History Analysis*. New York: Springer. doi:10.1007/978-0-387-68560-1.

Aalen, O. and Johansen, S. (1978). An Empirical Transition Matrix for Non-homogeneous Markov Chains Based on Censored Observations. *Scandinavian Journal of Statistics* 5: 141–150.

Allignol, A., Schumacher, M., and Beyersmann, J. (2011). Empirical Transition Matrix of Multi-State Models: The etm Package. *Journal of Statistical Software* 38(4): 1–15. http://www.jstatsoft.org/v38/i04/.

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*. doi:10.1007/s10985-013-9269-1.

Andersen, P.K., Borgan, Ø., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer. doi:10.1007/978-1-4612-4348-9.

Andersen, P.K., Hansen, L., and Keiding, N. (1991). Non-and Semi-Parametric Estimation of Transition Probabilities from Censored Observation of a Non-Homogeneous Markov Process. *Scandinavian Journal of Statistics* 18: 153–167.

Andersen, P.K. and Pohar Perme, M. (2008). Inference for outcome probabilities in multi-state models. *Lifetime Data Analysis* 14(4): 405–431. doi:10.1007/s10985-008-9097-x.

Beyersmann, J., di Termini, S., 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. doi:10.1111/j.1467-9469.2012.00817.x.

Brookmeyer, R. and Crowley, J. (1982). A Confidence Interval for the Median Survival Time. *Biometrics* 38(1): 29–41. doi:10.2307/2530286.

Cambois, E., Laborde, C., Romieu, I., and Robine, J.M. (2011). Occupational inequalities in health expectancies in France in the early 2000s: Unequal chances of reaching and living retirement in good health. *Demographic Research* 25(12): 407–436. doi:10.4054/DemRes.2011.25.12.

Crimmins, E.M., Hayward, M.D., Hagedorn, A., Saito, Y., and Brouard, N. (2009).

Change in Disability-Free Life Expectancy for Americans 70 Years Old and Older. *Demography* 46(3): 627–646. doi:10.1353/dem.0.0070.

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. doi:10.1016/S0167-7152(01)00155-9.

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): 1–30. http://www.jstatsoft.org/v38/i07/.

Gill, R. and Keilman, N. (1990). On the estimation of multidimensional demographic models with population registration data. *Mathematical Population Studies* 2(2): 119–143. doi:10.1080/08898489009525298.

Gill, R.D. and Johansen, S. (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis. *The Annals of Statistics* 18(4): 1501–1555. doi:10.1214/aos/1176347865.

Glidden, D.V. (2002). Robust Inference for Event Probabilities with Non-Markov Event Data. *Biometrics* 58(2): 361–368. doi:10.1111/j.0006-341X.2002.00361.x.

Izmirlian, G., Brock, D., Ferrucci, L., and Phillips, C. (2000). Active Life Expectancy from Annual Follow-Up Data with Missing Responses. *Biometrics* 56(1): 244–248. doi:10.1111/j.0006-341X.2000.00244.x.

Juster, F.T. and Suzman, R. (1995). An Overview of the Health and Retirement Study. *The Journal of Human Resources* 30: S7–S56. doi:10.2307/146277.

Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., and Jaffe, M.W. (1963). Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *JAMA* 185(12): 914–919. doi:10.1001/jama.1963.03060120024016.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. London: Chapman & Hall. doi:10.1007/978-1-4899-3132-0.

Kuo, T.M., Suchindran, C.M., and Koo, H.P. (2008). The multistate life table method: An application to contraceptive switching behavior. *Demography* 45(1): 157–171. doi:10.1353/dem.2008.0013.

Laditka, S.B. and Wolf, D.A. (1998). New Methods for Analyzing Active Life Expectancy. *Journal of Aging and Health* 10(2): 214–241. doi:10.1177/089826439801000206.

Lièvre, A., Brouard, N., and Heathcote, C. (2003). The estimation of health expectancies from cross-longitudinal surveys. *Mathematical Population Studies* 10(4): 211–248. doi:10.1080/713644739.

Martinussen, T. and Scheike, T. (2006). *Dynamic Regression Models for Survival Data*. New York: Springer.

Meira-Machado, L. and Pardinas, J. (2011). p3state.msm: Analyzing Survival Data from an Illness-Death Model. *Journal of Statistical Software* 38(3): 1–18. http://www.jstatsoft.org/v38/i03.

Meira-Machado, L., 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. doi:10.1007/s10985-006-9009-x.

Schoen, R. (2005). Intrinsically dynamic population models. *Demographic Research* 12(3): 51–76. doi:10.4054/DemRes.2005.12.3.

Spitoni, C., Verduijn, M., and Putter, H. (2012). Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi-State Models. *The International Journal of Biostatistics* 8(1): 51–76. doi:10.1515/1557-4679.1375.

Touraine, C., Helmer, C., and Joly, P. (2013). Predictions in an illness-death model. *Statistical Methods in Medical Research* doi:10.1177/0962280213489234.

Van Den Hout, A., Jagger, C., and Matthews, F.E. (2009). Estimating life expectancy in health and ill health by using a hidden Markov model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58(4): 449–465. doi:10.1111/j.1467-9876.2008.00659.x.

van der Vaart, A. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer.

Wolf, D.A. and Gill, T.M. (2009). Modeling transition rates using panel current-status data: How serious is the bias? *Demography* 46(2): 371–386. doi:10.1353/dem.0.0057.