The effect of stroke on dementia onset: Left-truncation and right-censoring

R Weißbach 1, A Dörre 1, D Wied 2, G Doblhammer 3,4, and A Fink 4

1 Chair in Statistics and Econometrics, University of Rostock, Germany
2 Chair in Statistics and Econometrics, University of Cologne, Germany
3 Chair in Empirical Social Research/Demography, University of Rostock, Germany
4 German Center for Neurodegenerative Diseases, Bonn, Germany

Received 8 March 2021, revised 1 June 2021, accepted zzz

We monitor a sample of a quarter million people over a period of nine years, and are interested in the effect of a stroke on the probability of dementia onset. People already deceased before the period are randomly left-truncated. The ages at a stroke event or dementia onset are conditionally fixed right-censored, when either event may still occur, but only after the period. We incorporate death, and model the history of the three events by a homogeneous Markov process. The compensator for the respective counting processes is derived, and Jacod’s formula yields the likelihood contribution, conditional on observation. An Appendix is devoted to the role of filtrations in deriving the likelihood, starting with the simplification of merged non-dead states. Asymptotic normality of the estimated intensities, relative to the size of the sample including the truncated persons, is derived by using martingale theory. Data from a German health insurance company reveals that after a stroke, the intensity of dementia onset is increased from 0.02 to 0.07, for Germans born in the first half of the 20th century. The intensity difference has a 95%-confidence interval of [0.048, 0.051] and the difference halves, due to Simpson’s paradox, when we extend the analysis to an age-inhomogeneous model.

Key words: Truncation; Morbidity; Filtering; Confidence interval; Multi states

1 Introduction

In the ageing populations of the western hemisphere, ageing-related conditions such as dementia and cardiovascular diseases will become more prevalent. The future number of dementia cases worldwide is forecasted to increase from 46 million in 2015 to 131.5 million by 2050 (Prince et al., 2015). For the aging
country Germany, Doblhammer et al. (2018) forecast an increase up to 2.8 million in 2050. Dementia is associated with high individual and societal burden as it is a very high-care and high-cost disease (see Michalowsky et al., 2019). Established risk factors for dementia are vascular diseases such as hypertension, diabetes mellitus and cerebrovascular diseases (see Mangialasche et al., 2012). We are interested in the consequences of a stroke (as e.g. Hbid et al., 2020) on dementia. Several community-based studies have shown that stroke in particular is associated with an increased risk of subsequent dementia (see Desmond et al., 2002; Ivan et al., 2002; Reitz et al., 2008). The aim of the current study is to quantify the effect of stroke on the incidence of dementia in a large cohort of German Health Claims Data (HCD) which also includes the institutionalized population living in nursing homes. Persons living in nursing homes are at a high risk for both diseases, stroke and dementia, but are often under-represented in community-dwelling studies. Individual conditions typically evolve in time from healthy to having had a stroke, for some, and further to dementia, for even fewer. Finally everyone becomes invisible after death. Together with a limited period of monitoring, missing data defects of left-truncation and right-censoring must occur and we model the health history as sequence of Markovian events to efficiently estimate the post-stroke dementia incidence. For the analysis of event histories with missing data, Andersen et al. (1993) has become a standard.

A prerequisite for our aim is a confidence interval for the incidence parameter, which we construct in Wald-type by means of a centralised estimator. An estimator is a statistic of the data so that determining a centre for the estimator requires defining a centre for the data, in our case of a stochastic history. In order to derive the point estimate we maximise the likelihood, or to be precise, a conditional likelihood. The data are stochastic processes, so is any of the statistics that counts state changes. We think of the latter as sum of increments. The increments are Benoulli experiments and the probability distribution of an increment depends on the expectation, which is inherently associated with a probability measure. We need some effort, going into defining an expectation, or other centre, for a counting process. Here, as centre, a ‘trend’ of a counting process, with respect to a filtrations must be derived. The filtration represents our available information. In the Doob-Meyer decomposition of the counting processes, the trend is named compensator and the remaining additive noise is a martingale.

The described methods allow to account for the types of missing values that typically arise when studying morbidity (and co-morbidity). These are right-censoring, if at least one of the events is past the monitoring period, and left-truncation, if the death event is prior to the period. All this is easiest explained by merging the not-dead states to an ‘alive’ state and we therefore included a mortality model. Of course, age is a natural covariate for the analysis of morbidity, including stroke and dementia. We discuss the age-homogeneous transition behaviour, where the probability to suffer either event, stroke or dementia onset,
Population: Germans born 1900–1954 (Size 76,239,006)

Figure 1  Data: Trajectory from population to the left-truncated mass and right-censored persons (Population size: German Statistical Office (2004), without stillborn)

given the disease has not yet occurred, is equal for all ages. Medically such equality is very doubtful, and we extend the model to age-inhomogeneous transition intensities.

The current study neglects a potential cohort trend, in contrast to Weißbach et al. (2021), who use the same data. Also our study does not consider the left-censoring information of those with dementia prior to the monitoring period (with dementia onset at unknown age). On the other hand, the current study is superior, in the sense that Weißbach et al. (2021) only perform a survival analysis and here an event history analysis is performed. Already the two states ‘dementia’ and ‘death’ of the study in Weißbach et al. (2021) do not clearly qualify for a survival analysis. With the two states at interest ‘stroke’ and ‘dementia’ a survival analysis is impossible, because of the retrospective sampling design, is governed by the two additional states ‘alive’ and ‘dead’, especially that of left-truncation. To this end, the current study uses counting process notation and theory.

Our result is for Germany, and we estimate the risk of a dementia onset approximately doubles after a stroke, which is similar to that for Korea reported by Kim and Lee (2018).

2 MODEL AND ESTIMATION TECHNIQUE

The population we aim at are Germans born in the first half of the 20th century (see first line in Figure 1). Let $t$ count the years after a person’s 50th birthday, typically the earliest age at which a stroke occurs, and we simply call $t$ ‘age’. In our multi-morbidity analysis, we consider the following three increasingly poor deviations from health ($H$): having had at least one stroke ($S^1$), having dementia ($D$) and being dead ($d$). Let $X_t$ indicate a person’s state at the age of $t$. For a person with dementia we say $X_t = D$, irrespective of whether a stroke has preceded dementia onset at that age or will occur later. Our multi-state model about the connection between stroke and dementia for Germany is similar to that for South Korea by Kim and Lee (2018), both use national Health Claims Data (HCD). We collect the possible state transitions in $\mathcal{I} := \{HS^1, S^1D, HD, Hd, S^1d, Dd\}$. We do not look any further than $\tau$ years, and let, in continuous time, the history $X = \{X_t, t \in [0, \tau]\}$ be defined on the probability spaces $(\Omega, \mathcal{F}, P_\lambda)$. We assume throughout the Markov property, so that the history is determined by the transition intensities $\lambda_{hj}(t)$ :=

© 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

www.biometrical-journal.com
lim_{u \to 0} P_X (X_{t+u} = j \mid X_t = h)/u. With the exception of Section 3.3, we model the population of Germany (at that time) as age-homogeneous, i.e. assume \( \lambda_{hj} (t) \equiv \lambda_{hj} \). By parameter we mean the vector \( \lambda := (\lambda_{hj}, hj \in I) \) and denote true parameters as \( \lambda_{hj}^0 \). Precise definitions of what follows for the age-homogeneous model will be given in Appendix B. We are neither in possession of the population data, nor of a simple sample of size \( n_{all} \). However, from the latter we observe those \( n < n_{all} \) histories, that occur – completely or in part – during our monitoring period between 2003 and 2013 (see Figure 1). We aim at a statistical analysis that takes the missing information into account. Abstractly speaking, a change in information changes the \( \sigma \)-algebra \( F \) of the model. What we also observe, for each observed history, is the ‘age-at-study-begin’, a positive duration \( U \) (presumably independent of \( X \)), between the calendar dates of the 50\(^{th}\) birthday and the study begin on 01/01/2004. Let us become more specific and for the most specific display of the theory we analyse a preliminary model, in Appendix A, where we specialize to mortality, i.e. collapse all ‘non-dead’ states to one ‘alive’ state.

For the stochastic process, \( X_t \), adapted to the filtration \( N_t \in F \), a change in information must change the filtration. And we need to distinguish between existing information and information available to the researcher. On the one hand, the age-at-study-begin is, as a second measurement compared to a model without \( U \), new information and must lead to a finer \( \sigma \)-algebra and hence to a finer filtration. On the other hand, less information available, in our case the loss of left-truncated persons deceased before 2004 and right-censored events after 2013 (see again Figure 1), must on the contrary lead to a coarser filtration. The filtration is derived in Appendix B. We name the filtration \( U^F \) that finally models our available information. By Jacod’s formula, the logarithm of the likelihood of \( (X, U, 1_{\{X \neq d\}}) \), conditional on the last two coordinates, up to \( \tau \), is shown to be (see especially (15) in Section B)

\[
\log U L_c^e (X, U | \Lambda) = \int_0^\tau \sum_{hj \in I} (\log U Y^c_h (t) + \log \lambda_{hj}) d U N^c_{hj} (t) - \sum_{hj \in I} \int_0^\tau U Y^c_h (t) \lambda_{hj} dt. \tag{1}
\]

(When, in an expression with \( hj \in I \), only \( h \) appears, the first position of \( hj \) it meant. Especially \( h \neq d \), because death is absorbing.) The expression uses the Doob-Meyer decomposition of \( N_{hj} (t) \), stacked to \( N_X \). Let \( N_{hj} (t) \) ‘count’ the transition between states up to age \( t \). Furthermore, we indicate residence in state \( h \) at the age of \( t \) by \( Y^c_h (t) \). Note also that, for the ease of notion and notation, we assume a fifty-year old to be healthy, i.e. \( X_0 = H \), otherwise conditioning would need extension to include \( X_0 \) as condition. The observed left-truncated and right-censored version thereof is

\[
U N^c (t) := \int_0^t C(s) d U N (s).
\]

and \( U Y^c_h (t) := C(t) 1_{\{U < t\}} Y^c_h (t) \) with \( C(t) \) being one, as long as the unit is not censored, i.e. for \( t \leq u + 9 \) and \( U N (t) := N_X (t) - N_X (\min(t, U)) \). The data, unobserved and observed statistics are sketched in Figure 2 for \( n_{all} = 4 \).
The log-likelihood contribution (1) for one person of the sample holds for any person, truncated or not. A truncated unit will contribute an ineffective numerical zero to the expression, as argued in detail in Appendix A. Now we consider all persons in the simple sample (of size $n_{\text{all}}$). Without loss of generality, we sort those who are truncated to the end of the unobserved sample and name the (random) size of the observed sample $n$. Now, due to the simple sample assumption, the conditional density of the sample is the product of the persons’ conditional densities (see Section A.3),

$$\log U_L^c(\text{data}; \lambda) = \sum_{i=1}^n \log U_L^c(X_i, U_i | \lambda).$$  \hspace{1cm} (2)

This also requires $U_i$ to be random. The unique root of its derivative is simply

$$\hat{\lambda}_{hj} = \frac{U^{N_1^c}}{\int_0^{U_{Y_1^c}(t)} dt},$$  \hspace{1cm} (3)

where each counting process carries a third index $i$, identifying the not truncated unit. The dot signals the sum over the units, $\sum_{i=1}^n$. One can avoid integration in the denominator in (3). Of the interesting states for $h$, $H$ and $S^1$, rewrite e.g. for $h = H$:

$$\int_0^{\tau} u Y_{H^c}^c(t) dt = \sum_{i=1}^n \int_{U_i}^{U_i+9} \mathbb{1}\{X_i(t^-) = H\} dt = \sum_{i=1}^n \mathbb{1}\{X_i(U_i) \geq H\} \left\{9 \mathbb{1}\{X_i(U_i+9) = H\} + \mathbb{1}\{X_i(U_i+9) \in \{S^1, D\}\} \sup_{t \geq X_i(U_i)} \{X_i(t) = H\} - U_i\right\}$$

Figure 2 Four health histories, unobservable $N_{S^1 D^*}(t) := \sum_{i=1}^n N_{S^1 D^*}(t)$ and observed $U_{N_{S^1 D^*}}(t)$ (offset between $N_{S^1 D^*}$ and $U_{N_{S^1 D^*}}$ for clarity)
2.1 Asymptotic properties

By verifying regularity conditions, we arrive at the (joint) asymptotic distribution of the estimators \( \hat{\lambda}_{HD} \) and \( \hat{\lambda}_{S1D} \) by standard results on martingales. It depends on \( m_h(t) \), the prevalence of state \( h \) at age \( t \) in the population, and \( \beta_{h0} \), the probability of a person from the sample to be observed, i.e. not to be truncated.

**Theorem 2.1** Under Conditions (B1)-(B3) and \( m_h(t) \) defined in (17) it is \( \hat{\lambda} \), composed of (3), consistent and \( \sqrt{n_{all}}(\hat{\lambda} - \lambda_0) \overset{D}{\rightarrow} N(0, \Sigma^{-1}(\lambda_0)) \) with diagonal matrix \( \Sigma(\lambda_0) \) of diagonal elements

\[
\sigma_{hj,hj}(\lambda_0) := \beta_{h0} \int_0^\tau \sum_{hj \in \mathcal{I}} \frac{m_h(t)}{\lambda_{hj,0}} dt, \quad \text{for} \quad hj \in \mathcal{I}.
\]

Roughly speaking, the arguments of the proof (given in Appendix B) are similar to the case of the univariate parameter space in Appendix A. In addition the multivariate parameter space here results in a diagonal matrix of asymptotic variances, and positive definiteness follows from the positivity of the diagonal elements.

It remains to consistently estimate \( \Sigma(\lambda_0) \), in order to construct a confidence interval for the difference \( \hat{\lambda}_{S1D^0} - \hat{\lambda}_{HD^0} \) with the standard error. This is a Wald-type test for the effect of a stroke \( S1 \) on the intensity of dementia onset. By Theorem 2.1, \( Var(\hat{\lambda}) = Var(\sqrt{n_{all}}\hat{\lambda})/n_{all} = \Sigma^{-1}(\lambda_0)/n_{all} \), so that, for estimating the asymptotic variance in Theorem 2.1, define \( -J_\tau(\lambda_0) \) as

\[
diag \left( \frac{\partial^2}{\partial \lambda_{hj}^2} \log u \mathbb{L}_n^c(\text{data}|\lambda=\lambda_0; hj \in \mathcal{I}) \right) = -\diag \left( \frac{U_{N_{hj,0}}(\tau)}{\lambda_{hj,0}^2} ; hj \in \mathcal{I} \right).
\]

Now, as \( J_\tau(\lambda_0)/n_{all} \overset{n_{all} \to \infty}{\rightarrow} \Sigma(\lambda_0) \) (see Andersen et al., 1993, Formula (6.1.11)), it is

\[
Var(\hat{\lambda}_{hj}) = J_\tau^{-1}(\lambda)_{hj,hj} = \frac{U_{N_{hj,0}}(\tau)}{\int_0^\tau U_{Y_{hj}^c}(t) dt}, \quad \text{due to} \quad J_\tau^{-1}(\lambda_0)_{hj,hj} = \lambda_{hj,0}^2 / U_{N_{hj,0}}(\tau), \quad (3) \quad \text{and the CMT. The estimated standard error of} \ \hat{\lambda}_{hj} \ \text{is the square root thereof. Note that even though} \ n_{all}, \ \text{and with it the standard error, is not observable, the standard error estimate is indeed observed.}
\]

2.2 Finite sample properties

We conduct a Monte Carlo simulation in order to study the degree of consistency, measured in (root) mean square error, and asymptotic normality of the estimators claimed by Theorems 2.1 and A.1, for small but realistic sample sizes. We refrain from indicating the true parameter by the sub/superscript and drop 0 in this section. Especially we will mimic the situation of the up-coming data example in Section 3. It is tempting to use as true parameter the estimate of the example. However, one must keep in mind that if the method is flawed in the situation of the application, then using the application’s estimate as parameter for
a simulation study may also let one conclude a valid procedure. Fortunately, the relation between health, stroke, dementia and death is a well studied area and we think that, consolidating our estimates with values from different literature sources, we can hope to be in a region of the parameter space for the simulation study to be meaningful. Of course, literature results are in many cases less specific or more specific, e.g. for age-groups and/or other countries, and their values are hence only a rough indication.

2.2.1 Simulation for mortality model

For the simplification to the mortality model in Appendix A, we simulate $T$ from an Exponential distribution. The value (for (10)) that will be given in Section 3.1, and is used now, is $\hat{\lambda} = 0.02$. In order to consolidate it with the literature, note first that $P(T \leq 1) = 1 - e^{-\lambda} \approx \lambda$ for small $\lambda$ (by the well-known $e^x \approx 1 + x$, for small $x$). The one-year death incidence (and hence intensity) in Germany are $\approx 0.96$ mio deaths in 2020, restricted to the over 50 year-old’s (Destatis 2021b), relative to $\approx 35.5$ mio alive over 50 year old’s in 2019 (Destatis 2021a). The ratio is $F(1) \approx \lambda \approx 0.027$ and roughly verifies our value.

Furthermore we simulate, independent thereof, $U \sim \text{Exp}(0.004)$, by this we have (on average) for $n = 250,000$ observations, $n_{uncens} = 40,000$ uncensored, resembling our data situation. Table 1 shows, for 2,000 repetitions, that the mean square error decreases to an irrelevant size with increasing sample since, already for sample sizes below that of our example. That confirms that consistency, as stated in Theorem A.1, will already have kicked-in for our data example’s size. For the assessment of the normality, to be realistic for our example, we draw $n_{all} = 1.5$ million units, with 2,000 repetitions. The plot of the kernel smoothed histogram is in Figure 3 (left panel) and confirms the normal shape.

2.2.2 Simulation for age-homogeneous morbidity model

For simulating from the morbidity model we collect the $\lambda_{hj}$ in some matrix (left side):

$$Q = \begin{pmatrix}
-\lambda_H & \lambda_{HS}^1 & \lambda_{HD} & \lambda_{Hd} \\
0 & -\lambda_{S}^1 & \lambda_{S1}^D & \lambda_{S1}^d \\
0 & 0 & -\lambda_{Dd} & \lambda_{Dd} \\
0 & 0 & 0 & 0
\end{pmatrix} = \begin{pmatrix}
-0.086 & 1/30 & 0.02 & 1/30 \\
0 & -0.17 & 0.07 & 1/10 \\
0 & 0 & -0.1 & 1/10 \\
0 & 0 & 0 & 0
\end{pmatrix}$$

Here, the small dot signals summation over the respective index $\sum_j$. In order to sample paths of the latent histories, we describe the homogeneous Markov process with generator $Q$ as in Albert (1962). Let
Figure 3  Simulated finite distribution of estimator for \( n = 250,000 \) observations (uncensored thereof: \( 40,000 \)) in mortality model, (10) for \( \lambda = 0.02 \) (left), and in morbidity model, (3) for \( \lambda_{HD} = 0.02 \) and \( \lambda_{S^1D} = 0.07 \) (middle/right) \( \approx 80\% \) truncated units), kernel smoothed

\[ X_t \equiv H \text{ on } [0, T_1] \text{ with } T_1 \text{ having cumulative hazard function } A_H(t) = \lambda_H t \text{ (i.e. } T_1 \sim \text{Exp}(\lambda_H)). \]

Then in \( t = T_1 \), \( X \) migrates from \( H \) to \( j \in \{S^1, D, d\} \) with \( p_{Hj} = Pr(X_{T_1} = j|X_{T_1} = H) \), with \( p_{hj} = \lambda_{hj}/\lambda_h \). Then (if \( X_{T_1} \neq d \)), \( X_t \equiv j \) on \( [T_1, T_2[ \text{ with } T_2 \text{ (and } j \in \{S^1, D\}) \text{ having cumulative hazard function } A_j(t + T_1) - A_j(T_1) = \lambda_j t \text{ (i.e. } T_2 \sim \text{Exp}(\lambda_j)). \)

Then in \( t = T_2 \), \( X \) migrates from \( j \) to \( k \in \{D, d\} \) with \( p_{jk} \). Then (if \( X_{T_2} \neq d \)), \( X_t \equiv D \) on \( [T_2, T_3[ \text{ with } T_3 \text{ having cumulative hazard function } A_{Dd}(t + T_2) - A_{Dd}(T_2) = -\lambda_{Dd}t \text{ (i.e. } T_3 \sim \text{Exp}(\lambda_{Dd})). \]

Again we aim at using the example’s estimates as input, and foster them by literature values. As in the previous Section 2.2.1, it may even be difficult to find the rough intensities in the literature. Again, one-year transition rates might be available, so-called incidences. Given all incidences, in order to specify \( Q \), we may use the relation \( P(t) = e^{tQ} \) (see e.g. Weißbach et al., 2009, Formula (2)), where \( P(t) \) denotes the matrix of \( t \)-year probabilities \( P(X(t) = j|X(0) = h) \). Now one can invert the relation by a spectral decomposition of \( P(t) \) (see Webel and Wied, 2012, Sect. 4.3). Similar to Section 2.2.1, an approximation for the case of \( t = 1 \), namely \( e^{Q} \approx I + Q \) (for \( Q \) ‘small’) suggests one-year incidences as approximation for the intensities.

In the data example we will only estimate two of the \( \lambda_{hj} \). The other four, and the distribution of \( U \), are nuisance parameters. For the intensity of dementia onset for a person after a stroke, for instance, Leys et al. (2005) find a one-year incidence of 7%. Hence \( \lambda_{S^1D} = 0.07 \) is conceivable, and we will have the same value in our example, using (4). For the intensity of dementia onset without a stroke, we will calculate \( \lambda_{HD} = 0.02 \). Similar, Vieira et al. (2013) collect, however independent on whether a stroke preceded, one-year incidences of 0.008, 0.001 and 0.002 (dependent on the country and age range) for persons below age 65. Our value of 0.02 is larger but aims also at high ages and we still use 0.02.
All other parameters are nuisance for our question, but necessary for the simulation, and estimation of the nuisance parameters also follows (4). Our data reveal \( \hat{\lambda}_{S1,d} = 0.07 \), but van den Bussche et al. (2010) find for Germany, from health claims data, that 17% die within one year after a stroke. We consider only the first stroke, which should lower the value and we use as value in between 0.1. Our data contain the estimate \( \hat{\lambda}_{H,d} \) of 9%, whereas Garcia-Ptaceka et al. (2014) find for Sweden from registry data that 11% die yearly with dementia. Of course, conceptionally, those that had died from other causes would need to be subtracted, but we use as value in between 0.1. We have for \( \lambda_{HS1} \) the value 0.02. Garcia-Ptaceka et al. (2012) find incidences above 1% only for French above age 80 and for Italian and British above age 75. We increase slightly to \( 1/30 \approx 0.03 \). For the death intensity without stroke or dementia \( \lambda_{H,d} \), we did not find a relevant study. Our general death hazard from Section 2.2.1 is \( \lambda = 0.02 \) and we increase slightly to \( 1/30 \approx 0.03 \). As distribution for the age-at-study-begin \( U \) we follow Dörre (2017) and assume for the number of births a homogeneous Poisson process.

Our final choice is collected in (5), right hand side. In order to reduce the computational burden, we let \( n_{all} \) vary from 1,000, 5,000, 10,000 and 20,000, all below the example’s sample size. The homogeneous Poisson process for births implies the distribution of \( U \) to be Uniform (see Dörre, 2017, Lemma 2). The longer birth period, the more persons are left-truncated and in our example of Section 3 the population will be born within 54 years. However, to start with, we only use 30 years, i.e. \( U \sim U[0,30] \). Combined with (5), on average, 48.7.3% of the simulated sample are unobserved due to left-truncation. The number of simulation replications is 10,000.

The simulation confirms, overall, consistency of \( \hat{\lambda}_{hj} \), as stated in Theorem 2.1. Especially, the root mean squared error drops, as a function of in sample size (see Table 2). In general, the maximum likelihood method is not unbiased in a finite sample, being equally true for CML, but vanishes for large samples in case of consistency. The simulation averages of \( \hat{\lambda}_{hj} - \lambda_{hj} \), the bias, show that the bias is small, already in finite samples (see Table 2). Also the estimator (4) of the standard error seems to be consistent for the same

<table>
<thead>
<tr>
<th>( n_{all} )</th>
<th>( \lambda_{H,D} )</th>
<th>( \lambda_{S1,D} )</th>
<th>( \lambda_{H,D} )</th>
<th>( \lambda_{S1,D} )</th>
<th>( \lambda_{H,D} )</th>
<th>( \lambda_{S1,D} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>-0.0008</td>
<td>0.0352</td>
<td>0.3029</td>
<td>1.1435</td>
<td>94.39%</td>
<td>94.88%</td>
</tr>
<tr>
<td>5,000</td>
<td>0.0003</td>
<td>0.0075</td>
<td>0.1345</td>
<td>0.5124</td>
<td>94.51%</td>
<td>95.20%</td>
</tr>
<tr>
<td>10,000</td>
<td>-0.0009</td>
<td>0.0055</td>
<td>0.0947</td>
<td>0.3630</td>
<td>95.16%</td>
<td>94.82%</td>
</tr>
<tr>
<td>20,000</td>
<td>-0.0005</td>
<td>-0.0013</td>
<td>0.0663</td>
<td>0.2562</td>
<td>95.20%</td>
<td>95.01%</td>
</tr>
</tbody>
</table>
reason (see Table 3), without a formal proof in the above section. Simulations show similar behaviour for all other $\hat{\lambda}_{hj}$ (and their standard errors).

**Table 3** Bias, root means squared error (rMSE) for standard error estimator (4), $n \approx 0.513 \times n_{all}$

<table>
<thead>
<tr>
<th>$n_{all}$</th>
<th>$\text{Var}(\hat{\lambda}_{HD})$</th>
<th>$\text{Var}(\hat{\lambda}_{S1D})$</th>
<th>$\text{Var}(\hat{\lambda}_{HD})$</th>
<th>$\text{Var}(\hat{\lambda}_{S1D})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>-0.0015</td>
<td>0.0415</td>
<td>0.0156</td>
<td>0.3022</td>
</tr>
<tr>
<td>5,000</td>
<td>-0.0001</td>
<td>0.0025</td>
<td>0.0014</td>
<td>0.0262</td>
</tr>
<tr>
<td>10,000</td>
<td>0.0000</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0092</td>
</tr>
<tr>
<td>20,000</td>
<td>0.0001</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

We now come to the conditions of our example, where the population are births of the cohorts 1900 to 1954. We still assume homogeneous Poisson process for the births, i.e. now $U \sim U[0, 54]$, and use (5). We use $n_{all} = 100,000$, which is still below the sample size of the application, and (only) 2,000 simulation loops now. Middle and right panel of Figure 3 now confirm the asymptotic normality of $\hat{\lambda}_{HD}$ and $\hat{\lambda}_{S1D}$ (of (3)) stated in Theorem 2.1. The theorem also states asymptotic independence of the two estimators, which will be important when deriving a confidence interval for the difference later on. The simulated correlation of $-0.02$ confirms the independence.

### 3 Result: Stroke and Dementia

As the population, we consider Germans born between January 1st, 1900, and December 31st, 1954. On 01/01/2004, we drew a simple sample of size $n_{all}$ from that population. Truncated (and not observed) is that part of the sample, that did not survive to 01/01/2004 (see again Figure 1). Those 250,000 people made accessible to us, were reduced by 4121 persons with implausible information over time on sex, birth year or region of living. The $n = 245,879$ contribute to the conditional likelihood (2). To be a bit more specific, the persons are all insurants of the Allgemeine Ortskrankenkasse (AOK), the with 25 million at the time of sampling largest public health insurance company in Germany. Studies have shown that the population insured by the AOK had, on average, higher morbidity rates and a higher proportion of persons with low socio-economic status compared with other statutory health insurance funds and also compared with private health insurance funds, which are not considered here (see Schnee, 2008). We ignore any possible selection bias in this respect. Also simplifying, we have assumed in Section 2 that $X_0 = H$, i.e. that a person cannot die before, and stays healthy in our sense, until age 50. Hence, we exclude those alive from the sample that at the age for 50 already incurred a stroke or suffer from dementia. Bias and loss in efficiency are beyond our scope here, but clearly small. We monitor histories for transitions up to
31/12/2013, i.e. up to τ = 54 + 9 = 63. By doing so, the n persons are at most followed until the age of 113 (see Figures 1 and 5).

### 3.1 Mortality

The AOK Health Claims Data (HCD) data include information about age, year of birth and date of exit (death or switch to another insurance company). For the preliminary mortality model, we analyse in Appendix A, \( n_{\text{uncens}} = 43,457 \) persons, out of \( n \), die within the monitoring period between 2004 and 2013. For the Exponentially distributed age-at-death, \( T \), the point estimate (10), with narrow confidence interval, is \( \hat{\lambda} = 0.022 \) (see Table 4) and results in an expected lifetime of \( 1/0.022 \approx 45 \) (plus 50) years. That is apparently too large, primarily due to the assumption of a constant hazard, and to some extent because death is impossible before age 50 in this model. We account for the well-documented age-inhomogeneous increase in the hazard, in demography usually modelled as a Gompertz distribution, by piecewise constant hazards in Section 3.3. Now, we first disentangle the three non-dead states ‘Healthy’, ‘Stroke’ and ‘Dementia’, still age-homogeneously.

### 3.2 Age-homogeneous transition of stroke to dementia

The model is very similar to ‘Mortality of Diabetics in the County of Fyn’ (see Andersen et al., 1993). However, the age-at-study-begin, \( U \), and the age-at-study-end, \( U + 9 \), share a simpler relation than in (Andersen et al., 1993, Examples III.3.6, IV.1.7 and VI.1.4). Other than the life-threatening effect of diabetes, we investigate the well-known association of a stroke with all-cause dementia. On the one hand, cerebrovascular lesions may substantially affect cognition through several mechanisms such as changes in blood flow and oxygen supply or chronic inflammation (Hu and Chen, 2017). On the other hand, dementia and stroke share common risk factors such as hypertension, diabetes or atrial fibrillations (Pendlebury and Rothwell, 2009). Desmond et al. (2002) reveal an increased relative risk (RR) for dementia of 3.8 after a stroke, adjusted for several demographic factors and cognitive status. Within the Framingham Study (Ivan et al., 2002), the adjusted RR of dementia is estimated to be 2.0. The Rotterdam Study (Reitz et al., 2008) shows that a stroke doubles the risk of dementia (Hazard ratio: HR=2.1). A systematic review and meta-analysis reveals a pooled HR of between 1.7 and 2.2 (Kužma et al., 2018). Another result, but

<table>
<thead>
<tr>
<th>'Survivors' ( n_{\text{cens}} )</th>
<th>'Deaths' ( n_{\text{uncens}} )</th>
<th>'Time at Risk' (in years) ( \sum_{i=1}^{n} I_{{u_i \leq t_i &lt; u_i + 9}} (t_i - u_i) )</th>
<th>Point ( \hat{\lambda} )</th>
<th>SE ( \sqrt{\hat{J}_{-1}(\hat{\lambda})} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>202,407</td>
<td>43,472</td>
<td>180,163</td>
<td>0.0217</td>
<td>0.000104</td>
</tr>
</tbody>
</table>

Table 4 Statistics, point estimate (10) and standard error (SE) by (13) for mortality
Table 5  Statistics, point estimates (3) and estimated standard errors (SE, (4)) for age-homogeneous model

<table>
<thead>
<tr>
<th></th>
<th>'Dementia after Stroke'</th>
<th>'Time after Stroke'</th>
<th>Point</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\int_0^T U y_{S1,D}(t) dt$</td>
<td>$\hat{\lambda}_{S1,D}$</td>
<td>$\sqrt{J^{-1}(\hat{\lambda}_{S1,D})}$</td>
</tr>
<tr>
<td>Unconditional</td>
<td>8,105</td>
<td>115,566</td>
<td>0.0701</td>
<td>0.00078</td>
</tr>
<tr>
<td>'Dementia without Stroke'</td>
<td>'Healthy Times'</td>
<td>$\int_0^T U y_{H,D}(t) dt$</td>
<td>$\hat{\lambda}_{HD}$</td>
<td>$\sqrt{J^{-1}(\hat{\lambda}_{HD})}$</td>
</tr>
<tr>
<td>Unconditional</td>
<td>41,775</td>
<td>1,997,092</td>
<td>0.0209</td>
<td>0.000102</td>
</tr>
</tbody>
</table>

without multi-states, is that of Savva and Blossom (2010) who report a hazard ratio of 2. Based on South Korean health claims data, and also using multi-state methods, Kim and Lee (2018) find a 2.4-fold risk of subsequent dementia after a stroke.

Here, additionally to mortality information, the AOK HCD data contain information about outpatient and inpatient diagnoses, on a quarterly basis, for each insured person, with at least one day of insurance coverage, regardless of whether or not they sought medical treatment. All inpatient and outpatient diagnoses are coded in the International Statistical Classification of Diseases and Related Health Problems (ICD), revision 10, issued by the World Health Organization. Dementia was defined as having at least one of the following diagnoses coded by ICD-10: G30, G31.0, G31.82, G23.1, F00, F01, F02, F03, and F05.1. A stroke was defined as having at least one diagnosis of I63 or I64. With interest solely in the inferential comparison of $\lambda_{HD}$ with $\lambda_{S1,D}$, only their risk sets and counts are needed (see (3)). The data contains 49,880 people with onset of dementia in the monitoring period. Point estimates (3) and standard errors (roots of (4)) are given in Table 5. We find that a stroke increases the intensity of suffering from dementia from 0.02 to 0.07. Due to the asymptotic independence of $\hat{\lambda}_{HD}$ and $\hat{\lambda}_{S1,D}$, by Theorem 2.1 it is

$$\text{Var}(\hat{\lambda}_{S1,D} - \hat{\lambda}_{HD}) = \text{Var}(\hat{\lambda}_{S1,D}) + \text{Var}(\hat{\lambda}_{HD})$$

and (see Table 5) estimated to be $0.000000617$. Together with the asymptotic normality of Theorem 2.1, it is approximately

$$\hat{\lambda}_{S1,D} - \hat{\lambda}_{HD} \sim N(\lambda_{S1,D0} - \lambda_{HD0}, 0.000785^2).$$

So that a 95%-confidence interval (CI) of the intensity difference is approximately $[0.492 \pm 0.00154] = [4.8\%, 5.1\%]$. The point estimates’ ratio of 3.5 may exceed recent studies, because we do not adjust for further risk factors of dementia, so far. As one risk factor, we now include age-inhomogeneous intensities as piecewise constant. We will see that the effect of a stroke on dementia onset is markedly smaller because, simultaneously, intensities increase with age and a stroke is more likely at advanced ages.
3.3 Age-inhomogeneous transition from stroke to dementia

In order to account for age-inhomogeneous intensities (also found for the data at hand in Weißbach et al. (2021)), we here define (as in Weißbach et al., 2009; Weißbach and Walter, 2010), with partition $0 = t_0, \ldots, t_b = \tau$, $X$ as a Markov process with intensities $\lambda_{hj}(t) := \sum_{l=1}^{b} 1_{[t_{l-1}, t_l]}(t) \lambda_{hlj}$. Formulae for point estimates and standard errors are given in Appendix C.

Figure 4 and Table 6 exhibit point estimates, estimated standard errors and confidence intervals, using five-year age-groups, i.e. for $b = 13$. For instance, in the age-group with the most events, namely from 80 to 85 years, the intensity after stroke of $\hat{\lambda}_{S^1,D}(t) = 0.086$ exceeds that without stroke of $\hat{\lambda}_{H,D}(t) = 0.049$ (see framed numbers in Table 6). The ratio of 1.8 is now half of the ratio $0.07/0.02 = 3.5$ in Table 5, and more in line with the recent literature. The reason is Simpson’s paradox; in the case for persons with stroke, the age-homogeneous $\hat{\lambda}_{S^1,D} = 0.07$ of Section 3.2 is implicitly an average over a later part of the time span $[0, \tau]$, because a stroke generally occurs at higher ages. The age-homogeneous intensity is hence larger than an average of the age-group intensities over the entire time span, due to the increasing estimates of dementia onset in age. This is not the case for healthy persons’ intensities.
Table 6

Statistics ((2),(3),(5),(6)), point estimates with standard errors (SE) (Formulae: (20) and root of (21)) ((4),(7)) and intensity ratio ((8)) for 5-year age-classes (Exception: oldest class)

<table>
<thead>
<tr>
<th>Age interval</th>
<th># With dementia</th>
<th>Time after stroke</th>
<th>Point Intensity</th>
<th># With dementia</th>
<th>Intensity ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>t0−15</td>
<td>0.0030</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>t15−25</td>
<td>0.0014</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t25−35</td>
<td>0.0015</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t35−45</td>
<td>0.0025</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>t45−55</td>
<td>0.0013</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t55−65</td>
<td>0.0013</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t65−75</td>
<td>0.0013</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t75−85</td>
<td>0.0025</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>t85−95</td>
<td>0.0013</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>t95−105</td>
<td>0.0013</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
</tbody>
</table>

Note: NA = Not available.
In order to interpret the role of age more broadly, note the decreasing stroke-effect in age here (Table 6 or Figure 4). In the age group of the 55 to 60 years old the intensity ratio is 8.5. The higher the age is, the smaller is the intensity ratio. Respective Wald-type tests for pairwise differences (suppressed here) show that there is no significant difference in the risk of dementia between persons with and without stroke for the highest age groups (90 years and older). This coincides with the Framingham Study (Ivan et al., 2002) where the adjusted RR was higher for those younger than 80 (RR=2.6) compared to those aged 80 or older (RR=1.6). Similarly, the systematic review by Savva and Blossom (2010) also does not find an excess risk of dementia after stroke in those at ages 85 years or older. Because age is the major risk factor for dementia, the disadvantage of stroke, including the subsequent excess risk for dementia, is more pronounced in younger and less selected populations.

4 Conclusion

It is encouraging to see that our model, applied to Germany, has similar point estimates as the related study of Kim and Lee (2018), for Korea. That the later study takes more covariates into account clouds the enjoyment over the similarity, because they will typically decrease the effect size. Note that more exogeneous covariates could be integrated algorithmically in our event history analysis, also using e.g. Kim et al. (2012). However, in order to reduce bias, one should probably commence by including left-censored persons, i.e. with dementia prevalent before 2004 (see Weißbach et al., 2021). However, our thrust is more about precision and therefore takes truncation into account. Our precision claim is holstered by assuming three sorts of independency. First of all, within pairs \((T_i, U_i)\), and tests already exist on whether censoring (and with it \(U\)) is independent of the primary measure (see e.g. Sun and Lee, 2011; Tanzer et al., 2021). Second, we assume independence between pairs \((T_i, U_i)\), and generalizing to the conditions (17) and (18) would be an ineffective improvement. More specifically, on the one hand we assume independence between persons in the latent sample (of size \(n_{all}\)), which can be defended in survival analysis with the argument of unforeseeable, staggered entry. And if it holds, selecting a person into the observed sample (of size \(n\)) relies only on a person’s information alone, so that - in contrast to, say, selecting the obviously dependent largest and smallest values - we are confident that the observations are also independent. On the other hand, it would still be sufficient to assume pairs \((T_i, U_i)\) to be independent conditionally on \(A_i\), with no reason as yet. Third, close to longitudinal independence, probably the most critical assumption of our modelling strategy seems to be the Markovian. This is especially the case, as for our application Pendlebury and Rothwell (2009) and Corraini et al. (2017) claim that the time elapsed since stroke is a risk factor for the intensity of dementia onset. Such duration-dependence especially violates the assumption of multiplicative intensities (18) and thus requires a different strategy (see e.g. Weißbach and Schmal, 2019).
Acknowledgment: The financial support from the Deutsche Forschungsgemeinschaft (DFG) of R. Weißbach and G. Doblhammer is gratefully acknowledged (Grant 386913674 ‘Multi-state, multi-time, multi-level analysis of health-related demographic events: Statistical aspects and applications’). For discussions at an earlier stage of the study, we are grateful to O. Gefeller and for valuable suggestions to the editor M. Schmid. For support with the data we thank the AOK Research Institute (WIdO) and for literature research we thank E. Rakusa. The linguistic and idiomatic advice of Brian Bloch is also gratefully acknowledged.

References


A DERIVATION OF CONDITIONAL LIKELIHOOD FOR MORTALITY MODEL

The analysis of the data, simplified to a survival analysis, is the blueprint for all three models and after constructing the filtration we derive the conditional likelihood, the estimator and its consistency and asymptotic normality.

A.1 Filtration in population model

The population comprises Germans born between 1900 and 1954, and we assume insurance by Germany’s largest health insurer AOK to be representative. We consider a simple sample thereof of size \( n_{all} \). The lifetime \( T \) after the 50th birthday, called ‘age-at-death’ in the following, has hazard rate \( \lambda(\cdot) \equiv \lambda \) and CDF \( F(\cdot) \). For a person drawn randomly from the population, the age-at-death \( T \) is defined on the probability space \( (\Omega, \mathcal{F}, P_\lambda) \).

(A1) It is for the true parameter \( \lambda_0 \in \Lambda := [\varepsilon; 1/\varepsilon] \) for some small \( \varepsilon \in \{0; 1\}. \)
In terms of a process, $N_T(t) := 1_{\{T \leq t\}}$ is adapted to $N_t := \sigma\{1_{\{T \leq s\}}, 0 \leq s \leq t\}$. $N_T$ has a compensator with intensity $\alpha(t) := 1_{\{N_T(t-) = 0\}} \lambda = Y(t)\lambda$, with respect to $N_t$ and $P_\lambda$ and with $Y(t) := 1_{\{T \geq t\}}$.

The population model is further described by acknowledging that $T$ is accompanied by the time elapsed since the age of 50 at study begin, $U$ (if random independent of $T$, or $u$ if not, and denoted ‘age at study begin’).

### A.2 Non-random left-truncation and right-censoring (Figure 5)

![Figure 5](image)

**Figure 5** (top: uncensored/untruncated) Path for person born 1/1/1925 with death 1/1/2009, i.e. with $T = 84 - 50, u \leq T \leq u + 9$

(middle: left-truncated) Path for person born 1/1/1915 with death 1/1/2000, i.e. with $T = 85 - 50, T < u, u = 89 - 50$

(bottom: right-censored) Path for person born 1/1/1915 with death 1/1/2015, i.e. with $T = 100 - 50, T \geq u + 9, u + 9 = 98 - 50$

In our study, we begin to monitor at 1/1/2004 and by left-truncation, information deteriorates even more than by left-censoring, where only the age at an event gets lost for concerned persons (see e.g. Kremer et al., 2014; Weißbach et al., 2021). Here also, the age-at-censoring gets lost, and even the existence of a truncated person,

As a preparation (and as in Andersen et al., 1993, Example III.3.6), we start with deterministic $u$. Increasing the filtration (as will be necessary for random truncation), is not necessary for deterministic (and hence universally known) $u$. The (non) left-truncation event is $A := \{T > u\}$. We assume $P_\lambda(A) > 0$. The observable filtration is now

$$uG_t := \sigma\{1_{\{u \leq T \leq s\}}, u \leq s \leq t\} = \sigma\{uN_T(s), u \leq s \leq t\}$$
with \( u N_T(t) := N_T(t) - N_T(\min(t, u)) \), i.e. \( u N_T \) is \( N_T \) if \( u \leq T \) and zero otherwise. The intensity process of \( u N_T(t) \) is \( u \alpha(t) := 1_{\{u < t \leq T\}} \lambda \), with respect to the probability measure

\[
P^A(F) := \frac{P_A(F \cap A)}{P_A(A)} \quad \text{for} \quad F \in \mathcal{F}.
\]

The deterministic \( u \) is a special case of the random \( U \), the Example 4.1 in Andersen et al. (1988).

Right-censored is the age-at-death if it occurs after 2013, or after having left the AOK (see Figure 1). As in (Andersen et al., 1993, Examples III.3.6, IV.1.7 and VI.1.4) we superimpose right-censoring on left-truncation (and not vice versa). The observed left-truncated and right-censored counting process is

\[
u N^c_T(t) := \int_0^t C(s) d \nu N_T(s) = 1_{\{u \leq t \leq \min(t, u+9)\}}
\]

with \( C(t) := 1_{\{t \leq u+9\}} \) (compare Fleming and Harrington, 1991, Example 1.4.2). It has intensity \( \nu \alpha^c_T(t) := \lambda 1_{\{u \leq t \leq \min(T, u+9)\}} \) with respect to \( P^A \) and observed filtration

\[
u F^c_T := \sigma\{\nu N^c_T(s), u \leq s \leq t\} = \sigma\{1_{\{u \leq t \leq \min(s, u+9)\}}, u \leq s \leq t\}.
\]

The \( \nu F^c_T \) is self-exciting (see Andersen et al., 1988, p4), as required (see Andersen et al., 1988, p23), so that the likelihood is determined by Jacod’s formula (see Andersen et al., 1988, Formula (4.3)), (see also Andersen et al., 1993, Formula (2.7.2’)) and (in extension 3.2.8):

\[
dP = \prod_{u < t \leq u+9} \left\{ (1 - u \alpha^c_T(t)dt)^{1 - d} \nu N^c_T(t) \right\} d\nu N^c_T(t)
\]

\[
= \begin{cases} 
1^{1 \cdot 0^0} = 1 \quad & \text{for } T < u \\
\lambda(T)e^{-\int_u^T \lambda(t)dt} = \lambda e^{-\lambda(T-u)} \quad & \text{for } u \leq T \leq u + 9 \\
e^{-\int_u^{u+9} \lambda(s)ds} = e^{-9\lambda} \quad & \text{for } u + 9 < T
\end{cases}
\]

(6)

For an explanation of the first and third case, see page 24 and, respectively, Example 2.2 in Andersen et al. (1988). Note that in the second case \( d\nu N_T(t) = 0 \) and \( 1_{\{u \leq T \leq \min(t, u+9)\}} = 0 \).

Independent persons are, due to different \( u_i \), not identically distributed. Each likelihood is a Radon-Nikodym derivative with respect to a different measure, all of which are even dependent on the parameter, \( P^A \). Even worse, measures are conditional on observation. The usual formula from introductory statistics that the density, for a collection of independent persons is the product of the persons’ densities relies on the equal (and parameter-independent) dominating measure (usually being Lebesgues). Unconditionally, such fixed covariate regression could be analysed as in Weißbach and Radloff (2020). To achieve equal dominating measures for all persons, we follow Examples IV.1.7 and VI.1.4 of Andersen et al. (1993) in using a random \( U \). (At the end, we will draw a simple sample of size \( n_{All} \) from the population and begin to monitor on 1/1/2004.)
A.3 Random left-truncation and conditional right-censoring

The probability space for \((T, U)\) is \((\Omega, \mathcal{F}, \tilde{P}_\lambda)\), where the distribution of \(U\) will not be important and we suppress its parameter (and indicate the difference with the tilde instead).

\[(A2)\] \(U\) and \(T\) are independent and \(\beta_{\lambda_0} := \tilde{P}_\lambda(\{T > U\}) > 0\) with redefined \(A := \{T > U\}\).

A.3.1 Left-truncation for age-at-death (Figure 6)

Ignoring censoring for the moment, \(T\) is recorded when larger than \(U\), the age-at-study-begin (see Figure 6). Consider the unobservable filtration \(\mathcal{G}_t := \sigma\{1_{\{T \leq s\}}, 1_{\{U \leq s\}}, 0 \leq s \leq t\}\). For left-truncation

\[
\begin{align*}
T + 50 &= 85 \\
U + 50 &= 89 \\
T + 50 &= 84
\end{align*}
\]

\(1/1/1900\) \hspace{1cm} \(1/1/1910\) \hspace{1cm} \(1/1/1920\) \hspace{1cm} \(1/1/1930\) \hspace{1cm} \(1/1/1940\) \hspace{1cm} \(1/1/1950\) \hspace{1cm} \(1/1/1955\) \hspace{1cm} \(1/1/2004\)

**Figure 6** (top: left-censored/truncated) Path for person born 1/1/1915 with death 1/1/2000, i.e. with \(T = 85 - 50 = 35, T < U, U = 89 - 50 = 39\)

(bottom: uncensored/truncated) Path for person born 1/1/1925 with death 1/1/2009, i.e. with \(T = 84 - 50 = 34, T \geq U\) (see Andersen et al., 1993, Example III.3.2), similar to non-random left-truncation, define \(U N_T(t) := N_T(t) - N_T(t \land U)\) with \(\mathcal{G}_t\)-intensity, with respect to \(\tilde{P}_\lambda\), again being \(\lambda 1_{\{T \geq t\}}\) (because \(U\) is independent of \(T\)). We observe durations \(U\) and \(T\) in the case of \(A\), and neither measurement \(U\) nor \(T\) - nor the person at all - when \(T < U\). Now define \(U Y(t) := Y(t) 1_{\{t > U\}} = 1_{\{T \geq t > U\}}\), and as \(\mathcal{G}_t\) is unobservable, but due to \(U\) being a \(\mathcal{G}_t\)-stopping time, \(U F_t := \{U N_T(s), U Y(s), U \leq s \leq t\}\) is observable. The intensity process of \(U N_T(t)\) is \(U \alpha_T(t) = 1_{\{U < t \leq T\}} \lambda\), with respect to \(U F_t\) and the probability measure

\[
\tilde{P}_\lambda^A(F) := \frac{\tilde{P}_\lambda(F \cap A)}{\tilde{P}_\lambda(A)} \quad \text{for} \quad F \in \mathcal{F}.
\]

Intuitively, \(U N_T\) will not yet jump prior to \(U\), and no longer after \(T\), and in between, at the hazard rate of \(T\). Now, as \(U F_t\) is self-exciting, we may apply Jacod’s formula (see Andersen et al., 1988, Formula (2.1))
in order to determine the conditional likelihood (see Andersen et al., 1988, Formula (4.3)). And as $U$ is independent of $T$ and the hazard rate has the form $\lambda(\cdot) \equiv \lambda$, the conditional likelihood is for $T \leq U$ one and else (see also Andersen et al., 1993, Formula (3.3.3)):

$$
UL(\lambda) = \prod_{t > U} \{ \mathbb{U} \alpha_T(t)^{\Delta U N_T(t)}(1 - \mathbb{U} \alpha_T(t)dt)^{1-\Delta U N_T(t)} \}
$$

$$
= \lambda(T)e^{-\int_T^U \lambda(t)dt} = \frac{[1 - F(T)]\lambda(T)}{1 - F(U)} = \lambda e^{-\lambda(T-U)}
$$

(A.3.2) Acknowledging non-random censoring

First note that now $T > U + 9$ is known, conditionally on $A$, because $U$ is observable, rendering unnecessary a further increase in filtration. Hence, similar to A.2, we apply Definition III.2.1 in Andersen et al. (1993), with $U N_T$ in the role of $N$, $(U, F)$ in the role of $(F_t)$, $U \alpha_T$ in the role of $\lambda$ and $\tilde{P}_\lambda$ in the role of $P_{\phi \psi}$. Being conditionally deterministic, $C(t) := \mathbb{1}_{\{t \leq U + 9\}}$ is independent and predictable. We define

$$
U Y^c(t) := C(t) U Y(t) = \mathbb{1}_{\{U \leq t < \min(T, U + 9)\}}
$$

so that

$$
U N^c_T(t) := \int_U^t C(s) U N_T(s) = \mathbb{1}_{\{U \leq T \leq \min(t, U + 9)\}}
$$

has intensity $U \alpha^c_T(t) = \mathbb{1}_{\{U < t \leq \min(T, U + 9)\}} \lambda$, with respect to the observable filtration

$$
U F^c_t := \{ U N^c_T(s), U Y^c(s+), U \leq s \leq t \}
$$

and conditional distribution $\tilde{P}_\lambda$ (see Andersen et al., 1993, Section III.2.2). By Formula (3.2.8) in Andersen et al. (1993), the partial (and here only conditional) likelihood is for $T \leq U$ again one and otherwise:

$$
UL^c_T(\lambda) = \prod_{t > U} \{ (U \alpha^c_T(t))^{\Delta U N^c_T(t)}(1 - U \alpha^c_T(t)dt)^{1-\Delta U N^c_T(t)} \}
$$

$$
= \begin{cases} 
\lambda(T)e^{-\int_T^U \lambda(t)dt} = \lambda e^{-\lambda(T-U)} & \text{for } U \leq T \leq U + 9 < \tau \\
\lambda e^{-\lambda(T-U)} & \text{for } U + 9 < T < \tau 
\end{cases}
$$

(A.4) Sample result

We now switch from one person to the simple sample of unobserved $n_{all}$. Now, the size of the observed sample persons surviving 31/12/2003 is $n := \sum_{i=1}^{n_{all}} 1_{A_i}$, where $i$ enumerates the persons (see Figure 1). Again, without loss of generality, we have sorted those not truncated at the beginning of the sample, and all other summands are zero. Due to the assumption of a simple sample (and hence the same dominating measure), the log conditional likelihood for the data is the sum of the log conditional likelihood contributions for each observed person. We denote by $N_{cens} := \sum_{i=1}^{n} 1_{\{T_i > U_i + 9\}}$ the number of right-censored and by
\( N_{\text{uncens}} := n - N_{\text{cens}} = \sum_{i=1}^{n} \mathbb{I}_{\{U_i < T_i \leq U_i + 9\}} \) the number of neither truncated nor censored people and can write \( \log U L_C(\lambda) \) as
\[
\sum_{i=1}^{n} \log U L_C(\lambda)_i = N_{\text{uncens}} \log(\lambda) - \lambda \sum_{i=1}^{n} \mathbb{I}_{\{U_i \leq T_i \leq U_i + 9\}} (T_i - U_i) - 9\lambda N_{\text{cens}},
\] (9)
so that the unique estimator becomes
\[
\hat{\lambda} = \arg\max_{\lambda \in \Lambda} \log U L_C(\lambda) = \frac{N_{\text{uncens}}}{\sum_{i=1}^{n} \mathbb{I}_{\{U_i \leq T_i < U_i + 9\}} (T_i - U_i) + 9N_{\text{cens}}}. \tag{10}
\]

Note that due to the simple sample, among those who survive \( U_i \) (i.e. 2004), for the portion in the study period alive at the age \( t \) in observed sample, it is, with
\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{U_i, t \leq \min(T_i, U_i + 9)\}} \xrightarrow{P} \tilde{P}_\lambda(U < t \leq \min(T, U + 9)|A), \tag{11}
\]
the latter is that portion of the population, and analogue to \( m_h(t) \). Again, for our parametric model, we only assume:

(A3) \( \int_{0}^{\tau} \tilde{P}_\lambda(U < t \leq \min(T, U + 9)|A) dt > 0 \)

**Lemma A.1** Under Conditions (A1)-(A3) it is \( \hat{\lambda} \) of (10) consistent and \( \sqrt{n_{\text{all}}}(\hat{\lambda} - \lambda_0) \xrightarrow{D} N(0, \sigma^{-1}(\lambda_0)) \) with
\[
\sigma(\lambda_0) := \frac{\beta_{\lambda_0}}{\lambda_0} \int_{0}^{\tau} \tilde{P}_\lambda(U < t \leq \min(T, U + 9)|T > U) dt.
\]

**Proof:** Due to the uniqueness of \( \hat{\lambda} \), for both consistency and asymptotic normality, we need to verify Conditions (A1)-(E) (see Andersen et al., 1993, Theorems VI.1.1+2). In order to map the notations, note that \( n \) becomes \( n_{\text{all}} \), \( a_n := \sqrt{n_{\text{all}}} \), \( \theta \) becomes \( \lambda \) and \( \beta(t) \) is redundant. The \( \lambda(t; \theta) \) becomes, by interchanging conditional expectations with summation, the multiplicative intensity process of \( U N_{T \bullet}(t) := \sum_{i=1}^{n} U N_{T_i}(t) \) (where \( i \) enumerates the persons):
\[
U \alpha_{\bullet}(t; \lambda) = \lambda \sum_{i=1}^{n} \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}}
\]
By (A1), (A) is fulfilled for intensity \( U \alpha_{\bullet}(t) \) and therefore for \( U \alpha_{T \bullet}(t; \lambda) \) and the logarithm of likelihood (8). For (B), because by the LLN, for fixed \( t \in [0, \tau] \), (11) and \( n/n_{\text{all}} \xrightarrow{P} \tilde{P}_\lambda(A) = \beta_{\lambda_0} \), so that Slutzky’s Lemma and the CMT yield
\[
\frac{1}{n_{\text{all}}} \int_{0}^{\tau} \left( \frac{d}{d\lambda} (\log \lambda) \right)^2 \big|_{\lambda = \lambda_0} \int_{0}^{n} \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}} dt
\]
\[
= \frac{n}{n_{\text{all}}} \sum_{i=1}^{n} \int_{0}^{\tau} \frac{1}{\lambda_0} \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}} dt \to \sigma(\lambda_0).
\]
For (C), because (i) \( n \leq n_{\text{all}} \), (ii) \( 1/(n\lambda_0) \rightarrow P \rightarrow 0 \) by (A1) and \( n \rightarrow P \rightarrow \infty \) (due to \( n \) following a Binomial distribution with parameters \( n_{\text{all}} \) and the selection probability \( \beta \lambda_0 \)) and (iii) (A3), we have

\[
\frac{n}{n_{\text{all}}} \int_0^\tau \left( \frac{d}{d\lambda} \log(\lambda) \right)_{\lambda=\lambda_0}^2 \mathbb{I}_{\{1/(n\lambda_0) > \varepsilon\}} \lambda_0 \sum_{i=1}^n \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}} dt = \frac{n}{n_{\text{all}}} \int_0^\tau \frac{1}{n_{\text{all}}} \lambda_0 \int_0^\tau \frac{1}{n_{\text{all}}} \lambda_0 \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}} dt \rightarrow 0. \quad (12)
\]

For (D), by (A3), it is \( \sigma(\lambda_0) \) positive. Condition (E) consists of six conditions: For the first four, note that

\[
\sup_{\lambda} \left| \frac{\partial^3}{d\lambda^3} \left( \lambda \sum_{i=1}^n \mathbb{I}_{\{U_i < t \leq \min(T_i, U_i + 9)\}} \right) \right| = 0,
\]

\[
\sup_{\lambda} \left| \frac{\partial^3}{d\lambda^3} \log(\lambda) \right| = \sup_{\lambda} |\lambda^{-3/2}| =: H < \infty \quad \text{and}
\]

\[
\frac{1}{n_{\text{all}}} \int_0^\tau H dt \rightarrow 0.
\]

For the fifth, note (12) (and also (A1)). For the sixth, note (12) and then \( H/\sqrt{n_{\text{all}}} \rightarrow 0 \).

It remains to consistently estimate \( \sigma(\lambda_0) \), in order to construct a confidence interval:

\[
-J_{\tau}(\lambda_0) := \frac{d^2}{d\lambda^2} \log U L_{\tau}(\lambda)_{\lambda=\lambda_0} = -\frac{N_{\text{uncens}}}{\lambda_0^2} \quad \text{(Andersen et al., 1993, Formula (6.1.11))}
\]

\[
\frac{1}{n_{\text{all}}} \int_0^\tau \frac{d^2}{d\lambda^2} \log U L_{\tau}(\lambda)_{\lambda=\lambda_0} \rightarrow \sigma(\lambda_0) \quad \text{and} \quad n_{\text{all}} J_{\tau}(\lambda_0) \rightarrow \sigma^{-1}(\lambda_0)
\]

\[
\Rightarrow J_{\tau}^{-1}(\lambda_0) = \frac{\lambda_0^2}{N_{\text{uncens}}} \quad \text{and} \quad n_{\text{all}} J_{\tau}^{-1}(\lambda_0) \equiv \sigma^{-1}(\lambda_0)
\]

\[
\Rightarrow Var(\hat{\lambda}) = J_{\tau}^{-1}(\lambda) \quad \text{(by (10) and CMT as \( \hat{\lambda} \) is consistent)}
\]

\[
= \frac{N_{\text{uncens}}}{(\sum_{i=1}^n \mathbb{I}_{\{T_i \geq U_i\}} (T_i - U_i) + 9N_{\text{cens}})^2}
\]

(13)

Note that by Lemma A.1, \( Var(\hat{\lambda}) = Var(\sqrt{n_{\text{all}}} \hat{\lambda})/n_{\text{all}} = \sigma^{-1}(\lambda_0)/n_{\text{all}} \) depends on the unobservable \( n_{\text{all}} \). However, the estimated standard error of \( \hat{\lambda} \) (SE), as the square root of (13), is observable.

### B Derivation of Conditional Likelihood for Age-Homogeneous Morbidity Model

The analysis of the model, simplified to age-homogeneity is the centre of the paper, and the important arguments of the survival analysis in Appendix A are repeated, when differences need to be accounted for. Especially, we state the filtration, derive the conditional likelihood and the estimator, including consistency and asymptotic normality.
B.1 Filtration, left-truncation, right-censoring

In the survival analysis we initially simplify to a non-random age-at-study-begin, \( u \), so that the measurement of one sample person becomes a univariate random variable. Now, a unit is a history \( X \) which we reformulate in the counting processes

\[
\mathbf{N}_{hj}(t) := \sum_{s \leq t} 1\{X(s) = h, X(s) = j\}
\]

and

\[
Y_{h}(t) := 1\{X(t) = h\}.
\]

The vector \( \mathbf{N}_{X}(t) := (\mathbf{N}_{hj}(t), hj \in \mathcal{I})' \) is adapted to the filtration \( \mathcal{N}_t := \sigma\{\mathbf{N}_{X}(s), 0 \leq s \leq t\} \), because we assume \( X_{0} = H \). We assume for the true parameter \( \lambda_{0} \):

(B1) It is \( \lambda_{hj}^{0} \in \Lambda_{hj} := [\varepsilon_{hj}; 1/\varepsilon_{hj}] \) for small \( \varepsilon_{hj} \in ]0, 1[ \).

The compensator of \( \mathbf{N}_{X} \) has an intensity with respect to \( \mathcal{N}_{\lambda} \) and \( P_{\lambda}(\alpha(t) := (Y_{h}(t)\lambda_{hj}, hj \in \mathcal{I})' \).

A person is not left-truncated when surviving 2003, i.e. is observed in case of \( A := \{X_{u} \neq d\} \). We assume \( P_{\lambda}(A) > 0 \), and the observable filtration is

\[
u_{\mathcal{G}_{t}} := \sigma\{u\mathbf{N}(s), u \leq s \leq t\}
\]

with

\[
u_{\mathbf{N}}(t) := \mathbf{N}_{X}(t) - \mathbf{N}_{X}(t \land u),
\]

where \( t \land u := \min(t, u) \).

**Lemma B.1** With respect to the probability measure \( P_{\lambda}^{A}(F) := P_{\lambda}(F \cap A)/P_{\lambda}(A) \) for \( F \in \mathcal{F} \), the intensity process of \( u\mathbf{N}(t) \) is \( u\alpha(t) := 1_{\{u < t\}}\alpha(t) \).

The deterministic \( u \) is a special case of the random \( U \), so that the proof of the latter can be found in Proposition 4.1 and Example 4.3 of Andersen et al. (1988), for details see A.1.

Note that \( P_{\lambda}^{u} \) depends on the parameter \( \lambda \) and on \( u \). With \( u\gamma_{h}(t) := 1_{\{u < t\}}Y_{h}(t) \), the coordinates of \( u\alpha \) are \( u\alpha_{hj}(t) := 1_{\{u < t\}}\alpha_{hj}(t) = u\gamma(t)\lambda_{hj} \). The observed left-truncated and right-censored counting process is \( u\mathbf{N}^{c}(t) := \int_{0}^{t} C(s)d\mathbf{uN}(s) \), with \( C(t) := 1_{\{t \leq u + 9\}} \) and \( u\mathbf{Y}^{c}(t) := C(t)u\mathbf{Y}(t) \) (compare Fleming and Harrington, 1991, Example 1.4.2). It has intensity

\[
u\alpha^{c}(t) := 1_{\{t \leq u + 9\}} u\alpha(t) = 1_{\{u < t \leq u + 9\}} \alpha(t)
\]

with respect to \( P_{\lambda}^{A} \) and observed filtration \( u\mathcal{F}^{c}_{t} := \sigma\{u\mathbf{N}^{c}(s), u \leq s \leq t\} \).
As \( uF^c_t \) is a required self-exciting filtration, by Jacod’s formula (see Andersen et al., 1988, Formula (4.3)), likelihood and log-likelihood are:

\[
dP = \prod_{u < t \leq u + 9} \{ (1 - u\alpha^c(t))dt \}^{1 - d} \prod_{h \in \{H,S,D\}} \prod_{j \neq h} \{ (u\alpha^c_{hj}(t))dt \} \prod_{j \neq h} \{ (uN^c_{hj}(t))dt \}
\]

\[
= \left[ \prod_{t \in [u,u+9], \exists hj \in \mathcal{I}} (uN^c_{hj}(t)\lambda_{hj})^{\Delta N^c_{hj}(t)} \right] \exp \left( -\sum_{hj \in \mathcal{I}} \int_{u}^{u+9} uY^c_{hj}(t)\lambda_{hj} dt \right)
\]

\[
\log dP = \int_0^\tau \sum_{hj \in \mathcal{I}} (\log uY^c_{hj}(t) + \log \lambda_{hj}) d\left( uN^c_{hj}(t) \right) - \sum_{hj \in \mathcal{I}} \int_0^\tau uY^c_{hj}(t)\lambda_{hj} dt \tag{15}
\]

Note \( u\alpha^c(t) = \sum_{hj \in \mathcal{I}} u\alpha^c_{hj}(t) \) together with \( uN^c(t) := \sum_{hj \in \mathcal{I}} uN^c_{hj}(t) \). The double-use of the integration symbol \( dt \) in the first line is explained in (7). For the second line, see e.g. Formula (11) in Weißbach et al. (2009): For whatever process \( \mathbf{Z} \), we define \( d\mathbf{Z}(t) := \mathbf{Z}(t) - \mathbf{Z}(t - dt) \) for some ‘small’ \( dt \) (see Fleming and Harrington, 1991, Sect. 1.4) and \( \Delta \mathbf{Z}(t) := \mathbf{Z}(t) - \mathbf{Z}(t -) \) (see Andersen et al., 1993, Sect. II.2), hence \( \Delta uN^c_{hj}(t) = uN^c_{hj}(t) - uN^c_{hj}(t-) \) is only not zero if \( uN^c_{hj}(t) \) jumps. These jumps are of height one, since the transition times are continuous. Further note that \( \int_u^{u+9} \) can be replaced by \( \int_0^\tau \), because \( uY^c_{hj}(t) \) already accounts for the limits, and similarly, in the product, \([u, u + 9] \) is accounted for in \( uN^c(t) \). Note that, because \( uN^c_{hj}(u) = 0 \),

\[
\frac{\partial}{\partial \lambda_{hj}} \log dP = \frac{uN^c_{hj}(\tau)}{\lambda_{hj}} - \int_0^\tau uY^c_{hj}(t) dt. \tag{16}
\]

However, we cannot assume \( u \) to be non-random. If we assume \( \mathbf{X} \) to be random, and with it the age-at-death, we should also assume the age-at-study-begin to be random, i.e. \( U \). Increasing the filtration to a finer level has not been necessary so far, for non-random (and universally known) \( u \). We impose as additional assumption, that not everybody is dead, prior to 2004:

(B2) \( \beta_{x_0} := P(X_U \neq d) > 0 \)

The additional information by \( U \), i.e. for \( (\mathbf{X}, U) \), and at the same time the loss in information by truncation, is reflected by including \( u\mathbf{Y}^c \) in the filtration \( uF^c_t, uF^c_t := \sigma \{ U, N^c(s), u\mathbf{Y}^c(s), u \leq s \leq t \} \), for details see A.2. Consult A.3 to see that, similar to non-random truncation (15), conditional on the last two coordinates, the logarithmic density of \( (X, U, \mathbb{1}(X_U \neq d)) \) up to \( \tau \) is (1) where \( U \) replaces \( u \) in the definitions of \( uN^c_{hj}(t) \), \( uY^c_{hj}(t) \) and \( u\alpha^c(t) \) of (14).
B.2 Sample result

Again, so far, we studied one person of the sample, truncated or not. Now we consider all persons in the simple sample (of size \( n_{\text{all}} \)). As in Appendix A, the size of the observed sample is \( n := \sum_{i=1}^{n_{\text{all}}} 1_{A_i} \), and we sort the truncated to the end of the unobserved sample. Their summands or factors in the sequel will then be ineffective. Again, due to the simple sample assumption and \( U_i \) being random, the conditional density of the sample is the product of the persons’ conditional densities (see A.3), (2). Due to (16), and in spite of (1), \( U_{N_{h_j}^c}(t) := \sum_{i=1}^{n_{\text{all}}} U_{N_{h_j}^c}(t) \) and \( U_{Y_{h}^c}(t) := \sum_{i=1}^{n_{\text{all}}} U_{Y_{h}^c}(t) \) suffice for representing the point estimator, because of \( \frac{\partial}{\partial \lambda} \log U_{L^c}(\text{data}|\lambda) \) the unique root is (3).

Note that by using the simple sample assumption, among those who survive \( U_i \) (i.e. 2003), the portion in the study period at age \( t \) in state \( h \) is asymptotically the same in the observed sample and in the entire population. By the LLN, for fixed \( t \),

\[
\frac{U_{Y_{h}^c}(t)}{n} \overset{P}{\longrightarrow} m_h(t).
\]

The latter will typically be positive, but for our parametric model, we only need to assume:

(B3) \( \int_0^\tau \sum_{h \in \{H,S,D\}} m_h(t) dt > 0 \)

As stated in the body of the text, assumptions (B1)-(B3) render the estimator asymptotically normal.

**Proof of Theorem 2.1:** Due to the uniqueness of \( \hat{\lambda} \), for both consistency (see Andersen et al., 1993, Theorem VI.1.1) and asymptotic normality (see Andersen et al., 1993, Theorem VI.1.2), we need to verify Conditions (A)-(E) (see Andersen et al., 1993, Condition VI.1.1). Specifically, in order to map the notations, note that \( n \) becomes \( n_{\text{all}} \), \( a_n := \sqrt{n_{\text{all}}} \), \( \theta \) becomes \( \lambda \), \( h \) becomes \( h_j \) and \( \lambda_h(t; \theta) \) becomes

\[
U_{\alpha_{h_j}^c}(t; \lambda) = \lambda_{h_j} U_{Y_{h}^c}(t), \ h_j \in \mathcal{I},
\]

because by interchanging conditional expectations, the multiplicative intensity process of \( U_{N_{h_j}^c}(t) \) is the sum of the compensators for each person’s counting process.

The fulfilment of (A) is now as in Lemma A.1 of A.4. For (B), note first that for mixed derivatives, terms are only non-zero when \( h_{1,j_1} = h_{2,j_2} = h_{3,j_3} \) due to (18). In that case,

\[
\frac{1}{n_{\text{all}}} \int_0^\tau \sum_{h_j \in \mathcal{I}} \left( \frac{\partial}{\partial \lambda_{h_j}} \log U_{\alpha_{h_j}^c}(t; \lambda_0) \right)^2 U_{\alpha_{h_j}^c}(t; \lambda_0) dt = \frac{n}{n_{\text{all}} n} \int_0^\tau \sum_{h_j} \frac{1}{\lambda_{h_j}^0} \lambda_{h_j} U_{Y_{h}^c}(t) dt \to \sigma_{h_j,h_j}(\lambda_0),
\]
because, by the LLN, for fixed $t \in [0, \tau]$, (including Slutzky's Lemma, the CMT, and (17) (as in Lemma A.1)) it is $n/n_{all} \xrightarrow{P} P(X_U \neq d) = \beta \lambda_0 > 0$ by (B2). For (C),

$$
\frac{1}{n_{all}} \int_0^\tau \sum_{h_j \in \mathbb{I}} \left( \frac{\partial}{\partial \lambda_{h_j}} \log U \alpha^{h_j}_{\bullet}(t; \lambda_0) \right)^2 \left\{ \frac{\partial}{\partial \lambda_{h_j}} \log U \alpha^{h_j}_{\bullet}(t; \lambda_0) \right\} \lambda^{h_j}_{\bullet}(t; \lambda_0) dt
$$

$$
= \frac{1}{n_{all}} \int_0^\tau \sum_{h_j \in \mathbb{I}} \left( \frac{\partial}{\partial \lambda_{h_j}} \left( \frac{\partial}{\partial \lambda_{h_j}} \log U \alpha^{h_j}_{\bullet}(t; \lambda_0) \right) \right) \lambda^{h_j}_{\bullet}(t; \lambda_0) dt
$$

because $1/\sqrt{n_{all}} \rightarrow 0$ and $\lambda_{h_j} > 0$ due to (B1). For (D), $\Sigma(\lambda_0)$ is diagonal with $\sigma_{h_j, h_j}(\lambda_0) > 0$ due to (B2) and (B3). Now in (E), generally (first line) and if $h_1 j_1 = h_2 j_2 = h_3 j_3$ (second line), it is:

$$
\sup_{\lambda \in \mathbb{A}} \left| \frac{\partial^3}{\partial \lambda_{h_1 j_1} \partial \lambda_{h_2 j_2} \partial \lambda_{h_3 j_3}} \lambda^{h_j}_{\bullet} U \alpha^{h_j}_{\bullet}(t) \right| = 0
$$

and

$$
\sup_{\lambda \in \mathbb{A}} \left| \frac{\partial^3}{\partial \lambda_{h_1 j_1} \partial \lambda_{h_2 j_2} \partial \lambda_{h_3 j_3}} \log(1/\lambda^{h_j}_{\bullet}(t; \lambda)) \right| = \sup_{\lambda_{h_j} \in \lambda_{h_j}} \frac{1}{2\lambda_{h_j}^3}
$$

Now $1/(2\lambda_{h_j}^3) \leq 1/(2\varepsilon_{h_j}^3) < \infty$ by (B1). (This explains the third and forth requirement.) For the fifth,

$$
\frac{1}{n_{all}} \int_0^\tau \sum_{h_j \in \mathbb{I}} \left( \frac{\partial^2}{\partial \lambda_{h_j}^2} \log U \alpha^{h_j}_{\bullet}(t; \lambda_0) \right)^2 \lambda^{h_j}_{\bullet}(t; \lambda_0) dt
$$

$$
= \frac{n}{n_{all} n} \int_0^\tau \sum_{(h_j, h_j)} \lambda^{h_j}_{\bullet}(t; \lambda_0) \lambda^{h_j,0}_{\bullet}(t) dt - \sigma_{h_j, h_j}(\lambda_0)^{-1}
$$

The latter are finite due to (B1) and (B3). The sixth is fulfilled with the same argument as in (C).

\section{Derivation of Conditional Likelihood for Age-Inhomogeneous Morbidity Model}

The model that suits most the applied view is age-inhomogeneous. We do give neither the very granular analysis of the survival analysis of Appendix A, nor the still complete analysis of the age-homogeneous model in Appendix B. We restrict here to the state of the conditional likelihood and derive the estimator and its standard errors. Again, the two counting processes $N_{h_j}(t)$ and $Y_{h}(t)$ reformulate a history. When stacking $\lambda_{h_j}(t)$ to $\lambda(t)$ in the same way as $N_{h_j}$ to $\mathbb{N}_X$, $\mathbb{N}_X$ has a compensator - with respect to $\mathbb{N}_I$ - with intensity $\alpha(t) := (Y_{h}(t)\lambda_{h_j}(t), h_j \in \mathbb{I})$. The compensator is with respect to the probability measure $P_h$, where $\lambda := (\lambda_{h_{S+1}}, \ldots, \lambda_{D_{db}})'$ collects the $6b$ parameters. Hence with little change, compared to (1):

$$
\log U \mathcal{L}^c(X| \lambda) = \int_U^{U+9} \sum_{h_j \in \mathbb{I}} \left( \log U Y_h(t) + \log \lambda_{h_j}(t) \right) dU N_{h_j}(t)
$$

\begin{equation}
- \sum_{h_j \in \mathbb{I}} \int_U^{U+9} U Y_h(t) \lambda_{h_j}(t) dt
\end{equation} (19)
Again, the Hessian of the log conditional likelihood is a diagonal matrix with diagonal elements

$$-\mathcal{J}_r(\lambda_0)_{h_{ij},h_{ij}} = \frac{\partial^2}{\partial \lambda_{h_{ij}}^2} \log L^c(data|\lambda)|_{\lambda=\lambda_0} = -\frac{A_t}{\lambda_{h_{ij}}^2},$$

and hence

$$\text{Var}(\hat{\lambda}_{h_{ij}}) = A_t / B_t^2.$$