Stockholms universitet

Confidence Intervals in Relative Survival Analysis

Oskar Gauffin

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2017:6
Matematisk statistik
Maj 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2017:6**
http://www.math.su.se

Confidence Intervals in Relative Survival Analysis

Oskar Gauffin*

May 2017

Abstract

Relative survival analysis measures the survival of individuals subject to a specific disease, relative to the expected survival of these individuals had they not been subjects to the disease. Variance of relative survival estimates is commonly calculated based on an assumption of negligible variance in the expected survival estimates, i.e. uncertainty in the population mortality rates is ignored. In this thesis we examine the impact of including this uncertainty for three estimators of relative survival. From bootstrap simulations we find that the confidence interval width for the 10 year Pohar Perme estimate increases with 16 % after including population mortality uncertainty. For age standardized Ederer II and Flexible Parametric Models, the changes in confidence interval width were found to be small. The uncertainty of population mortality is related to the size of the population, and we also investigate the impact of decreasing the original 5 million population to a size of 2.5 and 0.5 million. The results indicate that confidence interval widths for the smaller population sizes does not differ much compared to the original one.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: oskar.gauffin@hotmail.com. Supervisor: Mathias Lindholm.

Acknowledgements

I want to thank my supervisors at Karolinska Institutet (KI), Therese Andersson and Hannah Bower for giving me the opportunity to write my master thesis at KI, their warm support and for sharing their deep knowledge of relative survival analysis. I would also like to thank everyone else at KI who made the writing of this thesis enjoyable. Finally, I would like to thank my supervisor at Stockholm University, Mathias Lindholm, for good discussions, feedback and fresh perspectives.

Contents

1 Introduction

1.1 Overview

When evaluating healthcare a common question is how large the survival of patients with a specific disease is for some given time. To answer this seemingly simple question is not straightforward. For the medical doctor, it is not always a simple task to determine the cause of death, when a deceased patient had several potentially life ending diseases. And how should accidents, related to the specific disease be classified? Relative survival estimates compare all observed deaths of the diseased subpopulation, to what we would have expected had the patients not been subject to the disease. The expected survival is based on the overall population mortality, and by including this in the estimation, relative survival avoids the difficulties with determining the cause of death.

Once the estimate has been calculated, other questions might be of interest. How sure are we of the actual estimate? The uncertainty is usually measured by constructing confidence intervals. As relative survival is often calculated using the entire diseased population of a country, the need for confidence intervals has been discussed in Läkartidningen. There, Dal & Andersson (2004) argue for not constructing confidence intervals, as the entire target population is observed. Dickman, Palmgren & Pawitan (2004) suggest that although this is true, the observed outcomes should be thought of as an observation of an underlying random process. They illustrate their reasoning with the incidence of breast cancer cases in Sweden 2002, which was found to be 6623 cases. If year 2002 could be observed repeatedly, the authors argue that the counts would not be exactly the same due to underlying randomness in the causes of cancer. Taking the sample average of these repeated observations would instead estimate the true underlying process. In this thesis we will proceed from an underlying random process point of view.

To provide accurate confidence intervals of relative survival estimates is important. When deciding whether the relative survival of two subgroups differ, or assessing whether cancer survival has changed with new cancer treatment, confidence intervals play an important in distinguishing random fluctuations from statistically significant differences. The aim of this thesis is to investigate the impact of not making a common assumption when creating confidence intervals, i.e. whether uncertainty in the population mortality rates is negligible.

1.2 Outline

We will begin with introducing some basic concepts of survival analysis in section 2, before moving on to relative survival analysis in section 3 where we introduce the three estimators that we are to compare and some theory on the relationship between them. In section 4 we introduce an approximation for estimating the variance of ratios and illustrate the difficulties of this approach, before discussing resampling techniques. In section 5 we introduce a dataset of colon cancer patients, and smooth population mortality rates. In section 6 we present and

discuss variance and confidence interval estimates with and without including uncertainty from population mortality rates.

2 Survival Analysis

Survival analysis is a large subfield of statistics. The focus lies on the probability or hazard of *events* which could be for instance malfunction of a machine, death, pregnancy and many other things.

Survival data is often subject to *censoring* - that is, we do not observe the event of interest for all individuals. In a healthcare research context, depending on the aims of the study, causes of censoring could be death from other causes, drop-out of studies or end of study period, i.e. we do not have time to wait until all participants experience the event.

A similar complication is caused by late entries in a study. This is referred to as *truncation*, but will not be treated in any detail here as the data we will analyze was collected in such a way that all individuals are monitored starting from date of diagnosis, as compared to initiating the monitoring, e.g. a few years after diagnosis.

Even though survival analysis is applied in many settings, relative survival analysis is typically used for analysis of cancer survival. Because of this we will for convenience sometimes write *cancer* and *death* throughout the text, rather than disease and event of interest. That being said, we stress that estimates of relative survival typically concerns specific types of cancer, and not all types of cancer at once.

As our focus in this thesis is in relative survival, we settle for a minimal presentation of basic survival analysis concepts, needed for introducing the relative survival framework. For a more complete introuction, we direct the reader to introductory texts in survival analysis, such as Klein & Moeschberger (2003).

2.1 Hazard rates and Survival Functions

Let $T \geq 0$ denote a random survival time. Two quantities of fundamental interest in survival analysis are the survival function and the hazard rate. The following definitions of these quantities and related estimators can be found in Aalen et al (2008). The survival function is defined as

$$S(t) = \mathbb{P}(T > t),$$

while the hazard rate, under an assumption of T being absolutely continuous, is defined as

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}(t \leq T < t + \Delta t | T \geq t).$$

The most well known estimators of the survival function and the cumulative hazard rate are non-parametric. The Kaplan-Meier-estimator of the survival function is defined as

$$\tilde{S}(t) = \prod_{T_j < t} \left(1 - \frac{1}{r(T_j)}\right)$$

where $r(T_j)$ denotes the number at risk "just before" time t. The Nelson-Aalen estimator of cumulative hazard rate is defined as

$$\tilde{A}(t) = \sum_{T_j \leq t} \frac{1}{r(T_j)}.$$

3 Relative Survival Analysis

The relative survival ratio was introduced by Berkson (1942) as a ratio of one observed and one expected survival function. It is defined in Dickman & Coviello (2015) as

$$RS(t) = \frac{\sum_{i=1}^{n} S_i(t)}{\sum_{i=1}^{n} S_i^*(t)}$$

where n is the sample size of the cancer cohort and $S_i(t)$ denotes the individual survival function of an invidual i subject to cancer, while $S_i^*(t)$ is the expected survival of the same individual had he or she not had cancer.

The interpretation of the relative survival ratio relates to a fictional world, where the only possible way of dying is from cancer. That is, a relative survival of 1 does not mean that a group of patients will survive, but that they will be subject to the same mortality as the overall population. The relative survival ratio could also indicate a large relative difference in survival for a rare cancer type, while the majority of the mortality in a population could be related to other, more common diseases.

Although it is not clear from the original papers, several authors (Dickman & Coviello (2015), Esteve et al (1990) and Pohar Perme et al(2012)) have suggested that the intention of the relative survival ratio was to estimate a quantity known as net survival. We give some background before defining net survival.

For the hazard rate, one could base relative survival on both multiplicative and additive hazard models. For a comparison of the two, see Buckley (1984). Dickman et al (2003, p. 61) claims that additive models are generally considered "most appropriate for population-based cancer survival". Thus we will not discuss multiplicative hazard models further here. Instead we assume that the individual observed hazard rate $\alpha_i(t)$ follow an additive hazard model, defined in Andersen & Vaeth (1989) as

$$\alpha_i(t) = \alpha_{Pi}(t) + \alpha_{Ei}(t),$$

where $\alpha_{Pi}(t)$ is the overall population hazard for individual i, while $\alpha_{Ei}(t)$ is the excess hazard due to cancer for the same individual. We define *net survival* for individual i, as the survival function constructed from the individual excess hazard. The definition is given in Pohar Perme et al (2012) as

$$S_{Ei}(t) = \exp\left(-\int_0^t \alpha_{Ei}(s)ds\right).$$

These individual survival functions are then averaged into the overall net survival as $S_E(t) = 1/n \sum_{i=1}^{n} S_{Ei}$. The terminology "net" and later on "excess", is

due to the extra risk from having cancer. After this introduction of the central concepts in relative survival analysis, we will now discuss different estimators of the relative survival components.

3.1 Estimating Observed Survival

When working with large scale registry data, both events and censorings are typically recorded at discrete timepoints, typically on yearly or monthly basis. This produces ties, of both censored and event survival times. The tie corrected version of the Kaplan-Meier-estimator is defined in Aalen et al (2008) as

$$\tilde{S}^{TC}(t) = \prod_{T_j \leq t} (1 - \frac{d(T_j)}{r(T_j)}) \tag{1}$$

where superscript TC denotes tie corrected, and $d(T_j)$ the number of events occuring at T_j. Here, tie corrected version is to be understood as with respect to tied events, rather than ties between events and censorings. For the latter kind, ties are solved by placing the censored times just after the time of the tied observed events. As this inferred order is unlikely to hold in reality, another estimator is commonly used for discrete timepoints data.

Cutler & Ederer (1958) describe the so called actuarial estimator or life-table method. It is defined in Kalbfleisch & Prentice (2002) as

$$\hat{S}(t) = \prod_{T_j \leq t} (1 - \frac{d_j}{r_j - c_j/2}) \tag{2}$$

where quantities with subscript j denotes number of individuals censored (c_j), deceased (d_j) or at risk (r_j) during time interval $I_j = (T_{j-1}, T_j]$. Note that it is sufficient to be at risk during the beginning of the interval to be counted in r_j. This estimator differs from (1), as it is defined for discrete time, but also as it attributes only half of the at risk-time for censored individuals. The intuition is that the censorings occur uniformly within the interval. If one agrees that this is a better way way of handling ties of censorings and events, the actuarial estimator is a reasonable estimator.

The following derivation of the actuarial estimator is from Cox & Oakes (1984), but as we will see, more or less arbitrary approximations are used to end up with the actuarial estimator. First, assume that data are recorded at discrete, prespecified time-points T_j, e.g. at the end of each year. Let times of censoring and events in interval j follow two separate hazard rates α_{Dj} and α_{Cj}, with corresponding survival times T_D and T_C denoting death and censoring, with $T = \min(T_D, T_C)$. The hazard rates are assumed piecewisely constant on each interval, and each interval is of length ℓ_j. We consider the contribution to the joint likelihood $\mathcal{L}(\alpha_{Dj}, \alpha_{Cj})$ for each interval I_j, from the following three outcomes

1. $r_j - d_j - c_j$ individuals survive beyond I_j

2. c_j individuals are censored during I_j
3. d_j individuals die during I_j.

In respective order, we give the probabilities of each outcome, conditioned on being at risk just before interval I_j as

1. $\mathbb{P}(T > T_j | T > T_{j-1}) = \exp(-\int_0^{\ell_j} \alpha_{Dj} + \alpha_{Cj} du) = \exp(-\ell_j(\alpha_{Dj} + \alpha_{Cj}))$
2. $\mathbb{P}(T_C \leq T_j | T > T_{j-1}) = (1 - \exp(-\ell_j(\alpha_{Dj} + \alpha_{Cj}))) \times \frac{\alpha_{Cj}}{\alpha_{Dj}+\alpha_{Cj}}$
3. $\mathbb{P}(T_D \leq T_j | T > T_{j-1}) = (1 - \exp(-\ell_j(\alpha_{Dj} + \alpha_{Cj}))) \times \frac{\alpha_{Dj}}{\alpha_{Dj}+\alpha_{Cj}}$

where the first probability follows from the standard relationship between hazard rates and survival functions, i.e. $S(t) = exp(-\int_0^t \alpha(u)du)$. The second and third conditional probabilities are derived as complements of the first probability, weighted according to each hazard rate. Writing up the complete likelihood and taking derivatives with respect to the two hazard rates in each interval, followed by substitution yields the following maximum likelihood estimator

$$\hat{\alpha}_{Dj} = -\frac{d_j}{\ell(d_j + c_j)} \log\left(\frac{r_j - d_j - m_j}{r_j}\right).$$

If the number of censorings and deaths in one interval is small relative to the number at risk, we can use Taylor expansion of $\log(1 - x)$ on the logarithm-factor, which gives

$$\ell_j \hat{\alpha}_{Dj} = \frac{d_j}{r_j} + \frac{d_j(d_j + c_j)}{2r_j^2} + O\left[\left(\frac{d_j + c_j}{r_j}\right)^3\right] =$$

$$\frac{d_j}{r_j - \frac{1}{2}(d_j + c_j)}\left(1 + O\left[\left(\frac{d_j + c_j}{r_j}\right)^2\right]\right)$$

where $O[x^k]$ denotes a kth order term which goes to zero as quickly as x^k when x $\to$ 0. Omitting the second order term gives an approximation of the hazard rate. Integrating this over I_j for the cumulative hazard, cancels the ℓ_j. From the standard relation between survival function and hazard rate, the conditional probability of failure becomes $1 - \exp(-\hat{\alpha}_{Dj})$. We approximate this using the Taylor expansion of $1 - \exp(-x) \approx x$, as

$$1 - \exp(-\hat{\alpha}_{Dj}) \approx \hat{\alpha}_{Dj} = \frac{d_j}{r_j - \frac{c_j}{2}}.$$

Note that apart from the Taylor approximation, $-d_j/2$ is removed from the denumerator in this step. Cox and Oakes (1984) claim that the approximating step is still of order $O(n^2)$, as the first approximation, without providing any further details. Finally we take the complement for the conditional probability of surviving the jth interval, and combine these for the actuarial estimator of $S(t)$. Keeping $-d_j/2$ might seem more straightforward, and produces a version of the actuarial estimator where both deaths and censorings occur uniformly within each interval. However, due to tradition we will use the actuarial estimator here.

3.2 Estimating Expected Survival

When estimating the expected survival, S^*, the quantity we want to estimate is the expected survival of the cancer cohort, had they not been subject to the disease of interest. Thus, each individual in the cohort is matched with a population mortality rate based on their covariates, typically calendar year, age and sex.

We note that the population mortality rates are typically not corrected for including individuals with the disease of interest. This is based on an assumption of the disease of interest being rare in the overall population. The idea of correcting the population mortality rates is discussed in Ederer et al (1961) who from previous studies concluded that such adjustments seemed to have negligible effect. More recently, Talbäck & Dickman (2011) do explore the impact of removing individuals with cancer diagnosis from the population mortality, but find that the differences are small as long as the disease is relatively rare. As an illustration they suggest that if one was to estimate relative survival for all cancer types combined as a single disease, this would require adjustment.

Given a cohort and matching population mortality rates, there are several suggestions on how to proceed to estimate relative survival. To our knowledge, four nonparametric estimators have been suggested: Ederer I, Ederer II, Hakulinen and Pohar Perme. Ederer I is not widely used according to Dickman et al (2013), while the proposer of the Hakulinen estimator suggested that Ederer II should be preferred in Hakulinen et al (2011). Thus, we will focus on Ederer II and Pohar Perme.

Remontet et al (2006) criticize nonparametric procedures for not being clinically probable, as they produce stepwise constant estimates. A parametric suggestion from Royston & Parmar (2002) is the so called Flexible Parametric Models, which incorporate splines and maximum likelihood estimation, which produce smooth estimates. There are numerous other parametric suggestions, but even though Lambert et al (2015, p. 4) claims that there is a "growing use of statistical models for excess mortality", nonparametric estimators are typically used, see e.g. Roche et al (2013). Here we will consider Flexible Parametric Models as a complement to the nonparametric estimators.

We continue with presenting our three selected estimators, Ederer II, Pohar Perme and Flexible Parametric Models in closer detail.

3.3 Ederer II-estimator

Ederer II provides an estimate of $S^*(t)$, which combined with the estimate of $S(t)$ from the actuarial estimator provides an estimator of net survival. For a yearly recorded time scale, the estimate of $S^*(t)$ is defined as the product of annual averaged conditional probabilities of surviving a time interval T_j for those still at risk in the cohort. It is defined in Ederer & Heise (1959) as

$$S^*_{EdII}(t) = \prod_{t=1}^{Y} \left(\frac{1}{r_j} \sum_{i=1}^{r_j} \mathbb{P}(T_i > t | T_i > t-1) \right) \tag{3}$$

where t is the time from 1 to Y years, sum up to r_j denotes that we are summing over individuals at risk in the beginning of the jth year, and T_i the time when individual i is either deceased or censored. Matching population mortality probabilities on covariates for individual i gives the individual conditional probabilities. If population mortality rates θ are assumed to be known for the demographic covariates calendar year, age and sex found in the cohort, we match these to the conditional probabilities for individual i at specific calendar years and ages as

$$\mathbb{P}(T_i > t | T_i > t-1) = \exp\left(-\int_0^1 \theta_i(u)du\right).$$

The raw Ederer II estimator is not commonly used, due to differences in expected survival between different age groups, i.e. older people have lower overall survival than younger. This causes bias, see e.g. Lambert et al (2015) and Pohar Perme et al (2012). To adjust for this, the so called internal age standardization is used, which corresponds to a weighted mean of G stratified Ederer II relative survival estimates RS_{EdIIa}, where a is for age groups $a = 1, \ldots, G$.

$$RS_{\text{St.}}(t) = \sum_{a=1}^{G} w_a RS_{\text{EdIIa}}(t),$$

where w_a equals the size of each age group in the population divided with the overall population size, thus $\sum_{a=1}^{G} w_a = 1$. The intuition for this approach is that bias in each age group will be smaller in the more similar age groups, compared to the overall population. The benefit of internal age standardization is supported from simulations, e.g. in Lambert et al (2015). The term *internal* is understood as opposed to *external* age standardization, which is a similar technique to correct for differences in age distribution, for instance when comparing two different countries.

Pohar Perme et al (2012) showed that continuous time versions of the Ederer I, II and Hakulinen-estimators are biased when it comes to estimating net survival. We follow Pohar Perme et al (2012) and present the proof for Ederer II being biased, and prepare for the proof of the Pohar Perme estimator being unbiased.

Let T_{Ei}, T_{Pi} and T_{Ci} be random variables for time of death due to cancer, general population mortality and censoring in respective order. Next, let T_{Di}, time of death, be defined as $T_{Di} = \min(T_{Ei}, T_{Pi})$ and $T_i = \min(T_{Di}, T_{Ci})$. We also let X_i be covariates for the individual, with a subset X_i' of demographic covariates. Demographic variables are typically sex, ages and calendar years for which the individual was observed.

Next we assume noninformative censoring, that is $S_{Ci}(t) = S_C(t)$ for all individuals i, where S_{Ci} is the survival function of T_{Ci}. We also let T_{Ei} and T_{Pi} be conditionally independent given X_i', that is

$$\mathbb{P}(T_{Ei} \leq t, T_{Pi} \leq t | X_i') = \mathbb{P}(T_{Ei} \leq t | X_i')\mathbb{P}(T_{Pi} \leq t | X_i').$$

We continue with defining the individual excess hazard rate α_{Ei},

$$\alpha_{Ei}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}(t \leq T_{Ei} < t + \Delta t | T_{Ei} \geq t),$$

with corresponding survival function $S_{Ei} = \exp(-\int_0^t \alpha_{Ei}(u)du)$. The overall excess survival function is defined as $S_E(t) = \frac{1}{n}\sum_{i=1}^n S_{Ei}(t)$, which we use to define the overall excess hazard rate α_E as

$$S_E(t) = \exp\left(-\int_0^t \alpha_E(u)du\right).$$

From this we derive another expression for the excess hazard $\alpha_E(t)$. After taking log and differentiating both sides with respect to t we find that

$$\alpha_E(t) = \frac{\sum_{i=1}^n S_{Ei}(t)\alpha_{Ei}(t)}{\sum_{i=1}^n S_{Ei}(t)} \tag{4}$$

Later we will show that this is the quantity estimated by the Pohar Perme estimator. Next, we define the cause specific hazard $\tilde{\alpha}_E(t)$

$$\tilde{\alpha}_E(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}(t \leq T_E < t + \Delta t | T \geq t),$$

where we note the difference in conditioning compared to the excess hazard. The same conditioning is found in the definition of the so called population hazard

$$\tilde{\alpha}_P(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}(t \leq T_P < t + \Delta t | T \geq t).$$

As an observed death must correspond to one of these two possible causes, these two quantities sum up to the observed hazard,

$$\alpha_O(t) = \tilde{\alpha}_E(t) + \tilde{\alpha}_D(t). \tag{5}$$

We note that this also holds for the individual hazards.

When considering individual hazards, and thus conditioning on the demographic covariates X', one could show that

$$\begin{aligned}
\tilde{\alpha}_{Ei}(t) &= \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T_{Ei} < t + \Delta t | \min(T_{Ei}, T_{Pi}) \geq t, X')}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T_{Ei} < t + \Delta t, T_{Pi} > t | X')}{\mathbb{P}(T_{Ei} \geq t, T_{Pi} \geq t | X')\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T_{Ei} < t + \Delta t | X')\mathbb{P}(T_{Pi} \geq t | X')}{\mathbb{P}(T_{Ei} \geq t | X')\mathbb{P}(T_{Pi} \geq t | X')\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T_{Ei} < t + \Delta t | X')}{\mathbb{P}(T_{Ei} \geq t | X')\Delta t} = \alpha_{Ei}(t)
\end{aligned}$$

from the conditional independence of T_{Ei} and T_{Pi} under X'. Thus there is no need in distinguishing $\tilde{\alpha}_{Ei}(t)$ from $\alpha_{Ei}(t)$.

Next, it follows from the definition of $\tilde{\alpha}_E(t)$ that

$$\tilde{\alpha}_E(t) = \frac{\sum_{i=1}^n S_{Oi}(t)\alpha_{Ei}(t)}{\sum_{i=1}^n S_{Oi}(t)} = \alpha_O - \frac{\sum_{i=1}^n S_{Oi}(t)\alpha_{Pi}(u)}{\sum_{i=1}^n S_{Oi}(t)} \tag{6}$$

where the second equality is due to (5) and

$$\alpha_O(t) = \frac{\sum_{i=1}^n S_{Oi}(u)\alpha_{Oi}(u)}{\sum_{i=1}^n S_{Oi}(u)} du.$$

From this, we define *observable net survival* as

$$\tilde{S}_E(t) = \exp\left(-\int_0^t \tilde{\alpha}_E(u)du\right)$$

From the expression for $\tilde{\alpha}_E(t)$, we note that unless α_{Ei} is constant across i, the observable net survival depends on S_{Oi}, which includes the population mortality. From this Pohar Perme et al (2012, p. 115) argues that observable net survival "cannot be used as a measure of cancer burden" as it depends on the population mortality from S_{Oi}. Written as in (6) we also note that observable net survival then differs from (4), i.e. net survival.

We will now show how a continuous time version of the Ederer II estimator is consistent for observable net survival. Let $N(t) = \sum_{i=1}^n \mathbb{1}(T_{Di} \leq t, T_{Di} \leq T_{Ci})$, i.e. the counting process of the number of observed deaths up to time t, and similarly, let the at risk-process be denoted $R(t) = \sum_{i=1}^n R_i(t)$ where $R_i(t) = \mathbb{1}(T_i \geq t)$. Pohar Perme et al (2012) then refers to Andersen & Vaeth (1989) for the following estimator of cumulative excess hazard, under the assumption of equal individual excess hazards α_{Ei}

$$\hat{A}_E(t) = \int_0^t \frac{dN(u)}{R(u)} - \int_0^t \frac{\sum_{i=1}^n R_i(u)dA_{Pi}(u)}{R(u)}, \tag{7}$$

where dA_{Pi} is the average change in cumulative population hazard. The first term of the estimator corresponds to the Nelson-Aalen estimator of observed hazard. Rewriting the estimator on the survival scale gives

$$\exp(-\hat{A}_E(t)) = \frac{\exp(-\int_0^t \frac{dN(u)}{R(u)}))}{\exp(-\int_0^t \frac{\sum_{i=1}^n R_i(u)dA_{Pi}(u)}{R(u)})},$$

where the numerator corresponds to an all-cause survival function estimate, while the denominator is a sum of indicators for being at risk, multiplied with the average change in cumulative population hazard over that interval, for individuals at risk. Thus, this is a continuous time-version of the Ederer II estimator.

The Nelson-Aalen estimator is consistent for the observed hazard, while

$$\mathbb{E}(R_i(t)) = \mathbb{P}(R_i(t) = 1) = S_{Oi}(t)S_{Ci}(t)$$

as you must be alive and uncensored to be at risk. Next, we assume $S_{Ci}(t) = S_C(t)$, i.e. noninformative censoring, which gives that the numerator of the last term of (7)

$$(1/n)\sum_{i=1}^{n} R_i(t)dA_{Pi}(t)$$

is consistent for $S_C(t)/n\sum_{i=1}^{n} S_{Oi}(t)dA_{Pi}(t)$, while the denominator of the last term of (7)

$$(1/n)\sum_{i=1}^{n} R_i(t)$$

is consistent for $S_C(t)/n\sum_{i=1}^{n} S_{Oi}(t)$. Combining the parts shows that $\hat{A}_E(t)$, after cancelling $S_C(t)$, is consistent for

$$\int_0^t \alpha_O(u) - \frac{\sum_{i=1}^{n} S_{Oi}(u)\alpha_{Pi}(u)}{\sum_{i=1}^{n} S_{Oi}(u)}du$$

which is a cumulative version of (6). This proves that Ederer II estimator is consistent for observable net survival. As we noted earlier, observable net survival is not the same as net survival, and hence Ederer II does provide an unbiased estimate of net survival.

Simulations in Lambert et al (2015) do however suggest that the bias of the Ederer II estimator is small, and that the larger variance of the Pohar Perme estimator compared to internally age standardized Ederer II is a reason to prefer internally age standardized Ederer II.

3.4 Pohar Perme-estimator

The Pohar Perme estimator was first suggested for continuous time by Pohar Perme et al (2012), as a reweighted version of the Ederer II estimator. We will begin with defining a discrete version, to illustrate the reweighting.

Let $\mathbb{1}_{ij} = \mathbb{1}(T_{Di} \in I_j)$, i.e. an indicator of death for individual i in interval I_j. Next, let D^*_{ij} be the expected number of deaths for individual i during I_j, had he or she not had cancer. D^*_{ij} is derived from matching population mortality rates of individual i during calendar year and age matching I_j, as $D^*_{ij} = -\log(\theta_i)r_{ij}$, where θ_i is the conditional probability of surviving interval I_j given that you survived interval I_{j-1}, for an individual without cancer. Using this notation, a discrete time version of the Pohar Perme estimator, for time interval I_j, is defined in Lambert et al (2015) as

$$\alpha_j^{\mathrm{PP}} = \frac{\sum_{i=1}^{r_j} w_{ij}\mathbb{1}_{ij} - \sum_{i=1}^{r_j} w_{ij}D^*_{ij}}{\sum_{i=1}^{r_j} w_{ij}r_{ij}},$$

where r_{ij} is the time at risk for individual i during I_j, and w_{ij} the inverse of the expected survival of individual i in time interval I_j, that is $w_{ij} = 1/S_{Pij}(t)$

evaluated at the midpoint of the jth interval. The estimate is transformed to the cumulative hazard scale by summing over I_j as

$$A_j^{\mathrm{PP}}(t) = \sum_{I_j} \ell_j \alpha_j^{\mathrm{PP}}$$

where ℓ_j is the length of I_j and to the probability scale by using $S(t) = \exp(-A(t))$. If all weights w_{ij} were set to 1, the numerator of α_j^{PP} is an averaged difference of mortality in the cohort and the overall population, over the individuals at risk, while the denominator is the at risk-time for those at risk. Lambert et al (2015) argues that setting weights w_{ij} equal to one gives a hazard scale version of the Ederer II estimator. Hence the Pohar Perme estimator is a reweighted version of the Ederer II estimator.

Next, we give a proof of the so called Pohar Perme-estimator being consistent for (4), which after transformation to the probability scale corresponds to net survival. We follow Pohar Perme et al (2012) and define the continuous time definition of the Pohar Perme estimator on the cumulative hazard scale.

$$\hat{A}^{\mathrm{PP}}(t) = \int_0^t \frac{dN_w(u)}{Y_w(u)} - \int_0^t \frac{\sum_{i=1}^n R_{iw}(u) dA_{Pi}(u)}{R_w(u)} \tag{8}$$

where subscript w denotes a reweighting of the counting process with the inverse population survival function, i.e. $R_w(t) = 1/n \sum_{i=1}^n R_{iw}(t)$ where $R_{iw}(t) = R_i(t)/S_{Pi}(t)$ and $N_w(t) = 1/n \sum_{i=1}^n N_{iw}(t)$ where $N_{iw}(t) = N_i(t)/S_{Pi}(t)$.

Next we take the expectation of $R_{iw}(t)$

$$\mathbb{E}(R_{iw}(t)) = S_C(t) S_{Oi}(t)/S_{Pi}(t) = S_C(t) S_{Ei}(t),$$

since $S_{Oi}(t) = S_{Pi}(t) S_{Ei}(t)$. The expectation of dN_{iw} is found as

$$\mathbb{E}(dN_{iw}(t)|\mathcal{F}_t) = \mathbb{E}(dN_i(t)|\mathcal{F}_t)/S_{Pi}(t) = R_i(t) dA_{Oi}(t)/S_{Pi}(t)$$

where $\mathcal{F}$ represents the history up to time t, see Aalen et al (2008). After assuming noninformative censoring, i.e. $S_{Ci}(t) = S_C(t)$, we note that

$$\frac{1}{n} \sum_{i=1}^n \frac{R_i(t) dA_{Oi}(t)}{S_{Pi}(t)}$$

is consistent for

$$\frac{S_C(t)}{n} \sum_{i=1}^n \frac{S_{Oi}(t) dA_{Oi}(t)}{S_{Pi}(t)} = \frac{S_C(t)}{n} \sum_{i=1}^n S_{Ei}(t) dA_{Oi}(t).$$

From this it follows that the two terms in (8) is consistent for

$$\frac{S_C(t) \sum_{i=1}^n S_{Ei}(t) dA_{Oi}(t)}{S_C(t) \sum_{i=1}^n S_{Ei}(t)} - \frac{S_C(t) \sum_{i=1}^n S_{Ei}(t) dA_{Pi}(t)}{S_C(t) \sum_{i=1}^n S_{Ei}(t)}.$$

After removing $S_C(t)$, we note that $\alpha_{Oi}(t) - \alpha_{Pi}(t) = \alpha_{Ei}(t)$. Thus, this is a consistent estimator of excess hazard, i.e. (4) and the proof is complete. As a summarizing remark we note that the reweighting of the Ederer II estimator changes $S_{Oi}(t)$ in (6), into $S_{Ei}(t)$, from which the unbiasedness follows.

In practical work, for instance in the STATA implementation **strs** used for the Pohar Perme estimation in this thesis, the following version of the Pohar Perme estimator is used. To emphasize that this estimator is unbiased for net survival, we write NS_j rather than RS_j for the estimate in the jth interval. The following definition can be found in Dickman & Coviello (2015).

$$NS_j = \frac{1 - \frac{d_{jw}}{r_{jw} - c_{jw}/2}}{\exp\left(-\frac{\sum_{i=1}^{r_j} \theta_{ijw} - \sum_{i=1}^{c_j} \theta_{ijw}/2 - \sum_{i=1}^{d_j} \theta_{ijw}/2}{r_{jw} - (d_{jw} + c_{jw})/2}\right)}$$

where subscript w as before denotes a reweighting with the inverse population mortality, and θ_{ijw} denotes a population hazard rate for individual i, in the specific jth interval. The numerator illustrates how the actuarial estimator is reweighted for an estimate of observed survival. The numerator of the denominator estimates the expected number of deaths by subtracting half of the mortality rates from deceased and censored individuals from the population mortality rates from those at risk, and divides with a similarly adjusted at risk-time. Dickman & Coviello (2015, p. 191) claim that estimates are "essentially identical" to a continuous time implementation of the Pohar Perme estimator. The product of NS_j over j yields the cumulative net survival estimate.

3.5 Flexible Parametric Models

Royston & Parmar (2002) generalize well known parametric models such as Weibull and LogNormal distribution, by using a spline as baseline to make the traditional models more flexible. The resulting models are called Flexible Parametric Models. We follow Royston & Lambert (2011) in our introduction of splines and the general Flexible Parametric Model framework. We end by presenting these models in the relative survival setting.

3.5.1 Splines

Splines are piecewise polynomial functions on which overall smoothness is imposed by restrictions on derivatives in certain positions, known as knots, as these positions ties the piecewise functions together into the overall fit. Depending on the number of knots, splines are typically divided into regression splines and smoothing splines, where the former uses a relatively small number of knots and the latter typically has a knot for each data point. We will focus on regression splines, and the subclass where both number and locations of knots are decided before fitting the actual splines to data.

In our setup this is achieved by first manually specifying k degrees of freedom, which results in $k+1$ knots. These knots are then evenly positioned according to

the percentiles of the uncensored log survival times, with two so called boundary knots, $k_{\max}$ and $k_{\min}$, at each outmost survival time. From k knots, we can define a spline, without continuity restrictions, of order n for covariate x as follows. The definition is found in Royston & Lambert (2011).

$$s(x) = \sum_{j=0}^{n} \beta_{0j} x^j + \sum_{i=1}^{k} \sum_{j=0}^{n} \beta_{ij} (x - k_i)_+^j$$

where "+"-subscript of an expression a denotes the product of a with an indicator function $\mathbb{1}(a > 0)$. Order $n = 3$ is a common choice of spline order, yielding $K+4$ parameters for a cubic regression spline. Smoothness across knots is imposed by demanding continuity and equality of first and second order derivatives at knots for piecewise functions sharing knots. To avoid variability in the fit for outer regions, the cubic splines are forced to be linear outside the external knots as well. This produces so called *restricted* cubic splines. To include a covariate x, one does not use the raw x-values, but transform x into z by the following relations

$$z_1(x) = x$$
$$z_j(x) = (x - k_j)_+^3 - \lambda_j (x - k_{\min})_+^3 - (1 - \lambda_j)(x - k_{\max})_+^3$$

where

$$\lambda_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}}$$

for $j \in \{2, \ldots, m+1\}$. The complete, covariate dependent spline is then written as

$$s(x) = \gamma_0 + \sum_{i=1}^{m+1} \gamma_i z_i \tag{9}$$

where γ_i, for $i \in \{0, \ldots, m+1\}$ are the coefficients of the spline.

3.5.2 Generalizing the Weibull distribution

We proceed with illustrating how flexible parametric models generalize parametric models, here the Weibull distribution, into a Flexible Parametric Model. The cumulative hazard of the Weibull distribution can be written as

$$A(t) = \lambda t^{\gamma^*},$$

where $\gamma^* > 0$ is a shape parameter, and superscript $*$ distinguishes γ^* from γ in the preceding splines section. We write this on the log scale as

$$\log A(t) = \log \lambda + \gamma^* \log t.$$

From here, we think of the two terms on the right hand side as the first two terms of the right hand side of (9), if z_1 equals $\log(t)$. For $m + 1$ number of knots, the generalization becomes

$$\log A(t) = s(\log(t), \gamma) = \gamma_0 + \sum_{i=1}^{m+1} \gamma_i z_i(\log(t)).$$

We write $\log(t)$ to emphasise that it is common to work on the log of time scale, rather than the time scale. We also note that in relative survival analysis, the log cumulative hazard is not neccesarily a monotonically increasing fuction, thus we do not need to impose any further restrictions on $s(\log(t), \gamma)$. From here, one could go on and include other covariates than survival time, but here we will only use survival time as covariate thus we move on to the relative survival setting.

3.5.3 Flexible Parametric Models in Relative Survival

We conclude this section following Nelson et al (2007) and present Flexible Parametric Models in a relative survival framework.

A well known result in survival analysis, see e.g. Klein & Moeschberger (2003) is that the contribution of individual i to the log likelihood, without truncation, is

$$\log \mathcal{L}_i = O_i \log(\alpha(t)) + \log(S(t))$$

where O_i is an indicator of death and t is the survival time for an individual. Inserting an individual additive hazard function, and adding and subtracting $\log(S^*(t))$ produces

$$\log \mathcal{L}_i = O_i \log(\alpha_{Pi}(t) + \alpha_{Ei}(t)) + \log(S^*(t)) + \log(R(t)) \tag{10}$$

where $R(t)$ is the relative survival of individual i. Under the assumption of $S^*(t)$ being known, we can remove it from the likelihood under proportionality. This corresponds to treating the population mortality estimates as constants. Note that $R(t)$ is the relative survival, rather than an estimate from e.g. the relative survival ratio, and hence the population mortality only enters the likelihood through $\alpha_{Pi}(t)$. Next, we model the log cumulative excess hazard with a spline, similar to the Weibull generalization as

$$\log(A_{Ei}(t)) = s(\log(t), \gamma)$$

where γ are the parameters of the spline. From this, we derive expressions for $R(t)$ and $\alpha_{Ei}(t)$ in (10) as

$$R(t) = \exp(-\exp(\log(A_{Ei}(t))))$$

and

$$\alpha_{Ei}(t) = \frac{1}{t} \frac{ds(\log(t), \gamma)}{d(\log(t))} \exp(\log(A_{Ei}(t))),$$

and plug these into (10) for the likelihood contribution of a single individual,

$$\log \mathcal{L}_i = O_i \log \left[\alpha_{Pi}(t) + \frac{1}{t} \frac{ds(\log(t), \gamma)}{d(\log(t))} \exp(s(\log(t), \gamma)) \right] - \exp(s(\log(t), \gamma)).$$

As before we are not interested in including other covariates than survival time, and direct the interested reader to Nelson et al (2007).

4 Variance of relative survival estimates

We will now discuss the variance of relative survival estimates. We will start with the general case of estimating the variance of a ratio of two random variables. That is, we consider

$$Var\left(\frac{X}{Y}\right)$$

If we make assumptions regarding the distribution of X and Y, some general results hold. It is a well known fact that the ratio of two independent standard gaussian variables produces a Cauchy distributed variable. The ratio of two sample averages of normally distributed samples is discussed in Fieller (1954). However, these results do not apply to our survival data situation.

Deriving an exact formula of the variance of a ratio of two arbitrary random variables has been commented upon by Casella & Berger (2002, p. 245) as "quite hopeless". Instead they suggest using an approximation known as the Delta method.

4.1 Delta method

We introduce the Delta method following Casella & Berger (2002) and begin with defining univariate Taylor polynomials. Let $f(x)$ be a function with derivatives of order n_T. For a constant c, the Taylor polynomial of order n_T of $f(x)$ is

$$T_{n_T}(x) = \sum_{i=0}^{n_T} \frac{f^{(i)}(c)}{i!}(x-c)^i.$$

The difference between $f(x)$ and $T_{n_T}(x)$ is called the remainder, and when omitted, the Taylor polynomial approximates $f(x)$.

Next we present a multivariate version of the Taylor polynomials, in a random variable setting. Let $T = (T_1, \ldots, T_k)$ be statistics with expectations $\mu = (\mu_1, \ldots, \mu_k)$, and let $f(x)$ be a differentiable function, with

$$f_i'(\mu) = \frac{\partial f(x)}{\partial x_i}|_{x=\mu}$$

We can then, by omitting the remainder after the first order derivatives approximate $f(x)$ as

$$f(x) \approx f(\theta) + \sum_{i=1}^{k} f_i'(\mu)(x_i - \mu_i) \tag{11}$$

Taking expectation on both sides yields

$$\mathbb{E}(f(T)) \approx \mathbb{E}(f(\theta)) + 0 \tag{12}$$

as the expectation of x_i equals μ_i. From these two approximations, we can derive an approximation of the variance of $f(T)$ as

$$Var(f(T)) \approx \mathbb{E}\left[(f(T) - f(\mu))^2\right] \approx \mathbb{E}\left[(\sum_{i=1}^{k} f_i'(\mu)(T_i - \mu_i))^2\right] =$$

$$\sum_{i=1}^{k}(f_i'(\mu))^2 Var(T_i) + 2\sum_{i>j} f_i'(\mu)f_j'(\mu)Cov(T_i,T_j)$$

where the first approximation is due to (12) and the second from rearranging (11). A generalization to the covariance of two functions $f_1(x)$ and $f_2(x)$ of random variables is found in Klein (1953) as

$$Cov(f_1(T), f_2(T)) \approx \sum_{i=1}^{k}\left(\frac{\partial f_1}{\partial x_i}\right)\left(\frac{\partial f_2}{\partial x_i}\right)Var(T_i) +$$

$$+2\sum_{i>j}\left(\frac{\partial f_1}{\partial x_i}\right)\left(\frac{\partial f_2}{\partial x_i}\right)Cov(T_i,T_j)$$

We omit the derivation. These approximations, both univariate and multivariate, will be referred to as the delta method.

For convenience we will omit time dependencies of $S(t)$ and $S^*(t)$ for now. In the relative survival ratio setting, we are interested in $T = (T_1, T_2)$ where $\mathbb{E}(T_1) = S$ and $\mathbb{E}(T_2) = S^*$ and $f(x) = S/S^*$, i.e. the relative survival ratio. The partial derivatives are

$$\frac{\partial f(S,S^*)}{\partial S} = \frac{1}{S^*}, \frac{\partial f(S,S^*)}{\partial S^*} = \frac{-S}{S^{*2}}$$

which produces, after inserting estimates of the true quantities

$$\widehat{Var}(S/S^*) \approx (\frac{\hat{S}}{\hat{S}^*})^2\left(\frac{\widehat{Var}(S)}{\hat{S}^2} + \frac{\widehat{Var}(S^*)}{\hat{S}^{*2}} - \frac{2\widehat{Cov}(S,S^*)}{\hat{S}\hat{S}^*}\right)$$

That is, to approximate the variance of the relative survival ratio, we need estimates of S, S^*, as well as estimates of variances and covariance of these quantities.

4.2 Delta method applied to $Var(S/S^*)$

We will now discuss the Delta method applied to the components of the Ederer II estimate, which is the simpler one of our three estimators. An estimate of $Var(S)$, where S is estimated by the actuarial method, is provided from a version of the so called Greenwoods formula. An asymptotic argument for the Kaplan-Meier estimator can be found in Cox & Oakes (1984). First, assume that event times are set before observing the data, e.g. montly or annual data, and a regularity condition on the censoring mechanism (it should allow the number of events to increase at the same rate as the sample size n). From this the standard large sample-sample maximum likelihood theory holds. The variance estimate from the inverse of the observed information matrix of the Kaplan-Meier estimator is identical to diagonal elements of a covariance matrix of several independent binomial distributions, and thus

$$Var(d_j/r_j)) = \frac{d_j(r_j - d_j)}{r_j^3},$$

From taking log of the actuarial estimator, and applying the delta method twice, we end up at Greenwood's formula for the Kaplan-Meier estimator. The result can be found in Cox & Oakes (1984).

$$Var(\tilde{S}(t)) = \tilde{S}(t)^2 \sum_{T_j \leq t} \frac{d_j}{(r_j)(r_j - d_j)}$$

To adapt this for the actuarial estimator, the Kaplan-Meier estimate $\tilde{S}(t)$ is replaced with the actuarial estimate $\hat{S}$, and the number at risk is adjusted as in the actuarial estimator, that is

$$Var(\hat{S}(t)) = \hat{S}(t)^2 \sum_{T_j \leq t} \frac{d_j}{(r_j - 1/2c_j)(r_j - 1/2c_j - d_j)}.$$

Next, we want to estimate the variance of S^*. Recall that Ederer II uses a sample average, where the number of terms depends on the at risk set in the cohort, while each term consists of population mortality rates, which by themselves include variance. Given population mortality predictions $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_i, \ldots, \hat{\theta}_n)$ as earlier, matched to each individual i, with variance-covariance matrix entries $Cov(\theta_i, \theta_j)$ for $i, j \in 1, \ldots, n$ we can write up the variance as follows

$$Var(S^*(t)) = Var\left(\frac{1}{r_j} \sum_{j=1}^{n} \mathbb{1}(T_j > t-1) \exp(-\hat{\theta}_j)\right).$$

Here we use an indicator to stress the dependence on the cohort, while the θ depend on the population mortality. Indicators equal 1 when an individual in the cohort was at risk at the beginning of the time period for which we are estimating the expected survival. That is, for the first year estimate, all individuals are included as they were all at risk during the beginning of the first year. For upcoming years, individual mortality rates are removed from the sum as individuals are no longer at risk in the beginning of that certain year. This implies that the first year estimates only contain stochasticity from the population mortality rates. Thus,

$$Var(S^*(1)) = \sum_{j=1}^{n} Var(\exp(-\hat{\theta}_j)) + 2 \sum_{j>i} Cov(\exp(\hat{\theta}_i), \exp(\hat{\theta}_j))$$

$$\approx \sum_{j=1}^{n} \exp(-2\hat{\theta}_j) Var(\hat{\theta}_j) + 2 \sum_{j>i} \exp(-(\hat{\theta}_i + \hat{\theta}_j)) Cov(\hat{\theta}_i, \hat{\theta}_j),$$

where the approximation is due to the delta method.

It is not trivial what covariance we are interested in estimating, in the sense that the composition of the cohort could be thought of as either fixed or random. Given the cohort, the covariance of S and S^* is 0 for the one year estimate. To see this, we imagine a cohort of one single person. Regardless of the outcome

for this individual (survives, censored or deceased during the first year), the individual was at risk in the beginning of the interval and thus the estimate of $S^*(1)$ is constant. The covariance of anything with a constant is 0.

Should we not consider the cohort as fixed, and instead consider variance from taking new cohorts, the matching from cohort to population mortality rates would produce a non-zero covariance, e.g. a single person cohort of a twenty year old would give a different expected survival than that of a fifty year old. We argue, from the derivation of Greenwood's formula, that the cohort should be considered as fixed as no alternative cohorts are considered in that derivation, rather the variation from Greenwood's formula is in the estimation of $Var(d_j/r_j)$. When estimating variance of binomial probabilities, one does not include variance from scenarios where the sampled had been replaced with other, unobserved samples.

Thus, if we consider the cohort to be fixed, it holds for $t = 1$ that

$$\widehat{Var}(S/S^*) \approx (\hat{S}/\hat{S}^*)^2(\widehat{Var}(S)/\hat{S}^2 + \widehat{Var}(S^*)/\hat{S}^{*2}),$$

where we know how to estimate each quantity in principle. However, beyond the first year, the number of terms in the Ederer II estimate of S^* will, apart from the uncertainty in population mortality rates, also depend on survival times of the cohort. We have not been successful in combining these two sources of variation into estimates of the variance of S^* or covariance of S and S^*, beyond the first year.

Since we are interested in estimates of survival times beyond the first year, the analytical approach is not very successful. As the at risk-individuals are stochastic beyond the first year for Pohar Perme as well, we would expect similar issues to arise for this estimator, but at a more involved level, as we weight each term with an inverse of a survival function which is no longer considered as known. Even if we were succesful, we would not expect this approach to be applicable for the likelihood-based Flexible Parametric Models, which would complicate comparisons between the three different estimates.

4.3 Current practice

What is done in practice, see e.g. Parkin & Hakulinen (1991) or Dickman & Coviello (2015), is to estimate the variance of the relative survival ratio, as

$$Var(S/S^*) = Var(S)/\hat{S}^{*2},$$

That is ignore uncertainty in the population mortality rates. The variance of the age standardized version of Ederer II is usually approximated using the Greenwood's formula-approach stratified over age groups. If we let RS_a denote the RS estimate in each age group $a = 1, \ldots, A$, the variance is combined under an assumption of independence between stratified estimates using the delta method, as stated in Corazziari et al (2004)

$$Var(\hat{RS}(t)_{St.}) = \sum_{a=1}^{A} w_a^2 Var(RS_{EdIIa}(t)).$$

Under the assumption of constant population mortality rates, the only variability is due to the estimation of the observed survival through the actuarial estimator. As this is (at least approximately) a likelihood estimator, asymptotic normality is used to create confidence intervals. The standard approach is to use the delta method to construct confidence intervals on the log cumulative hazard scale, and then backtransform those to the probability scale, see e.g. Dickman & Coviello (2015). This construction ensures a confidence interval on the survival scale with endpoints within $[0, 1]$. The variance of the continuous time version of the Pohar Perme estimator on the hazard scale, assuming known population mortality rates, is derived by Pohar Perme et al (2012) to be

$$\sigma^2_{PP}(t) = \int_0^t \frac{J(u) \sum_{i=1}^n dN_i(u)/S^2_{Pi}(u)}{\left(\sum_{i=1}^n R_i(u)/S_{Pi}(u)\right)^2}$$

where $J(u)$ handles the case of the denumerator being 0, then the entire ratio is interpreted as 0. Pohar Perme et al (2012) remarks that the variance of the Pohar Perme estimator is usually larger than the variance of the Ederer II estimator.

Finally, the variance of Flexible Parametric Model estimates are based on asymptotic results of general likelihood theory, see Royston & Parmar (2002). We omit the details.

4.4 Variance estimation using Rubin's Rule

Before discussing confidence intervals, an estimate of variance might still be useful as an indication of where to expect differences. When we describe our data in chapter 5, we will model or smooth population mortality rates which allows us to take resamples of them with slight variability in each resample. But for now, we simply assume that resampled population mortality files are available.

If we consider the situation with an underlying random process, our observed population mortality rates are a single observation. If we could, we would prefer having several such observations, and could then say that we are missing observations of population mortality rates. The resampled versions of the observed population mortality could then be thought of as imputed data sets, and from this, we could use the so called Rubin's Rule to combine our resamples into a combined variance estimate. An overview of Rubin's Rule and the following definition can be found in Marshall et al (2009).

$$Var(RS) = \bar{W} + (1 + 1/n_I)B,$$

where n_I is the number of imputed data sets, $\bar{W}$ is the sample average of the estimated variances within each imputed data set, and B is the sample variance calculated on the relative survival estimates from each resampled data set. Note that the $\bar{W}$, the within variance, is a sample average of variance estimates calculated under the assumption of no variance in the population mortality rates, e.g. using Greenwoods formula for the Ederer II estimates. We note that

this differs from the typical application of Rubin's Rule, where the variance on each imputed data set should be an estimate including the uncertainty of population mortality, if the overall estimate is supposed to include population mortality uncertainty as well. Hence this procedure will provide a lower bound of the variance, the components of $\bar{W}$ could be smaller than they should.

In estimating variance with resampling procedures, Efron & Tibshirani (1993) suggests that more than 200 samples is rarely needed, which is a recommendation we will follow. However, when including variability in the denominator of the relative survival ratio, e.g. in the Ederer II estimator, it is no longer clear that the asymptotic normality is reasonable. As our interest lies in the confidence intervals, rather than the standard errors, we might prefer to construct confidence intervals without the normality assumption.

4.5 Bootstrap

A non-analytic way out of estimating confidence intervals is to use bootstrap procedures. Efron (1981) discusses applying the bootstrap to censored data. One suggestion is to create *bootstrap samples*, that is sample individuals from a cohort with replacement until the original sample size is reached. The discrepancy between standard errors from bootstrap procedures and Greenwood's formula, is found to be small for the Kaplan-Meier estimate, e.g. in Efron (1981). As the actuarial estimate is a slightly modified version of the Kaplan-Meier estimate, this provides some basis for expecting similarities between the actuarial estimator and estimates based on bootstrapping individuals.

In the relative survival context Brenner & Hakulinen (2005) compared standard errors of relative survival estimates calculated with analytical methods, under assumption of known population mortality, with standard errors based on bootstrap samples of cancer cohorts. Although they used the Hakulinen estimator, making comparisons less straightforward, the overall finding was that Greenwood's formula often overestimated standard errors for many cancer types, i.e. the Greenwood estimate may differ from bootstrap estimates.

Even though the analytical methods assuming known population mortality are the common procedure, comparing these directly to a bootstrap which includes population mortality uncertainty, mixes differences from population uncertainty and analytical methods as compared to bootstrap. We would prefer to assess the increase in variance due to included variance from uncertainty in population mortality. That is, we suggest using one bootstrap where we bootstrap the cohort and use a single, non-resampled population mortality, and one bootstrap where we bootstrap both cohort and population mortality rates independently.

Given the resamples of the population mortality and a cohort, we can describe the bootstrap sampling algorithm in list form as follows

Bootstrap of cohort and population mortality

1. Draw a bootstrap sample from the cancer cohort C_i

2. Independently resample population mortality rates P_i

3. For each iteration of 1 and 2, apply the three selected estimators on C_i and P_i.

4. Repeat steps 1-3 B times, to produce B bootstrap estimates which we denote $\hat{\eta}^*$

The independence in step 2 assumes that the covariance of cancer and population mortality is negligible, an assumption based on the rarity of cancer compared to the overall population.

When using bootstrap to estimate the variance without including variance from the population mortality rates, step 2 is replaced with $2'$. Use non-resampled population mortality rates.

As we will discuss in section 5, smoothing is introduced to create resamples of population mortality rates. To make comparisons between non-resampled and resampled population mortality rates more straightforward, we note that smoothed, but non-resampled population mortality rates are used in step $2'$.

4.5.1 Bootstrap confidence intervals

We will now discuss two methods for bootstrapping confidence intervals, following Efron & Tibshirani (1993). A simple confidence level approach is the so called percentile method based on the sample quantiles of the bootstrap estimates $\hat{\eta}^*$, for a given significance level 2α

$$(\hat{I}_{lo}^{\text{Per.}}, \hat{I}_{up}^{\text{Per.}}) = (\hat{\eta}_{\alpha}^*, \hat{\eta}_{1-\alpha}^*),$$

where the superscript Per. indicates that the Percentile method was used. However, the percentile method is only first-order accurate. That is, for an even, two-sided confidence interval of significance level $1 - 2\alpha$, the estimated confidence interval $(\hat{I}_{lo}^{\text{Per.}}, \hat{I}_{up}^{\text{Per.}})$, from B bootstrap estimates relates to the true confidence intervals (I_{lo}, I_{up}) as

$$\mathbb{P}(I_{lo} < \hat{I}_{lo}^{Per.}) = \alpha + \frac{c_{lo}}{\sqrt{B}} \text{ and } \mathbb{P}(I_{up} > \hat{I}_{up}^{Per.}) = \alpha + \frac{c_{up}}{\sqrt{B}}$$

where c_{lo} and c_{up} are error constants. Thus, the error decreases as $1/\sqrt{n}$ as n grows. Another bootstrap procedure for estimating confidence intervals with faster convergence is the so called bias corrected and accelerated (BC_a) bootstrap. This approach is also based on the sample percentiles of the bootstrap samples, but with different percentiles, defined in Efron & Tibshirani (1993) as

$$(\hat{I}_{lo}^{BC_a}, \hat{I}_{up}^{BC_a}) = (\hat{\eta}_{\zeta_1}^*, \hat{\eta}_{\zeta_2}^*)$$

where

$$\zeta_1 = \Phi\left(\hat{b}_0 + \frac{\hat{b}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{b}_0 + z^{\alpha})}\right)$$

$$\zeta_2 = \Phi\left(\hat{b}_0 + \frac{\hat{b}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{b}_0 + z^{1-\alpha})}\right)$$

where z^α is the 100α percentile of a standard gaussian distribution, while $\Phi()$ is the standard gaussian cumulative distribution function. The two remaining constants $\hat{b}_0$ and $\hat{a}$ are the estimated bias correction and acceleration. The bias correction is defined in Efron & Tibshirani (1993) as

$$\hat{b}_0 = \Phi^{-1}\left(\frac{\sum_{i=1}^{B} \mathbb{1}(\hat{\eta}_i^* < \eta)}{B}\right)$$

where η is the original point estimate without any resampling or bootstrap samples. Thus, $\hat{b}_0$ is the inverse of the standard gaussian cumulative distribution function applied to the proportion of bootstrap estimates exceeding the original estimate, relating $\hat{b}_0$ to the bias of $\hat{\eta}^*$.

The acceleration $\hat{a}$ is based on jack-knife estimates of $\hat{\eta}$, which requires some notation. Let $x_{(i)}$ denote the original data set with observation i removed. We have interpreted this as removing the ith individual in the cohort, and calculating the relative survival estimate using the smoothed, non-resampled population mortality rates. Denote the estimator t and let $\hat{\eta}_{(i)} = t(x_{(i)})$, and $\hat{\eta}_{(.)} = \sum_{j=1}^{n} \hat{\eta}_{(j)}$ defined in Efron & Tibshirani (1993) as

$$\hat{a} = \frac{\sum_{i=1}^{n}(\hat{\eta}_{(.)} - \hat{\eta}_{(i)})^3}{6(\sum_{i=1}^{n}(\hat{\eta}_{(.)} - \hat{\eta}_{(i)})^2)^{3/2}}$$

A confidence interval constructed with BC_a can be shown to be second order accurate, i.e. the error decreases as $1/n$. Efron & Tibshirani (1993, p. 188) states that "BC_a intervals are recommended for general use", but due to the difficulties with accounting for the twofold dataset situation, i.e. one cohort and one for resampled population mortality rates, we will also provide percentile confidence intervals for comparisons. For further motivation of the bias and acceleration constants, we direct the reader to Efron & Tibshirani (1993).

A final matter to discuss is how many bootstrap samples we should use, i.e. the size of B. Efron & Tibshirani (1993) suggests 2000 bootstrap samples as a rule of thumb for estimating confidence intervals. We compared two sets of estimates, from 2500 bootstrap samples and found similar results, suggesting that convergence was reached, but decided to use the full set of 5000 samples.

5 Data description and modelling

In our simulations we use a cohort from an European cancer registry, containing all patients diagnosed with colon cancer from 1975 to 1994 in one specific country. The survival times, i.e. the dates of the data set have been permuted to protect personal integrity of the individuals, and this is also why we do not state the origins of the data set explicitly. Individuals of age higher than 99 or lower than 15 have been excluded, as the number of such individuals were few.

The resulting cohort is followed up to 1995 and consists of 15 563 individuals in total, 6 339 males and 9 224 females. As this gives us at most 20 years of follow up, we decided to estimate 1, 5 and 10 year relative survival. When age stratification is used, we use age groups of 15-44, 45-59, 60-74 and 75-99 years of age, in which we categorize individuals according to their age at entry, i.e. cancer diagnosis. The distribution of individuals in age groups is presented in table 1.

Table 1: Age distribution of the colon cancer cohort

Frequency	Age group
734	15-44
2 368	45-59
6 593	60-74
5 868	75-99
15 563	Total

For population mortality, we have used data from the Human Mortality Database matching the country of the cohort. The population mortality is not based on a yearly comprehensive survey including each and every person in the specific contry, as this would be cumbersome. Nor do we expect exact dates of immigration, deaths and births to be completely accurate on an individual level basis. This contributes to treating population mortality as a random quantity, even though the estimates are based on the entire population. For further details regarding the population mortality data, we refer to `http://www.mortality.org/`.

5.1 Modelling population mortality rates

From here, we assume population mortality rates to follow a piecewise constant hazard rate. This corresponds to survival times having an exponential distribution, and the likelihood from this model is proportional to the likelihood of a Generalized Linear Model for the Poisson family, see for instance Royston & Lambert (2011). From this, we use that the fitted coefficients of the model follow an asymptotical normal distribution and sample new resampled coefficients based on this normal distribution. Predictions based on these resampled coefficients are then used as another realization of the underlying population mortality process, i.e. we resample population mortality rates using the variance of the fitted coefficients.

We are interested in population mortality rates matching the cancer cohort. But we still consider rates of years close to those present in the cohort data useful for estimating the variance of rates at boundary years in the cohort, e.g. 1975 and 1995. Thus we include a five year margin, using years from 1970 to 2000 to fit our population mortality rate model.

As we expect a smooth underlying process, it is natural to use splines for age and calendar year. The sex variable is included as a binary variable, and consider which number of knots to use. Royston & Lambert suggest to inspect

Akaike information criterion (AIC) and Bayes information criterion (BIC) as a guideline for deciding the number of knots in spline models, i.e. the degrees of freedom. We inspect these two measures of fit for 1-9 degrees of freedom in tables 2 and 3.

Table 2: AIC for different degrees of freedom for year (vertical) and age (horizontal), subtracted with minimum AIC: 14.58

Year/Age	1	2	3	4	5	6	7	8	9
1	4.28	1.33	0.72	0.57	0.55	0.53	0.52	0.51	0.51
2	4.28	1.33	0.72	0.57	0.55	0.53	0.52	0.51	0.51
3	3.83	0.90	0.29	0.14	0.12	0.10	0.09	0.08	0.08
4	3.87	0.93	0.33	0.17	0.15	0.13	0.12	0.11	0.11
5	3.77	0.84	0.24	0.08	0.06	0.04	0.03	0.03	0.02
6	3.78	0.85	0.24	0.09	0.07	0.05	0.04	0.03	0.03
7	3.76	0.83	0.22	0.07	0.05	0.03	0.02	0.01	0.01
8	3.77	0.84	0.24	0.08	0.06	0.04	0.03	0.02	0.02
9	3.75	0.82	0.21	0.06	0.04	0.02	0.01	0.00	0.00

Table 3: BIC for different degrees of freedom for year (vertical) and age (horizontal), subtracted with minimum BIC: 10622

Year/Age	1	2	3	4	5	6	7	8	9
1	24063	7407	3983	3110	3011	2907	2863	2822	2815
2	24063	7407	3983	3110	3011	2907	2863	2822	2815
3	21517	4983	1574	700	602	498	454	414	406
4	21710	5167	1756	883	784	680	636	596	589
5	21195	4672	1268	395	297	192	148	108	101
6	21232	4707	1302	429	331	227	182	142	134
7	21126	4606	1202	330	232	127	83	46	36
8	21205	4681	1276	403	305	201	156	117	109
9	21083	4569	1166	294	196	92	47	8	0

Following the AIC or BIC-approach would have resulted in 8-9 degrees of freedom, which could have increased had we considered even higher options in our grid search. However, such relatively high degrees of freedom result in rapidly changing estimates of the population mortality rates, which we find unrealistic. Instead, visual inspection was conducted to find a model with as many degrees of freedom as possible, which still produced a visually smooth fit. This resulted in 4 degrees of freedom for age, and 2 degrees of freedom for calendar year. Some support for the choice of 4 degrees of freedom for age can be found in these tables, as there seems to be a drop in the fourth column of both the AIC- and BIC-table, but more than 2 degrees of freedom for year gives an implausibly volatile curve.

Once the main effects of the population mortality rate are decided, pairwise interactions of sex, age and calendar year were found to be significant using

likelihood ratio tests. As sample sizes are large for the population data, we have much power and these significances might be of small practical importance. However, after visual inspection, we decided to include the interactions in the model with two degrees of freedom for each interaction spline. The degrees of freedom were decided on from visual inspection, one degree of freedom for interactions introduced some bias and seemed unrealistically simple, while three produced implausible effects in the outer regions of the fit. The use of fewer degrees of freedom for interaction splines compared to the main effects is suggested in Royston & Lambert (2011) to produce a smooth fit. The final model is thus, in pseudo-code:

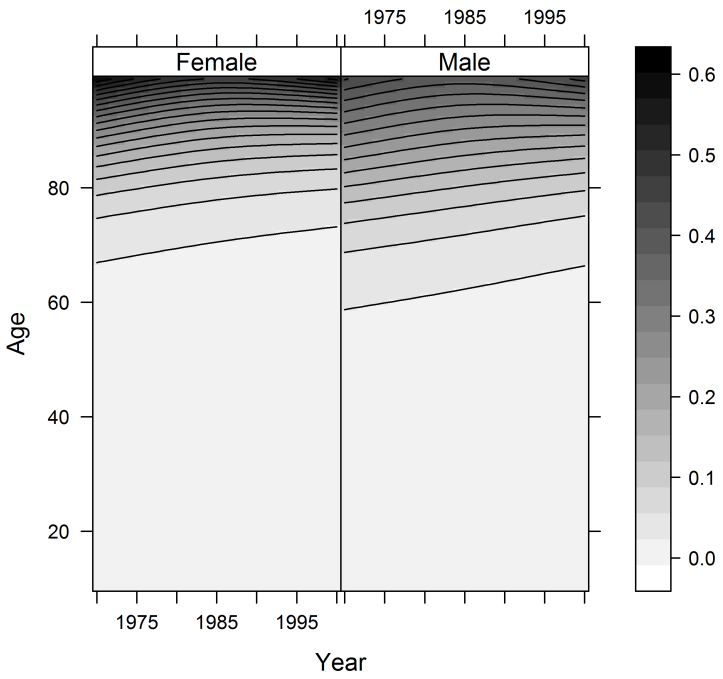
$$\log(rate) = s_4(age) + s_2(year) + sex \tag{13}$$

$$+s_2(age^*year) + s_2(age^*sex) + s_2(year^*sex),$$

where $s_{df}(x)$ denotes a spline of x, with df degrees of freedom, and star between two variable denotes an interaction.

We inspect the fit in a level plot of the predicted rates from the population mortality rate model in figure 1.

Figure 1: Level plot of the predicted rates from the population mortality rate model

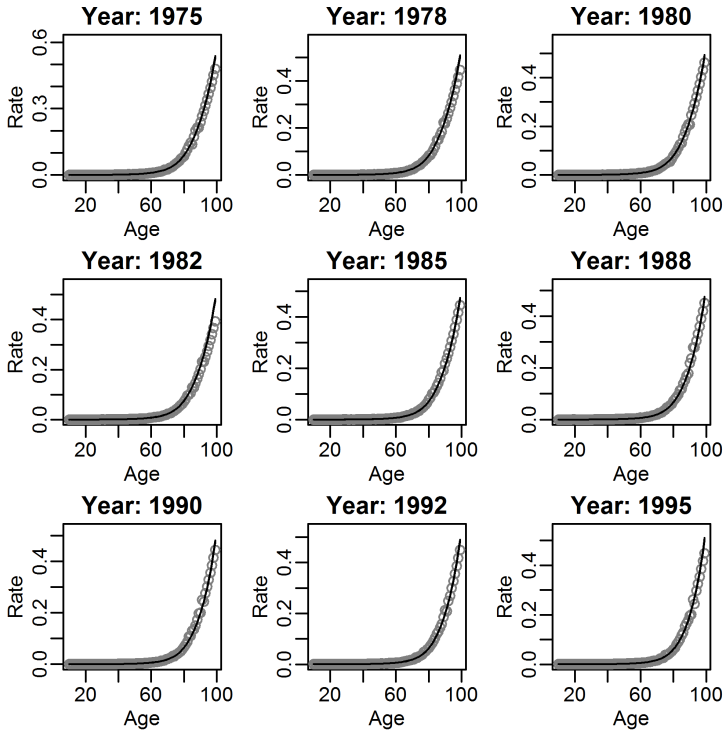


In figure 1 we see that the males reach a higher mortality rate at younger ages

compared to females. The impact of year in the resulting fit causes a lowering of rates as year increases, but age seem to have a large effect as well.

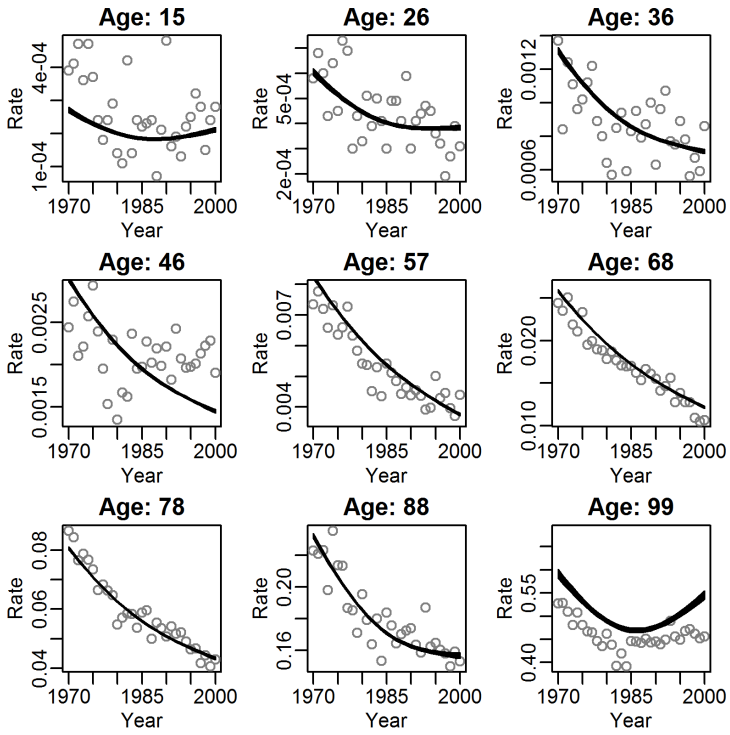
To visualize the magnitude of variance we introduce when resampling the coefficients of the population mortality rate model, we plot the fit for a subset of years and ages, for females together with the raw rates and 200 resampled predicted rates in figure 2. For corresponding plots for the male population, we refer readers to the Appendix. Note that we use different scales for each subplot, based on the dispersion of the raw rates, i.e. the dots.

Figure 2: Population mortality rates for females across age. The black lines are 200 resampled predictions, dots are the raw, unsmoothed rates.



From this plot, we note that the model, with resamples, sometimes overshoot the raw rates slightly, e.g. in year 1982. We give the same plot across years in figure 3.

Figure 3: Population mortality rates for males across calendar year. The black lines are 200 resampled predictions, dots are the raw, unsmoothed rates.



In figure 3 we also notice some bias, e.g. for the early years of age 15 and the entire 99-year old fit is slightly higher than the observed rates, and especially so for later years. However, as the purpose of this model is to introduce variance in resampled mortality rates rather than finding an unbiased fit of the raw rates, we do not find this bias problematic.

When it comes to assessing the variability in resampled rate predictions one could argue that the resampled rates do not come close to covering the the raw rates across years. Hence the variance is too small. On the other hand, one could argue that each age-year-sex-combination should be modelled separetely, which would give far less variance, as each pointwise estimate is based on relatively large sample sizes. We argue that the first point of view ignores that observations of the underlying random process will include more noise than the true mean of the underlying process i.e. a prediction interval is not the same as a confidence interval of the mean. The latter point of view ignores correlation between rates, i.e. we expect rates close in time and age to be observations of a similar mean, and thus we argue that they should be modelled jointly. The variance we want to incorporate is the uncertainty in the estimates of the underlying random process from the observed raw rates.

The population of the cancer registry country was approximately 5 million

during observation time. As there are countries with smaller populations, it would be interesting to see whether estimates in such countries are more uncertain, as we will have less certain population mortality estimates there, due to smaller sample sizes. One way to assess this is to scale down the original population mortality data. We do this by dividing the number of observed deaths and at risk with 2 and 10, for one 2.5 and one 0.5 million population. The corresponding population mortality rate predictions will be identical up to rounding, but we will increase the variance. One could argue both for and against changing the cohort size accordingly. After repeating the simulations using a random subsample of 10 % of the cohort and noting differences that were hard to explain from an increased variance, we decided to keep the cohort exactly the same during all three country-size scenarios. Although this is perhaps not the most clinically reasonable scenario, i.e. the cohort is ten times too large, the resulting comparison is restricted to the increase in variance of population mortality rates and not from differences in the cohorts. This might also be of interest if population mortality data is not available for more than a subset of the entire population. We illustrate the magnitude of the introduced variance for the three population scenarios, by plotting 200 resampled rates from ages 15, 42 and 99 for the three scenarios in figure 4.

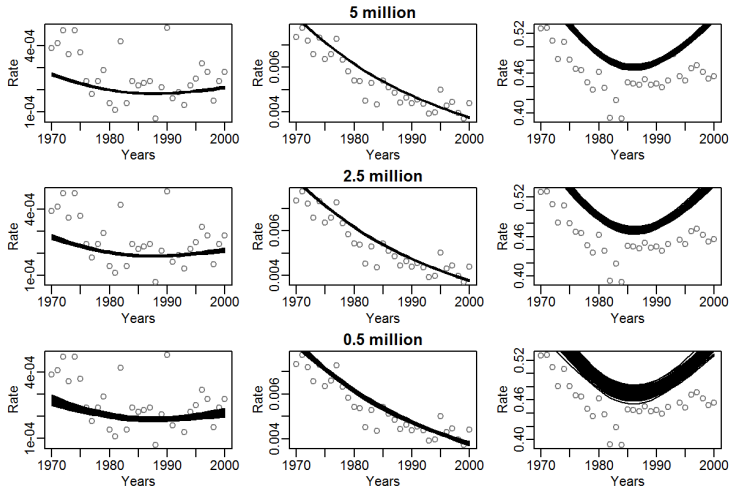


Figure 4: Population mortality rates for females across ages 15, 42 and 99 years from left to right. The black lines are 200 resampled predictions, dots are the raw, unsmoothed rates. The first row is the 5 million scenario, second row is 2.5 million, and third row is the 0.5 million scenario.

In figure 4 we note that there is an increase in variance, clearly visible for the 0.5 million scenario compared to the other two scenarios.

6 Results & Discussion

We begin with standard errors in the Rubin's Rule approach, and continue with a bootstrap-approach for confidence intervals. As we expect Rubin's rule to underestimate the variance, we only use the three population-size-scenarios in the bootstrap.

6.1 Variance estimates using Rubin's Rule

We begin with plotting the variance estimates from Rubin's Rule together with the analytic, standard versions.

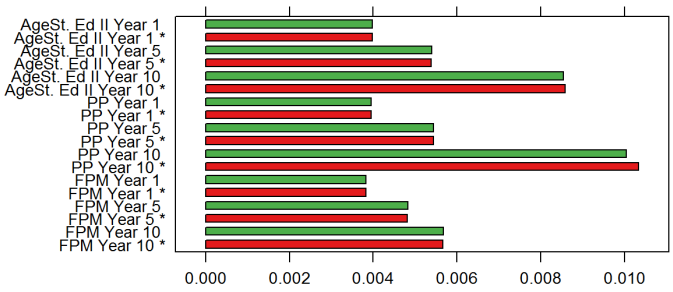


Figure 5: Plot of standard error estimates. AgeSt. Ed II, PP and FPM denotes internally age standardized Ederer II, Pohar Perme and Flexible Parametric Models in respective order, where star $*$ (and red colour) denotes Rubin's Rule-estimate, and no star (in green colour) denotes the analytic, assuming no variance in population mortality, estimate. Estimates are on the survival scale, and Rubin's Rule estimates are based on 200 resamples.

We note that the only visible difference is for the 10 year Pohar Perme estimate. We inspect the estimates in table 4.

Description	Estimate
AgeSt. Ed II Year 1	0.00399
AgeSt. Ed II Year 1 *	0.00398
AgeSt. Ed II Year 5	0.00540
AgeSt. Ed II Year 5 *	0.00539
AgeSt. Ed II Year 10	0.00854
AgeSt. Ed II Year 10 *	0.00858
PP Year 1	0.00395
PP Year 1 *	0.00395
PP Year 5	0.00544
PP Year 5 *	0.00544
PP Year 10	0.01004
PP Year 10 *	0.01033
FPM Year 1	0.00383
FPM Year 1 *	0.00383
FPM Year 5	0.00482
FPM Year 5 *	0.00481
FPM Year 10	0.00568
FPM Year 10 *	0.00566

Table 4: Estimates of standard error using analytical results, assuming no variance in population mortality compared to Rubin's Rule (marked with star). AgeSt. Ed II, PP and FPM denotes internally age standardized Ederer II, Pohar Perme and Flexible Parametric Model estimates in respective order. Estimates are on the survival scale and Rubin's Rule estimates are based on 200 resamples.

In table 4 we find that the only increase of standard error in the fourth decimal, when including uncertainty from population mortality, is for the 10 year Pohar Perme estimate. The increase is 0.00029 (3 %). We consider the smaller differences in the fifth decimal as artificial covariances in the resampling algorithm, and as indicated by figure 5, the variability relative to the interval length is small. To see an increase for the Pohar Perme estimate suggests that the population mortality uncertainty might at least have some effect. We continue to the bootstrap analysis.

6.2 Confidence interval estimates using Bootstrap

6.2.1 5 million - scenario

The results for the Ederer II estimator are seen in figure 6.

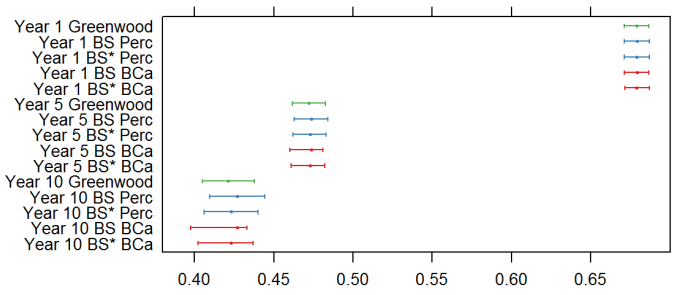


Figure 6: Confidence intervals for 1, 5 and 10 year age standardized Ederer II estimates. BS and BS denote bootstrap with single population mortality rate and bootstrap with resampled population rates, in respective order. Greenwood, BCa and Perc denotes which method was used for the confidence intervals, as indicated by the colours. Center is either the Ederer II estimate, or the sample mean of the bootstrap samples.*

From figure 6 we note that the estimates seem to be slightly shifted depending on how we construct the confidence interval. The point estimates differs slightly depending on whether we have used analytical results or bootstrap, and the point estimates also differs between the bootstraps, with a shift towards lower values when we include population mortality uncertainty. When it comes to the confidence interval endpoints, it is not surprising to see that the BC_a-intervals are shifted compared with the percentile intervals, in the direction of the analytic point estimate, as that is the effect of bias correction. We inspect the intervals closer in table 5.

Description	Lower	Center	Upper	Length
Year 1 Greenwood	0.6711	0.6789	0.6867	0.0156
Year 1 BS Perc	0.6713	0.6791	0.6869	0.0156
Year 1 BS* Perc	0.6712	0.6789	0.6868	0.0156
Year 1 BS BCa	0.6711	0.6791	0.6867	0.0156
Year 1 BS* BCa	0.6714	0.6789	0.6868	0.0155
Year 5 Greenwood	0.4617	0.4723	0.4828	0.0210
Year 5 BS Perc	0.4631	0.4738	0.4841	0.0210
Year 5 BS* Perc	0.4622	0.4728	0.4832	0.0210
Year 5 BS BCa	0.4601	0.4738	0.4812	0.0212
Year 5 BS* BCa	0.4610	0.4728	0.4821	0.0210
Year 10 Greenwood	0.4048	0.4214	0.4379	0.0332
Year 10 BS Perc	0.4095	0.4270	0.4442	0.0347
Year 10 BS* Perc	0.4061	0.4232	0.4402	0.0341
Year 10 BS BCa	0.3975	0.4270	0.4330	0.0356
Year 10 BS* BCa	0.4023	0.4232	0.4370	0.0347

Table 5: Confidence intervals for 1, 5 and 10 year age standardized Ederer II estimates. BS and BS* denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Greenwood, BCa and Perc denotes which method was used for the confidence intervals. Center is either the Ederer II estimate, or the sample mean of the bootstrap samples.

From table 5 we note that the bootstrap confidence intervals are slightly shorter in the 10 year estimate after including uncertainty from population mortality, 0.0006 (1.8 %) for the percentile interval, while the 1 and 5 year confidence interval lengths are similar. If anything we had expected an increase in variance, suggesting that the expected increase is dominated by random variation. We will return to this in our discussion. We also note that the 10 year bootstrap point estimates decrease by 0.0038 when including uncertainty from population mortality, and that this shift corresponds to 11 % of the percentile confidence interval length. Another observation is that the percentile and BC_a-interval lengths differ slightly, with longer confidence interval lengths for the BC_a-intervals. We move on to the Pohar Perme estimates in figure 7.

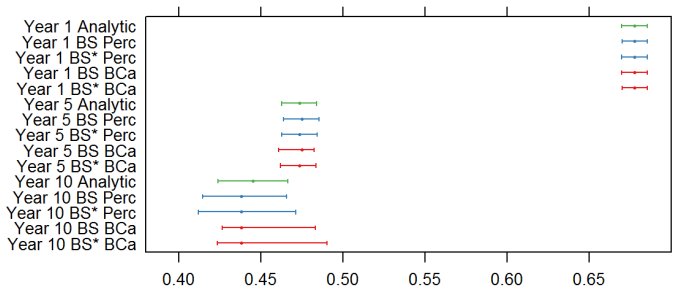


Figure 7: Confidence intervals for 1, 5 and 10 year Pohar Perme estimates. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), BCa and Perc. denotes which method was used for the confidence intervals, as indicated by the colours. Center is either the Pohar Perme estimate, or the sample mean of the bootstrap samples.*

From figure 7 we see that for the Pohar Perme estimator, estimates seem similar apart from the 10 year estimate where we note an increasing confidence interval length as we include population mortality uncertainty, for both bootstrap confidence interval methods. We give the details in table 6.

Description	Lower	Center	Upper	Length
Year 1 Analytic	0.6699	0.6777	0.6854	0.0155
Year 1 BS Perc	0.6701	0.6778	0.6857	0.0156
Year 1 BS* Perc	0.6700	0.6777	0.6855	0.0156
Year 1 BS BCa	0.6699	0.6778	0.6855	0.0156
Year 1 BS* BCa	0.6701	0.6777	0.6856	0.0155
Year 5 Analytic	0.4627	0.4734	0.4840	0.0213
Year 5 BS Perc	0.4640	0.4749	0.4855	0.0215
Year 5 BS* Perc	0.4628	0.4737	0.4843	0.0215
Year 5 BS BCa	0.4608	0.4749	0.4827	0.0220
Year 5 BS* BCa	0.4620	0.4737	0.4838	0.0218
Year 10 Analytic	0.4238	0.4452	0.4663	0.0425
Year 10 BS Perc	0.4146	0.4382	0.4656	0.0510
Year 10 BS* Perc	0.4121	0.4382	0.4712	0.0591
Year 10 BS BCa	0.4266	0.4382	0.4831	0.0565
Year 10 BS* BCa	0.4237	0.4382	0.4905	0.0669

Table 6: Confidence intervals for 1, 5 and 10 year Pohar Perme estimates. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), BCa and Perc denotes which method was used for the confidence intervals. Center is either the Pohar Perme estimate, or the sample mean of the bootstrap samples.*

In table 6 we note that most estimates are similar regardless of estimation procedure, expect for the 10 year estimate. There we see an increase of confidence interval length when including variance in population mortality rates of 0.0081 (16 %) for the percentile method and 0.0104 (18 %) for the BC_a-method, which could be considered quite similar. If we compare the percentile interval to the analytical confidence interval we find an increase of 0.0166 (39 %). But as we saw for the Ederer II estimates, the confidence interval lengths for the BC_a-intervals is often slightly larger compared to their percentile counterparts. We move on to the Flexible Parametric Model confidence intervals, seen in figure 8.

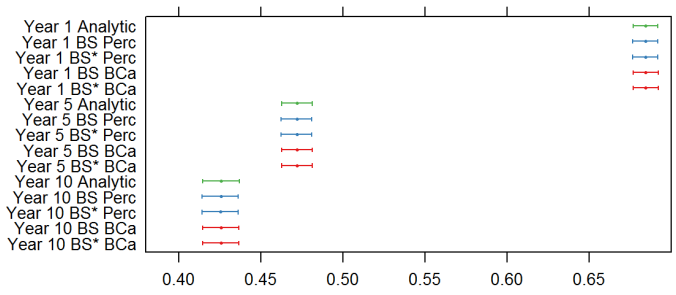


Figure 8: Confidence intervals for 1, 5 and 10 year Flexible Parametric Model estimates. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), BCa and Perc denotes the method used for the confidence intervals. Center is either the Flexible Parametric Model estimate, or the sample mean of the bootstrap samples.*

In figure 8 we do not see any differences between the different estimation procedures. We note that the confidence intervals does not seem to increase as much in variance for the 5 and 10 year estimates, as the non-parametric estimators. We present the confidence intervals in table format in table 7.

Description	Lower	Center	Upper	Length
Year 1 Analytic	0.6768	0.6844	0.6918	0.0150
Year 1 BS Perc	0.6767	0.6844	0.6920	0.0153
Year 1 BS* Perc	0.6767	0.6844	0.6920	0.0152
Year 1 BS BCa	0.6769	0.6844	0.6921	0.0152
Year 1 BS* BCa	0.6769	0.6844	0.6921	0.0152
Year 5 Analytic	0.4626	0.4721	0.4815	0.0189
Year 5 BS Perc	0.4624	0.4719	0.4811	0.0188
Year 5 BS* Perc	0.4623	0.4719	0.4811	0.0188
Year 5 BS BCa	0.4627	0.4719	0.4815	0.0188
Year 5 BS* BCa	0.4627	0.4719	0.4815	0.0188
Year 10 Analytic	0.4147	0.4258	0.4369	0.0222
Year 10 BS Perc	0.4143	0.4256	0.4364	0.0221
Year 10 BS* Perc	0.4143	0.4255	0.4364	0.0221
Year 10 BS BCa	0.4146	0.4256	0.4367	0.0221
Year 10 BS* BCa	0.4146	0.4256	0.4367	0.0221

Table 7: Confidence intervals for 1, 5 and 10 year Flexible Parametric Model estimates. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), BCa and Perc denotes which method was used for the confidence intervals. Center is either the Flexible Parametric Model estimate, or the sample mean of the bootstrap samples.*

We note that the two bootstrap estimates seem to produce similar confidence interval lengths here.

6.2.2 2.5 million - scenario

For the 2.5 million scenario we omit tables and figures, as they are very similar to the 5 million estimates. However, we do pick out two small details that we consider to be of interest. The width of the 2.5 million, 10 year Pohar Perme bootstrap percentile confidence interval, including the population unertainty increases with 0.0001 from 0.0591 in the 5 million scenario, to 0.0592 in the 2.5 million scenario. The corresponding BC_a interval does also increase with 0.0006 from 0.0669 to 0.0675. We will return to this in the discussion.

6.2.3 0.5 million - scenario

We continue with the last scenario, a population size of 0.5 million.

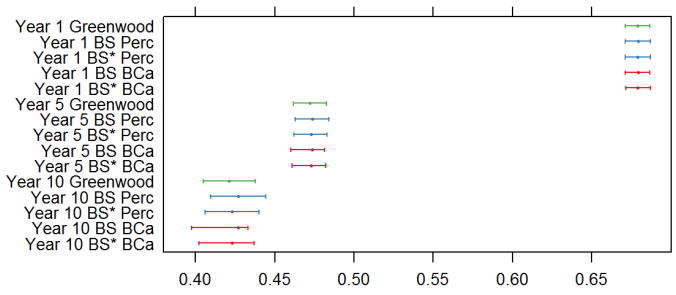


Figure 9: Confidence intervals for 1, 5 and 10 year age standardized Ederer II estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Greenwood (under assumption of negligible population mortality variance), BCa and Perc denotes the method used for the confidence intervals. Center is either the Ederer II estimate, or the sample mean of the bootstrap samples.*

From figure 9 we note that the impact of the smaller overall population seem negligible, the overall impression is similar to the previous scenarios. From table 8 we once again note that the 10 year interval decrease slightly in length, after including population mortality uncertainty.

Description	Lower	Center	Upper	Length
Year 1 Greenwood	0.6711	0.6790	0.6867	0.0156
Year 1 BS Perc	0.6713	0.6791	0.6869	0.0156
Year 1 BS* Perc	0.6712	0.6790	0.6868	0.0156
Year 1 BS BCa	0.6711	0.6791	0.6867	0.0157
Year 1 BS* BCa	0.6714	0.6790	0.6869	0.0155
Year 5 Greenwood	0.4618	0.4723	0.4828	0.0210
Year 5 BS Perc	0.4631	0.4738	0.4841	0.0210
Year 5 BS* Perc	0.4622	0.4729	0.4832	0.0210
Year 5 BS BCa	0.4601	0.4738	0.4813	0.0212
Year 5 BS* BCa	0.4611	0.4729	0.4821	0.0210
Year 10 Greenwood	0.4048	0.4214	0.4380	0.0332
Year 10 BS Perc	0.4095	0.4270	0.4442	0.0347
Year 10 BS* Perc	0.4060	0.4233	0.4403	0.0343
Year 10 BS BCa	0.3975	0.4270	0.4331	0.0356
Year 10 BS* BCa	0.4023	0.4233	0.4370	0.0346

Table 8: Confidence intervals for 1, 5 and 10 year age standardized Ederer II estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Greenwood (under assumption of negligible population mortality variance), BCa and Perc denotes the method used for the confidence intervals. Center is either the Ederer II estimate, or the sample mean of the bootstrap samples.*

The decrease for the 10 year estimate, after including uncertainty from population mortality is 0.001(2.9 %). We move on to the Pohar Perme estimator in figure 10.

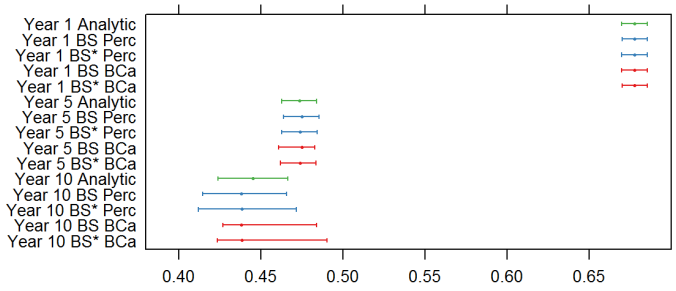


Figure 10: Confidence intervals for 1, 5 and 10 year Pohar Perme estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), BCa and Perc denotes the method used for the confidence intervals. Center is either the Pohar Perme estimate, or the sample mean of the bootstrap samples.*

From figure 10 we once again note an increased variance of the 10 year estimate after including variance of the population mortality rates. We inspect table 9 for more details.

Description	Lower	Center	Upper	Length
Year 1 Analytic	0.6699	0.6777	0.6854	0.0155
Year 1 BS Perc	0.6701	0.6778	0.6857	0.0156
Year 1 BS* Perc	0.6700	0.6777	0.6855	0.0156
Year 1 BS BCa	0.6699	0.6778	0.6856	0.0156
Year 1 BS* BCa	0.6701	0.6777	0.6856	0.0155
Year 5 Analytic	0.4627	0.4734	0.4840	0.0213
Year 5 BS Perc	0.4640	0.4749	0.4855	0.0215
Year 5 BS* Perc	0.4628	0.4738	0.4844	0.0216
Year 5 BS BCa	0.4608	0.4749	0.4828	0.0220
Year 5 BS* BCa	0.4621	0.4738	0.4837	0.0217
Year 10 Analytic	0.4239	0.4453	0.4665	0.0425
Year 10 BS Perc	0.4146	0.4382	0.4656	0.0510
Year 10 BS* Perc	0.4121	0.4384	0.4716	0.0595
Year 10 BS BCa	0.4268	0.4382	0.4839	0.0571
Year 10 BS* BCa	0.4236	0.4384	0.4905	0.0668

Table 9: Confidence intervals for 1, 5 and 10 year Pohar Perme estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), Perc and BCa denote the method used for the confidence intervals. Center is either the Pohar Perme estimate, or the sample mean of the bootstrap samples.*

From table 9, we find that the 10 year estimate increase is 0.0097 (17 %) for the BC_a-interval, which is slightly less compared to the other scenarios, but well within the variability of the fourth decimal which we have seen throughout the different scenarios. For the 10 year percentile interval we note that the BS^*-interval has increased with 0.0003 from 0.0592 to 0.0595 compared to the 2.5 million scenario, while corresponding BC_a interval has decreased from 0.0675 in the 2.5 million scenario to 0.0668. We note that for the 10 year estimates, the increase in confidence interval length when comparing with and without population mortality uncertainty is 0.0085 (17 %) for the percentile interval which is slightly more than for the previous scenarios. We finish with the Flexible Parametric Model estimates in figure 11.

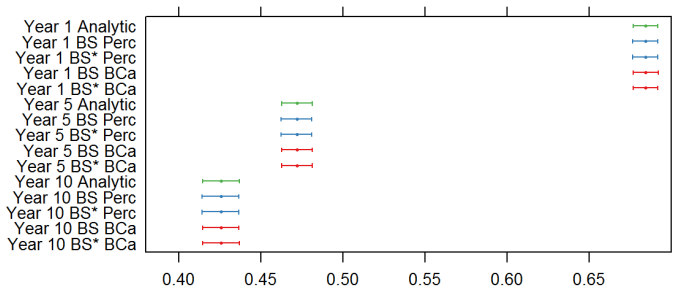


Figure 11: Confidence intervals for 1, 5 and 10 year Flexible Parametric Model estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), Perc and BCa denote the method used for the confidence intervals. Center is either the Flexible Parametric Model estimate, or the sample mean of the bootstrap samples.*

In figure 11 we see, as earlier, that the Flexible Parametric Models seem stable due to variability from population mortality rates. Once again we note that the estimates does not seem to increase as time increases, giving smaller confidence intervals for the 5 and 10 year estimates, compared to the non-parametric estimators. We give the results in table 10.

Description	Lower	Center	Upper	Length
Year 1 Analytic	0.6768	0.6844	0.6918	0.0150
Year 1 BS Perc	0.6767	0.6844	0.6920	0.0153
Year 1 BS* Perc	0.6767	0.6844	0.6919	0.0152
Year 1 BS BCa	0.6769	0.6844	0.6921	0.0152
Year 1 BS* BCa	0.6769	0.6844	0.6920	0.0152
Year 5 Analytic	0.4626	0.4721	0.4815	0.0189
Year 5 BS Perc	0.4624	0.4719	0.4811	0.0188
Year 5 BS* Perc	0.4624	0.4719	0.4812	0.0188
Year 5 BS BCa	0.4627	0.4719	0.4815	0.0188
Year 5 BS* BCa	0.4627	0.4719	0.4814	0.0187
Year 10 Analytic	0.4147	0.4258	0.4369	0.0222
Year 10 BS Perc	0.4143	0.4256	0.4365	0.0221
Year 10 BS* Perc	0.4142	0.4256	0.4365	0.0222
Year 10 BS BCa	0.4146	0.4256	0.4367	0.0221
Year 10 BS* BCa	0.4145	0.4256	0.4368	0.0223

Table 10: Confidence intervals for 1, 5 and 10 year Flexible Parametric Model estimates from a population of 0.5 million individuals. BS and BS denote bootstrap with single population rate and bootstrap with resampled population rates, in respective order. Analytic (under assumption of negligible population mortality variance), Perc and BCa denote the method used for the confidence intervals. Center is either the Flexible Parametric Model estimate, or the sample mean of the bootstrap samples.*

We note that for the 10 year estimates, there is a slight increase when including population mortality uncertainty, but comparing with the 1 and 5 year estimates indicate that this could be due to random variation, as interval lengths decrease slightly here, when including population mortality uncertainty.
For the upcoming discussion, we present a plot of a kernel density estimate from the 5000 bootstrap samples for the 5 million, 10 year estimates of Age Standardized Ederer II and Pohar Perme estimator.

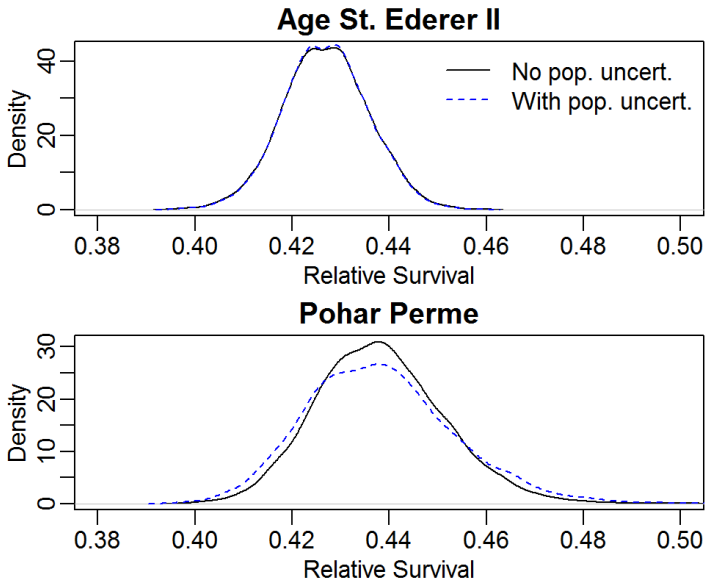


Figure 12: Kernel Density Estimates of bootstrap samples from Age Standardized Ederer II and Pohar Perme estimates, for the 5 million scenario, 10 year relative survival. Note that the dashed lines have been shifted onto the drawn curves in each subplot, to facilitate comparison of the shapes.

In figure 12 we note that the difference between the two Ederer II curves is found in the top of the density, where the bootstrap sample which included population uncertainty (dashed) is slightly bimodal. We suggest that the observed decrease in variance might be due to some artifact from the cohort, as we would have expected an overall lowering of the dashed curve if the variance did really differ, instead the shapes are rather similar apart from the bimodal peak. For the Pohar Perme estimator, the dashed curve clearly has a larger variability compared to the solid curve.

6.3 A shift in point estimates for Ederer II

We will now suggest an explanation to a shifting effect of point estimates in bootstrap estimates for the Ederer II-estimator, when uncertainty from population mortality is introduced. That is we try to explain results commented on below table 5. First we note that the numerator, $S(t)$, is simulated to be identical in both bootstrap procedures, hence the shift must be due to differences in expected survival $S^*(t)$. Since the ratio estimate in the results was found to be smaller when including population uncertainty, the estimate of $S^*(t)$ must increase with the extra uncertainty. Recall that we have modelled the population mortality rates as described in (13). If we for simplicity consider a single person cohort, this individual will have the non-resampled rate $\tilde{\theta}^*$ in the bootstrap where only the cohort is resampled. If we let θ^* be the random variable which

we draw samples from when resampling, it holds that $\mathbb{E}(\theta^*) = \tilde{\theta}^*$. However, the denominator of the Ederer II estimator is used on the probability scale, i.e. $\exp(-\theta^*)$, where $f(x) = \exp(-x)$ is a convex function. According to Jensen's inequality, see Grimmett & Stirzaker (2001), it holds that

$$\mathbb{E}(f(\theta^*)) \geq f(\mathbb{E}(\theta^*)) = f(\tilde{\theta}^*),$$

with equality when θ^* is deterministic. Thus estimates including variance could be expected to have a downwards shift on the probability scale. We have not been able to expand this line of reasoning for the other two, more complex estimators, but on the other hand we don't see a consistent shift for those either.

6.4 Discussion

In this thesis we have investigated whether confidence interval length is increased when including uncertainty from population mortality rates, for three different estimators. What constitutes a substantial increase in confidence interval length depends on the situation. In our first attempt, using Rubin's Rule, we saw that a small increase of interval length seemed present for the 10 year Pohar Perme estimate. For corresponding bootstrap, we found the percentile confidence interval of the Pohar Perme 10 year estimate to be 39 % longer when taking population mortality uncertainty into account, compared to standard procedures. The same estimate has a 16 % confidence interval increase when comparing percentile bootstraps with and without uncertainty from population mortality rates. Even though 39 % is larger, we prefer to stress the bootstrap estimate of confidence interval increase of 16 %, as bootstrap and standard procedures are known to produce slightly different estimates. If one considers this to be a substantial increase, it holds that at least for one estimator, there is a non-negligible increase of confidence interval length when taking uncertainty from population mortality into account.

We have calculated both percentile and BC_a confidence intervals from our bootstrap estimates. For the BC_a procedure, we know that we are missing the variation from population mortality in the calculation of the acceleration constant. This suggests that percentile confidence intervals might be a better choice. From our simulations, the main difference in confidence interval length when the two procedures differ is that the BC_a intervals are slightly wider than the percentile counterparts. However, when it comes to assessing the effect of including population mortality uncertainty, both procedures give similar results.

We argue that finding the increase for the Pohar Perme estimator at the 10 year estimate is not surprising. The Pohar Perme estimator has more variance than Ederer II, and we have seen that both non-parametric estimators produce longer confidence intervals than the Flexible Parametric Models, indicating more variance in the nonparametric estimators. To find the increase at the 10 year estimate rather than the first year is not surprising, as estimates of later years are based on less data, as fewer survivors are present. This suggests that the result for Pohar Perme is reasonable.

However, we have also seen that some confidence intervals decrease after inclusion of population mortality uncertainty, with the most extreme case for the Age Standardized Ederer II estimate. The decrease was found to be 2.9 % for the 0.5 million scenario. From a stochastic point of view, it can be hard to draw the line between actual trend and random variation, but we suggest from figure 12 to consider the decrease as due to random variation.

Another unexpected observation is the downwards shift on the probability scale of Ederer II bootstrap point estimates when including uncertainty from population mortality. This shifts the Ederer II 10 year confidence interval 11 % of its length towards zero. Although we have given some support from Jensen's inequality for this not simply being random noise, we stress that it's unclear how large the theoretically supported shift is. Since we are not aware of any other studies discussing such a shift, we simply note that it might be something worth investigating closer in future research.

It is worthwhile to consider how the population mortality rates were resampled. Here we have used a model which smooths before we can draw resamples. Under the assumption of interchangability of rates close in age and calendar year, other less parametric resampling algorithms could have been used. For instance one could sample the rate of year x, age y and a specific gender from the raw rates of years $[x-a, x+a]$, years $[y-a, y+a]$ for that gender, where the sampling window range a would have been tuned by some criteria. However, we consider the asymptotic normality of the GLM coefficients as a reasonable way of modelling the uncertainty, under the assumption of an underlying smooth population mortality process.

If we return to the length of the confidence intervals, a natural follow up question is how the interval lengths increase with the uncertainty of the population mortality rates. This is a question of practical importance, as not all countries have population of the 5 million size, considered in our original data set. Thus, we have simulated populations of half and a tenth of the original size of 5 millions. Although the extra variance introduced before estimation is clearly visible, e.g. in figure 4 it is hard to claim that there is a visible effect after estimation. The percentile bootstrap confidence interval for the Pohar Perme 10 year estimate increase slightly as we introduce more variance throughout the three scenarios. However, the increases are too small (0.0001 and 0.0003) to be more than a possible indication. The BC_a confidence intervals shift as 0.0669 to 0.0675 to 0.0668 as population size is decreased. That is, the length is shortest for the 0.5 scenario, which casts some doubts on the validity of BC_a approach in this setting. The overall interpretation is that a population of 5 or 0.5 million does not seem to make a big difference.

However, we must stress that although we simulated population of decreased size, we did not decrease cohort size, as we wanted a clean comparison of the effect from decreasing population size. This does however complicate the interpretation, as the 0.5 million scenario has a cohort of 10 times the size one would expect from a country of that particular size. Thus some caution is required in interpretations.

From a methodological point of view, the challenges of this thesis are mostly

due to combining variance from two separate datasets into a single confidence interval. We see it in our analytical delta method for ratios approach, in the underestimation of variance in Rubin's Rule and in the flawed acceleration constant in the BC_a bootstrap. We would not be surprised if the likelihood approach in Flexible Parametric Models could allow cohort and population rates to be estimated jointly. But as mentioned earlier, we were interested in comparing non-parametric estimators as well, and have thus used the bootstrap procedure for all three estimators.

For the broad perspective, we might ask what the implication of this thesis is. Should estimation of relative survival continue to assume a negligible insecurity when calculating confidence intervals in relative survival? The results presented here suggest that for Pohar Perme estimates of 10 year survival, and naturally beyond that point, the assumption is hard to support. For the other estimators, our results suggest that this particular source of uncertainty is of small concern. But we have only considered one single cancer site, one single cohort and one single overall population. It is unclear whether these results will hold in other settings, and we stress the need for more research.

7 Appendix

7.1 Male population mortality rates

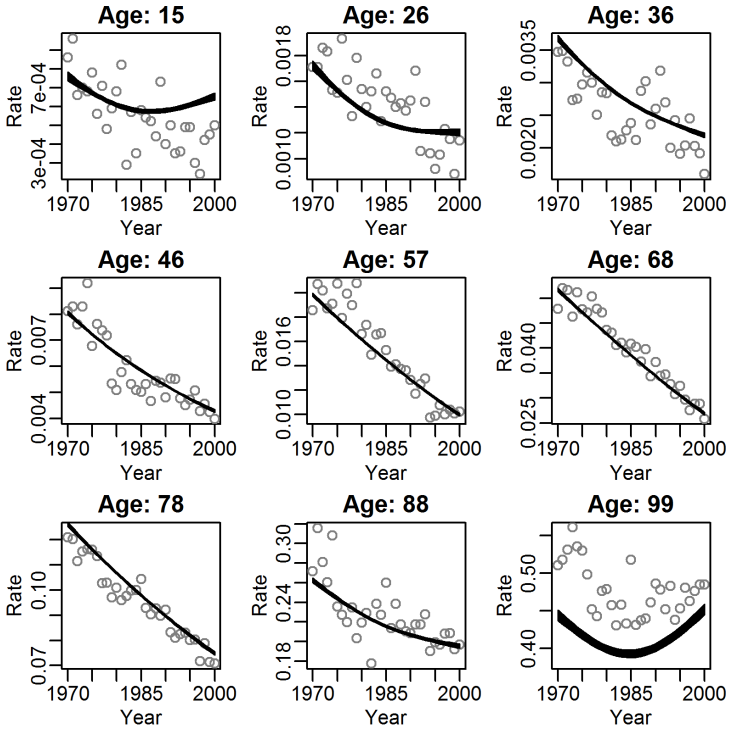


Figure 13: Population mortality rates for males across years. The black lines are 200 resampled predictions for the 5 million scenario, dots are the raw, unsmoothed rates.

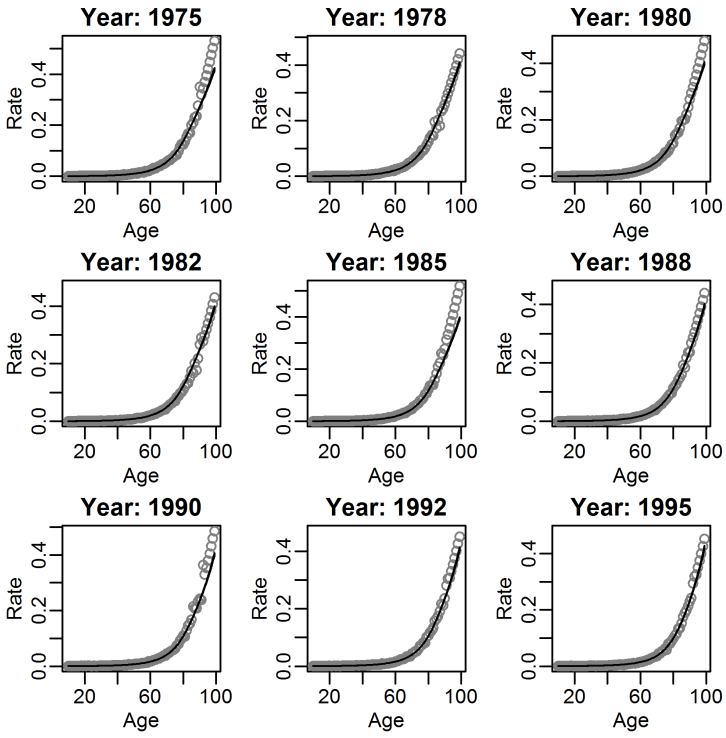


Figure 14: Population mortality rates for males across age. The black lines are 200 resampled predictions for the 5 million scenario, dots are the raw, unsmoothed rates.

References

Aalen, O., Borgan, O. & Gjessing, H. (2008). Survival and event history analysis: a point process point of view. Springer.

Andersen, P.K. & Vaeth, M. (1989). Simple Parametric and Nonparametric Models for Excess and Relative Mortality Source. Biometrics, Vol. 45, No. 2 (Jun., 1989), p. 523-535.

Berkson, J. (1942). The calculation of survival rates. In Carcinoma and Other Malignant Lesions of the Stomach, W. Walters, H. K. Gray and J. T. Priestley. p. 467-484.

Brenner, H. & Hakulinen, T. (2005) Substantial Overestimation of Standard Errors in Relative Survival Rates of Cancer Patients. American Journal of Epidemiology. 161. p. 781-786.

Buckley, J. D. 1984. Additive and Multiplicative Models for Relative Survival Rates. Biometrics, Vol. 40, No. 1 1984, pp. 51-62

Casella, G. & Berger, R.L. (2002). Statistical Inference. 2nd Edition.

Corazziari, I., Quinn, M. & Capocaccia, R. (2004). Standard cancer patient population for age standardising survival ratios. European Journal of Cancer 40. p. 2307-2316.

Cox, D. R., & Oakes, D. (1984). Analysis of Survival Data. Chapman and Hall.

Cutler S. J. & Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. Journal of Chronic Disorders 8, p. 699-712.

Dal, H. & Andersson, M. (2004). Cancerforskare, sluta redovisa konfidensintervall när det inte behövs! Läkartidningen 4, p. 314.

Dickman, P.W., Sloggett, A., Hills M. & Hakulinen, T. (2003) Regression models for relative survival. Stat. Med. 23, p. 51-64.

Dickman, P., Palmgren, J. & Pawitan, Y. (2004). Cancerforskare behöver statistisk osäkerhet! Läkartidningen 20, p. 1842

Dickman, P.W., Lambert, P.C., Coviello, E. & Rutherford, M.J. (2013). Estimating net survival in population-based cancer studies. Int. J. Cancer: 133, 519-522.

Dickman, P.W. & Coviello, E. (2015). Estimating and modeling relative survival. The Stata Journal 15:1, p. 186-215.

Ederer, F. & Heise, H. (1959). Instructions to IBM 650 programmers in processing survival computations. Methodological note no. 10. Bethesda, MD: End Results Evaluation Section, National Cancer Institute.

Ederer F., Axtell L.M. & Cutler S.J. (1961) The Relative survival rate: a statistical methodology. National Cancer Institute Monographs 6.

Efron, B. (1981). Censored Data and the Bootstrap. Journal of American Statistical Association. 76:374, p. 312-319.

Efron, B. & Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57. Chapman Hall

Fieller, E.C. (1954). Some Problems in Interval Estimation. Journal of the Royal Statistical Society. 16:2. p. 175-185.

Grimmett G.R. & Stirzaker. D.R. (2001). Probability and Random Processes. Oxford University Press. Third edition.

Hakulinen, T., Seppä, K. & Lambert, P.C. (2011) Choosing the relative survival method for cancer survival estimation. European Journal of Cancer 47. p. 2202-2210.

Kalbfleisch, J.D. & Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data, Second Edition. John Wiley & Sons.

Klein, J.P. & Moeschberger, M.L. (2003). Survival Analysis: Techniques for Censored and Truncated Data. Springer.

Klein, L.R. (1953). A Textbook of Econometrics. Prentice Hall, 2nd edition.

Lambert, P.C., Dickman, P.W. & Rutherford, M.J. (2015). Comparison of different approaches to estimating age standardized net survival. BMC Medical Research Methodology. 15:64.

Marshall, A., Altman, D.G., Holder, R.L. & Royston, P. (2009). Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Medical Research Methodology. 9:57.

Nelson, C.P., Lambert, P.C., Squire, I.B. & Jones, D.R. (2007). Flexible Parametric Models for Relative Survival, with Application in Coronary Heart Disease. Statistics in Medicine: 26. p. 5486-5498.

Parkin, D.M. & Hakulinen, T. (1991). Analysis of Survival. In: Cancer Registration: Principles and Methods. IARC Scientific Publication No. 95. Edited by Jensen, O.M., Parkin D.M., MacLennan, R., Muir C.S., and Skeet, R.G.

Pohar Perme, M., Stare, J. & Estève, J. (2012). On Estimation in Relative Survival. Biometrics: 68, p. 113-120.

Roche, L., Danieli, C., Belot, A., Grosclaude, P., Bouvier, A., Velten, M., Iwaz, J., Remontet, L. & Bossard, N. (2013). Cancer net survival on registry data: Use of the new unbiased Pohar-Perme estimator and magnitude of the bias with the classical methods. Int. J. Cancer: 132, 2359-2369.

Royston, P. & Parmar, M. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine: 21. p. 2175-2197.

Royston, P. & Lambert, P. (2011). Flexible parametric survival analysis using Stata. Stata Press.

Talbäck, M. & Dickman, P.W. (2011). Estimating expected survival probabilities for relative survival analysis - Exploring the impact of including cancer patient mortality in the calculations. European Journal of Cancer: 47, p. 2626-2632.

Data sources

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (Downloaded at 7th of February 2017).