Analysis of Relative Survival Patterns in Cancer Register Data

Jakobsen, Lasse Hjort

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

ANALYSIS OF RELATIVE SURVIVAL PATTERNS IN CANCER REGISTER DATA

BY
LASSE HJORT JAKOBSEN

DISSERTATION SUBMITTED 2018

AALBORG UNIVERSITY
DENMARK

Analysis of Relative Survival Patterns in Cancer Register Data

PhD Dissertation

Lasse Hjort Jakobsen

Dissertation submitted June, 2018

Dissertation submitted:	June 29, 2018
PhD supervisor:	Professor, Martin Bøgsted Department of Clinical Medicine Aalborg University, Denmark
Assistant PhD supervisors:	Clinical Associate Professor, Tarec C. El-Galaly Department of Clinical Medicine Aalborg University, Denmark
	Clinical Professor, Hans E. Johnsen Department of Clinical Medicine Aalborg University, Denmark
	Professor, Karen Dybkær Department of Clinical Medicine Aalborg University, Denmark
PhD committee:	Associate Professor Henrik Bøggild Aalborg University
	Professor Jacob von Bornemann Hjelmborg University of Southern Denmark
	Assistant Professor Matthew John Maurer College of Medicine, Mayo Clinic
PhD Series:	Faculty of Medicine, Aalborg University
Department:	Department of Clinical Medicine

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-221-4

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2018

Preface

This thesis was written during the period from July 1 2015 until June 30 2018 at the Department of Clinical Medicine, Aalborg University, and the Department of Hematology, Aalborg University Hospital. It consists of five scientific papers, a software package for the statistical programing language `R`, and an introduction to survival analysis and parametric relative survival models. The thesis is primarily intended for biostatisticians, epidemiologists, and clinicians as it includes both epidemiological and methodological studies.

During my time as a PhD student, I have met countless intelligent and extremely competent individuals, who have not only inspired me to pursue new ideas or optimize the ones already there, but also expanded my knowledge and encouraged me to continue researching. For this I am deeply grateful. However, I would like to point out a few people who have been vital to this project. First and foremost, I would like to thank my supervisor Martin Bøgsted for leading me into the world of hematological research and for giving me the opportunity to do this PhD. You have been extremely devoted to the project from the beginning and I have learned so much through our discussions over the years. My co-supervisor, Tarec El-Galaly, also deserves a special thanks for his extremely valuable supervision and for involving me in his projects right from the beginning. Also a special thanks to my co-supervisors, Karen Dybkær and the late Hans Erik Johnsen, for making the Hematological Research Laboratory a special place to work.

Thanks to my toothless partner in crime Jorne Biccler for vital, educational, entertaining, and, occasionally, odd discussions. It has been a pleasure working with you. I would also like to thank the rest of my office mates, Rasmus Brøndum and Mads Sønderkær for answering my occasionally trivial questions and for the good spirit. Also the rest of my colleagues at the Department of Hematology and the statistical group at Aalborg University Hospital deserve thanks. I would also like to express my gratitude towards Therese Andersson, Mark Clements, and the rest of the biostatistics department at Karolinska Institutet for welcoming me and making my visits comfortable and fruitful.

Additionally, I would like to thank Søren Vilsen and Johan Sørknæs for their friendship and for all the "hard" work we did together during our studies. Lastly I want to thank my friends and family, particularly Maja for her love and tolerance during the past nine years.

Lasse Hjort Jakobsen
Aalborg University, June 29, 2018

To Maja

Thesis outline

Thesis title:	Analysis of Relative Survival Patterns in Cancer Register Data
PhD student:	Lasse Hjort Jakobsen
Supervisor:	Professor, Martin Bøgsted, Aalborg University

This thesis consists of an introduction and the following five scientific papers, which can be found at the back of the thesis.

I) L.H. Jakobsen, M. Bøgsted, P.dN. Brown, B. Arboe, J. Jørgensen, T.S. Larsen, M.B. Juul, L. Schurmann, L. Højberg, O.J. Bergmann, T. Lassen, P.L. Josefsson, P. Jensen, H.E. Johnsen, and T.C. El-Galaly (2017). "Minimal loss of lifetime for patients with diffuse large B-cell lymphoma in remission and event free 24 months after treatment: a Danish population-based study". Journal of Clinical Oncology. 35(7):778–784.

II) L.H. Jakobsen, T.M-L. Andersson, J.L. Biccler, T.C. El-Galaly, and M. Bøgsted (2018). "Estimating the loss of lifetime function using flexible parametric relative survival models". Submitted to BMC Medical Research Methodology.

III) L.H. Jakobsen, T.M-L. Andersson, J. Biccler, L.Ø. Poulsen, M.T. Severinsen, T.C. El-Galaly, and M. Bøgsted (2018). "Estimation of time to statistical cure". Submitted to Statistics in Medicine.

IV) L.H. Jakobsen, M. Bøgsted, and M. Clements (2018). "Generalized parametric cure models". In preparation.

V) J.L. Biccler, T.C. El-Galaly, M. Bøgsted, J. Jørgensen, P.dN. Brown, C.B. Poulsen, J. Starklint, M.B. Juul, J.H. Christensen, P. Josefsson, A. Dessau, and L.H. Jakobsen (2018). "Clinical prognostic scores are poor predictors of overall survival in various types of malignant lymphomas". Submitted to Leukemia & Lymphoma.

In addition to the five articles, the following package for the statistical programming language `R` was developed.

VI) L.H. Jakobsen (2018): "**cuRe**: An R-package for parametric cure model estimation". `https://github.com/LasseHjort/cuRe`.

The following articles were also published during the course of the PhD, but are not considered a part of the thesis.

1. L.H. Jakobsen, M. Hutchings, P.dN. Brown, J. Linderoth, K.J. Mylam, D. Molin, H.E. Johnsen, M. Bøgsted, M. Jerkeman, and T.C. El-Galaly (2016). "No survival benefit associated with routine surveillance imaging for Hodgkin lymphoma in first remission: a Danish-Swedish population-based observational study". British Journal of Hematology. 173(2):236–244.

2. F. Ebbesen, P.H. Madsen, P.K. Vandborg, L.H. Jakobsen, T. Trydal, and H.J. Vreman (2016). "Bilirubin isomer distribution in jaundiced neonates during phototherapy with LED light centered at 497 nm (turquoise) vs. 459 nm (blue)". Pediatric Research. 80(4):511-515.

3. H. Cederleuf, L.H. Jakobsen, F. Ellin, P.dN. Brown, T.S. Larsen, M. Bøgsted, T. Relander, M. Jerkeman, and T.C. El-Galaly (2017). Outcome of peripheral T-cell lymphoma in first complete remission: a Danish-Swedish population-based study. Leukemia & Lymphoma. 58(12):2815-2823.

4. C.H. Nørgaard, L.H. Jakobsen, A. Gentles, K. Dybkær, T.C. El-Galaly, J.S. Bødker, A. Schmitz, P. Johansen, T. Herold, K. Spiekermann, J.R. Brown, J.L. Klitgaard, H.E. Johnsen, and M. Bøgsted (2018). "Subtype assignment of CLL based on B-cell subset associated gene signatures from normal bone marrow – A proof of concept study". PLOS ONE. 13(3):e0193249.

Abstract

Within cancer research population-based survival data are commonly used to identify clinical and molecular factors that are associated with the risk of death, progression, or similar end-points. As cancer patients may die from causes not related to cancer, it is sometimes of interest to quantify the patient survival relative to the survival of a relevant general population. From the time of diagnosis the patient mortality is typically elevated as compared to the general population, but for most cancer types the mortality difference decreases as patients remain alive. In paper I, we compare the conditional survival of diffuse large B-cell lymphoma patients in complete remission after first line R-CHOP(-like) therapy with the survival of the Danish general population.

As mortality risks can be quantified in various ways, survival trends and risk factors may be detected in multiple ways. The conditional number of life years lost due to the cancer, termed the *loss of lifetime* function, provides an intuitive measure of the disease severity. However, computing this function is generally difficult due to censoring. Paper II evaluates methods for estimating the loss of lifetime function through simulations and real-world data for which complete follow-up is available.

While cure is challenging to assess within cancer due to the risk of relapse, trends in the survival of cancer patients may be used to assess cure. In particular, if the patient mortality reaches the same level as the mortality of the general population, those patients who are still alive may be considered *statistically cured* of the disease. The time point at which this occurs is termed the *cure point* and provides useful information for health care planners as well as patients, particularly patients attending routine follow-up. However, its estimation is challenging for a number of reasons which are discussed in Paper III together with a new approach for estimating cure points.

Cure models have previously been introduced to model the proportion of patients who are statistically cured. The most applied models rely on either simple parametric survival distributions or spline-based distributions with the strict assumption of a finite cure point. Paper IV introduces a general formulation of parametric cure models, which allows for a wide range of functional forms and time/covariate effects.

While prognostic factors have been found for many lymphoma types, individual prognoses are often based on dichotomized clinical variables and grouped risks scores, which often result in a large loss of information. This can ultimately lead to patients receiving inaccurate information. Paper V evaluates the performance of commonly used prognostic scores within 11 frequent lymphoma types by comparing to a simple model relying on age and performance status alone, and a more refined modelling approach using

an extended list of clinical variables.

Resumé

Populationsbaseret overlevelsesdata bliver ofte brugt indenfor kræftforskning til at identificere kliniske og molekylære faktorer, som er associeret til risikoen for at dø, progrediere eller lignende. Eftersom kræftpatienter også kan dø af årsager, som ikke er relaterede til kræftsygdommen, kvantificeres patientoverlevelsen nogle gange relativt til overlevelsen i en relevant baggrundsbefolkningen. Sammenlignet med baggrundsbefolkningen er den kumulerede dødsrisiko, målt fra diagnosetidspunktet, ofte højere for kræftpatienter, men forskellen mellem patientoverlevelsen og overlevelsen i baggrundsbefolkning vil typisk aftage, jo længere patienterne overlever. I artikel I sammenligner vi den betingede overlevelse for patienter diagnosticeret med diffust storcellet B-celle lymfom, og som opnåede komplet remission efter førstelinje-behandling med R-CHOP, med overlevelsen i den danske baggrundsbefolkning.

Eftersom risici kan kvantificeres på forskellige måder, kan tendenser og sammenhænge i overlevelsesdata også findes på forskellig vis. Det betingede antal mistede leveår på grund af kræftdiagnosen kaldes levetidstabsfunktionen og er et intuitivt mål for sygdommens sværhedsgrad. Det er dog svært at udregne denne funktion på grund af censurering. I artikel II evaluerer vi forskellige metoder til at udregne levetidstabsfunktionen ved hjælp af simulationer og overlevelsesdata, hvor fuld opfølgning er tilgængelig.

Ved kræftpatienter kan det være svært at vurdere, hvorvidt en patient er kureret efter behandlingen, da der i mange tilfælde er en væsentlig risiko for, at sygdommen genopstår. Dog kan overlevelsesdata bruges til at vurdere kurering. Hvis patientoverlevelsen når det samme niveau som overlevelsen i baggrundsbefolkningen, kan de patienter, som fortsat er i live blive anset for at være *statistisk kurerede*. Tidspunktet hvor dette sker bliver kaldet kureringspunktet og indeholder brugbar information for både sundshedsplanlæggere og patienter. På grund af en række udfordringer kan det være kompliceret at beregne kureringspunktet. Disse diskuteres i artikel III sammen med en ny beregningsmetode.

Kureringsmodeller er tidligere blevet brugt til at modellere andelen i en patientpopulation, som er statistisk kurerede. De mest anvendte modeller bygger på simple parametriske overlevelsesmodeller eller spline-baserede overlevelsesfunktioner med antagelse om et endeligt kureringspunkt. Artikel IV introducerer en generel formulering af parametriske kureringsmodeller, som muliggør brugen af en lang række tids- eller kovariateffekter.

Selvom prognostiske faktorer er kendte indenfor mange lymfomtyper, bliver individuelle prognoser ofte baseret på dikotomiserede kliniske variable eller grupperede risikoscorer, hvilket typisk resulterer i et stort informationstab. Ultimativt kan dette lede til, at patienterne får upræcise oplysninger.

Artikel V evaluerer nøjagtigheden af hyppigt anvendte prognostiske scorer indenfor 11 lymfomtyper ved at sammenligne med en simple model, der bygger udelukkende på diagnostisk alder og performance status, samt en mere raffineret modelleringsteknik, som udnytter en udvidet liste af kliniske variable.

Contents

Part A

Background

Background

1 Cancer

Cancer spans a wide range of malignant diseases which differ in biology, treatment, and prognosis. Certain properties, described as the hallmarks of cancer, are required of a tumour for it to be considered cancerous [1]. The hallmarks include the ability to independently grow without external growth signals, avoid natural cell death, grow limitlessly where normal cells stop doubling, and spread to other organs or tissues in the body. In Denmark, more than 40,000 individuals are diagnosed with cancer each year of which the most common cancer types are cancer of the alimentary tract, skin, respiratory system, breast, and male genitalia [2].

As tumors vary in cellular composition and severity, the treatment of cancer is very heterogeneous. Even within specific cancer types, the treatment may depend on various clinical and molecular factors. As some patients remain refractory to the standard treatment or relapse after treatment, new treatment options are constantly being tested in clinical trials with the purpose of improving the cancer survival. Based on the reported risk-benefit balance, new therapies will, if appropriate, be included in the clinical guidelines associated with specific cancer types. While clinical trials provide a framework for testing new drugs rigorously, the conclusions, however, may not be generalizable to an entire patient population due to strict inclusion criteria [3]. Because the effect of new drugs is only reported in a selected group of patients, evidence of the drug effect in the general patient population is often lacking. Observational studies are commonly used to derive such clinical evidence outside randomized clinical trials. This thesis is concerned with the analysis of register data, which provide the basis for conducting observational studies for an entire patient population of a hospital, region, or country.

1.1 Is cancer curable?

The main goal of patients diagnosed with cancer is to become cured. In practice, clinicians perform response evaluations after completion of therapy to assess the amount of remaining malignant disease. If no residual disease is observed, patients are said to be in complete remission and may be considered cured of the disease. However, a substantial risk of relapse often remains while late toxicities and lethal side effects due to the treatment may also occur. Therefore, defining cure by complete remission can be misleading and some clinicians therefore refrain from using the term cure in general [4].

Another possibility is to base the assessment of cure upon population-based risks. If the risk of relapse is sufficiently low, patients who successfully complete the treatment may be considered cured of the disease. At the same time, late toxicities due to the treatment may lead to an elevated mortality, and so mortality risks may also be of relevance for assessing cure. Since cancer patients may also die from causes not related to the cancer, it has been suggested to base cure upon the comparison of the patient mortality and the mortality risk in the gender and age-matched general population [5]. If the patient mortality after some time point reaches the level of the mortality in the general population, we may consider patients who survive beyond this time point cured of the disease. This approach takes into account all excess mortality due to the cancer diagnosis in the assessment of cure, including immediate cancer-related death, death from side effects, and death from relapses. Because cure is based on the mortality risk computed from an entire patient population, we use the term *statistical cure* about this definition [6]. In relation to statistical cure, it is of interest to estimate the time at which the patient and the general population mortality become similar. This time point has previously been termed the *cure point* [7]

Following succesful completion of the treatment, cancer patients often attend routine follow-up programs with regular visits to the clinic. While the mortality, as evaluated from the end of treatment, is often elevated in comparison to the general population mortality, the individual prognosis is likely to change at every visit. For instance, the mortality of cancer patients who remain alive after completed therapy without experiencing adverse events often approaches that of the general population. To provide prognostic information for the patients at every visit, conditional mortality risks are particularly useful [8]. At every visit, the patients might be informed about their updated prognosis, possibly in relation to the general population mortality, given that they have survived without adverse events until the time of the visit. This is also known as *dynamic prediction* and constitutes an entire subfield within survival analysis [9]. Dynamic predictions can be carried out in countless ways by using different variables for the conditional statement and end-points.

1.2 Lymphoma

Lymphoma is cancer occurring in the lymphoid tissue and constitutes a relatively small proportion of all cancer cases [2]. Within lymphoma, the patients are further grouped according to the cellular characteristics of their disease [10]. The most common lymphoma type is diffuse large B-cell lymphoma (DLBCL), which is an aggressive disease of haematopoietic system. The treatment strategy within lymphoma is heterogeneous and ranges from immediate high dose treatment with immunochemotherapy and stem-cell transplantation to wait and watch programs [11, 12]. This comes as a result

of the varying disease severity and is often based on a number of clinical factors.

In practice, prognostic indices based on clinical variables are commonly used for risk assessment and, for some lymphoma types, guiding treatment [11]. Due to the heterogeneity of lymphoma, prognostic indices have been developed separately for the most frequent lymphoma types. These are based on variables regularly measured in clinical practice, e.g., age, gender, stage, performance status, and lactate dehydrogenase (LDH) value. The most prominent example is the international prognostic index (IPI) which is commonly used for patients diagnosed with DLBCL or T-cell lymphoma to categorize their mortality risk [13]. The IPI score ranges from 0 to 5 and is calculated as the number of true statements; age above 60, stage III-IV disease, elevated LDH, performance status above 1, and more than one extranodal sites. According to the IPI score, patients are considered at low risk (0-1), low-intermediate risk (2), high-intermediate risk (3), or high risk (4-5).

1.3 Cancer register data

In Denmark, the Danish Cancer Registry (DCR) collects information on all cancer incidences in the Danish population with the main purpose of conducting cancer statistics and quality controls [14]. Since 1987, it has been mandatory to report a number of different personal and tumour characteristics of individual cancer patients to the register. By using the Danish Civil Registration System [15], personal information such as age, gender, municipality, and the date of death is merged into the DCR. The DCR does not contain any treatment nor cause of death information, but this can be retrieved from the Danish National Patient Register [16] and the Danish Register of Causes of Death [17], respectively.

Because the clinical procedures may deviate substantially between different cancer types, cancer-specific registers have been established for numerous cancer types. In this thesis, data from the Danish Colorectal Cancer Group Database [18], the Danish National Acute Leukemia Registry [19], and the Danish National Lymphoma Registry (LYFO) [20], have been used. LYFO has been nationwide since 2000 and contains information on the vast majority (94.9%) of all lymphoma cases in Denmark. In addition to the variables in the DCR, LYFO contains more detailed baseline (at diagnosis) clinical information, treatment information, as well as response and relapse information. To ensure against missing registrations, LYFO is cross referenced with the Danish National Patient Registry [16] and the National Pathology Registry [21]. The Danish Colorectal Cancer Group Database and the National Acute Leukemia Registry are designed similarly, but with clinical variables specific to colon cancer and acute leukemia. In each register, vital status and date of death in deceased patients are obtained by merging with the Danish Civil

Registration System.

2 Basic concepts in survival analysis

Survival analysis is the analysis of the duration until the occurrence of an event, e.g., the time to death. The main goal is the same as in other types of statistical analyses, namely to conduct inference about a particular response of interest, but time-to-event analysis is typically challenged by the occurrence of censoring, which implies that conventional statistical methods cannot readily be applied. In clinical studies, censoring occurs when a study is terminated before the event of interest is observed for all patients. This type of censoring is known as right censoring. For censored patients, the exact time to event is unknown, but we know that the patients had not experienced the event at the time of censoring. This information has to be taken into account in order to avoid inefficient estimators and possibly biased results. Another challenge is truncation, which occurs when patients are only included in the analysis if their event happens within a certain time window [22]. Various types of censoring and truncation can affect the analysis of time-to-event data. However, the remaining of this thesis will mainly deal with right-censored survival data and we will assume that the response of interest is the time to death.

2.1 Commonly used functions

In medical studies, survival data are often used to measure the effect of an exposure by comparing to a control group. In particular, in randomized trials of cancer patients, the benefit of new drugs is often assessed based on the observed effect on the mortality. However, the effect of a certain drug can be quantified by various measures. In the following, we describe the most commonly used measures in survival analysis. We use the term *follow-up* as the period in which we monitor patients in a particular study. At the end of the follow-up, alive patients are censored.

First, let T be a random variable with density function f, denoting the survival time of a cancer patient. In many medical studies, the probability of surviving beyond some time t is reported. Treating this probability as a function of time, we obtain the *survival function*,

$$S(t) = \mathrm{P}(T > t) = \int_t^\infty f(u)du.$$

Naturally, we have that $S(0) = 1$ and $\lim_{t\to\infty} S(t) = 0$ if T is proper. In studies of cancer survivorship, the 5-year survival probability, i.e., the probability of surviving five years after the diagnosis, $S(5)$, is often reported. Although

the survival function can be estimated in various ways, the Kaplan-Meier (KM) estimator is typically used [23]. The KM estimator is only defined until the last available follow-up time and due to censoring, it is often seen that the KM estimator does not reach zero within the follow-up.

Another used measure is the *hazard function*,

$$h(t) = \lim_{h \to 0} \frac{\mathrm{P}(t \leq T < t + h | T \geq t)}{h}.$$

It provides the instantaneous risk of death given survival until time t and thus is a dynamic measure of the mortality risk. The definition of the hazard function implies the following relation between the hazard, density, and survival function:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log[S(t)]}{dt}.$$

While the hazard function, itself, is rarely reported in clinical studies, the hazard ratio, i.e., the ratio between two hazard functions, has gained widespread popularity in medical research. This is largely due to the Cox proportional hazards (CPH) model, which enables regression modelling of the hazard function by using the assumption that covariates affect the hazard function proportionally [24]. The model is specified by,

$$h(t|\boldsymbol{z}) = h_0(t) \exp\left(\boldsymbol{z}^T \boldsymbol{\beta}\right),$$

where $\boldsymbol{z}$ is a vector of covariates, $\boldsymbol{\beta}$ is a vector of coefficients, and $h_0(t)$ is a baseline hazard function, i.e., the hazard if $\boldsymbol{z} = \mathbf{0}$. Thus, if $\boldsymbol{z}$ denotes whether a patient has received treatment ($z = 1$) or placebo ($z = 0$), the ratio between the hazard function of the treated and the control group is a constant equal to $\exp(\beta)$. The estimated value of β and its standard error can then be used to assess the benefit of the treatment. The main advantage of the CPH model is that it can be fitted without making any assumptions about the form of h_0 by using a partial likelihood approach [24]. Therefore, this model is sometimes referred to as a semi-parametric survival model.

The *cumulative hazard function* given as

$$H(t) = \int_0^t h(u) du$$

is sometimes used to provide a measure of the cumulative mortality risk. The cumulative hazard function is interpreted as the expected number of events before time t. Given that the event of interest is death, this is under the assumption that patients can be resurrected immediately after each death, and thus the interpretation is not ideal for reporting mortality risks. However, the cumulative hazard function is related to the survival function by $H(t) =$

$-\log S(t)$.

2.2 Parametric survival models

This thesis is largely concerned with parametric survival models, which are survival models specified entirely by a finite number of parameters. A very simple and commonly used parametric survival model is the Weibull model. The survival function of the Weibull model is given as

$$S(t) = \exp\left(-\beta_1 t^{\beta_2}\right), \tag{1}$$

where $\beta_1 > 0$ and $\beta_2 > 0$ are parameters controlling the trajectory of the survival function.

To see how parametric models are estimated, we introduce some notation. We assume that n patients are followed and the time to event for the i^{th} patient is denoted X_i. Furthermore, patient i is censored at time C_i, implying that we observe only $T_i = \min(X_i, C_i)$ and a status indicator $\delta_i = \mathbb{1}[X_i \leq C_i]$ denoting whether patient i is censored before the event or not. Thus, we observe n independent triplets $(T_i, \delta_i, \boldsymbol{z}_i)$, where $\boldsymbol{z}_i$ is the covariates of the i^{th} patient. Often it is also assumed that X_i and C_i are independent.

The standard approach for estimating parametric survival models is by maximum likelihood estimation. In standard likelihood inference, the i^{th} contribution to the likelihood is the density function evaluated in the data of the i^{th} patient, but for censored patients, we only know that the event has not occured before the time of censoring. For these patients the survival function is used instead. Given that the survival model is described by parameters $\boldsymbol{\theta}$, the general likelihood for right-censored survival data [22] is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})^{\delta_i} S(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})^{1-\delta_i} = \prod_{i=1}^{n} h(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})^{\delta_i} S(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})$$

yielding the log-likelihood,

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log(h(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})) + \log(S(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})). \tag{2}$$

The model of interest is fitted by optimizing (2) with respect to $\boldsymbol{\theta}$. In the case of the Weibull model, (1) and $h(t) = \beta_2 \beta_1 t^{\beta_2 - 1}$ are inserted into (2) and the function is optimized with respect to $\boldsymbol{\theta} = (\beta_1, \beta_2)$.

2.3 Flexible parametric survival models

While the Weibull model has been used for many applications, the model is rather restrictive, which means that for most applications, the model will not be able to capture the true underlying survival function particularly well. To avoid the pitfalls of simple parametric survival models, a vast literature on flexible parametric survival models exists. The most popular alternative is the model by Royston and Parmar, which uses restricted cubic splines (RCSs) to model the survival function [25]. The Royston-Parmar model is specified by

$$S(t|\boldsymbol{z}) = \exp\left(-\exp\left(s_0(x,\boldsymbol{\gamma}) + \boldsymbol{z}^T\boldsymbol{\beta}\right)\right), \tag{3}$$

where $x = \log(t)$ and $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ are model parameters. The RCS is defined as a linear combination, $\sum_{i=1}^{K} v_i(x)\gamma_i$, of base functions, v_i (see [25] for precise formulation of the base functions), after selection of K knots on the log-time scale. Thus, the RCS is modelled on the log-time scale and the covariates are additive on the log-log scale. In fact, the Royston-Parmar model forms a proportional hazards model, which can easily be seen by

$$\begin{aligned} h(t|\boldsymbol{z}) &= -\frac{d}{dt}\log[S(t|\boldsymbol{z})] = \frac{d}{dt}\exp\left(s_0(x,\boldsymbol{\gamma}) + \boldsymbol{z}^T\boldsymbol{\beta}\right) \\ &= \exp\left(s_0(x,\boldsymbol{\gamma}) + \boldsymbol{z}^T\boldsymbol{\beta}\right)\frac{ds_0(x,\boldsymbol{\gamma})}{dt} \\ &= h_0(t)\exp\left(\boldsymbol{z}^T\boldsymbol{\beta}\right), \end{aligned}$$

where

$$h_0(t) = \exp\left(s_0(x,\boldsymbol{\gamma})\right)\frac{ds_0(x,\boldsymbol{\gamma})}{dt}.$$

This means that the covariate-specific parameters, $\boldsymbol{\beta}$, can be interpreted in the same manner as in the CPH model. However, the baseline hazard is modelled parametrically and thus smooth hazard and survival predictions are obtained from this model. The model can be fitted by maximum likelihood using (2). The flexibility of the Royston-Parmar model is controlled by the number of knots in the RCS. The Akaike information criterion (AIC) or similar measures were proposed to find the optimal number of knots [25].

The Royston-Parmar model was recently generalized by Liu et al. [26]. Using a bijective, monotone link function, $g : (0,1) \to (-\infty, \infty)$, and a time-varying linear predictor, they modelled the survival function by

$$g(S(t,\boldsymbol{z},\boldsymbol{\theta}) = \eta(t,\boldsymbol{z},\boldsymbol{\theta}) = X(t,\boldsymbol{z})\boldsymbol{\theta}.$$

That is, the g-transformed survival function is assumed to be linear in $\boldsymbol{\theta}$. The Royston-Parmar model is obtained if $g(x) = \log(-\log(x))$ and if the

time-effect is given by an RCS. Again, optimization of (2) is used to fit the model. The intent of this generalization was to enable the use of a wide range of smoothers, such as those described by Wood [27] and implemented in the `mgcv` `R`-package. In addition, the time effect is not restricted to be on the log-time scale, but may be formulated for t or $\sqrt{t}$ without additional considerations.

The approach by Liu et al. is implemented in the `R`-package `rstpm2`, which enables modelling of right-censored, left-truncated, and interval-censored time-to-event data. Also, Liu et al. proposed a penalized regression approach, such that the knot selection in knot-based smoothers becomes only a minor obstacle. This also avoids the use of AIC or similar measures for determining the flexibility of the survival model [26].

3 Mortality relative to the general population

In population-based studies of cancer survival, it is sometimes of interest to quantify the patient mortality relative to the mortality seen in the general population of a country or region. Life tables contain information on the periodical death rate in the general population stratified on a number of demographic variables such as gender and age. Under the assumption that the patient population of interest is a negligibly small part of the general population, life tables can thus be used to quantify the mortality among matched individuals without cancer. While life tables are available at central statistical offices, these are also readily available at the Human Mortality Database [28] for a wide range of countries.

Alternatively, the general population survival may be obtained by matching each patient to a number of "healthy" individuals with similar demographics. This enables matching on a larger number of demographical variables such as comorbidities and municipality, but requires additional data, which are not readily available. In Denmark, this approach is enabled through the central statistical office, Statistics Denmark.

In the following, different approaches for comparing the patient and general population mortality are described, with the general population mortality determined by life tables. Similarly to Pohar et al. [29], we describe both multiplicative and additive models incorporating the general population mortality.

3.1 Multiplicative models

An often used approach for comparing the patient and general population mortality is a proportional hazards model. For a general population hazard,

h^*, the model is given by

$$h(t) = h^*(t)\exp\left(\boldsymbol{z}^T\boldsymbol{\beta}\right), \tag{4}$$

where $\boldsymbol{z}$ is a vector of covariates and $\boldsymbol{\beta}$ are the corresponding coefficients. If $\boldsymbol{z} = \mathbf{1}$, then $\exp(\boldsymbol{\beta})$ denotes the ratio of the patient hazard to the general population hazard. Breslow et al. [30] showed that estimating $\exp(\boldsymbol{\beta})$ under $\boldsymbol{z} = \mathbf{1}$ corresponds simply to computing the *standardized mortality ratio* (SMR),

$$\text{SMR} = \frac{O}{E},$$

where O is the observed number of deaths during the follow-up and E is the expected number of deaths as determined by the general population. The observed number of deaths is simply $O = \sum_{i=1}^{n} \delta_i$ and utilizing the interpretation of the cumulative hazard function, as described in Section 2.1, the expected number of deaths may be computed by $E = \sum_{i=1}^{n} H^*(t_i|\boldsymbol{z}_i)$, where H^* is the general population cumulative hazard. The SMR provides a measure of the relative mortality throughout the entire follow-up. Whenever dynamic evaluations of the overall relative mortality is of interest, O and E may be computed conditionally on, say, survival until time t, which yields a conditional SMR. This quantifies the mortality ratio from time t throughout the remaining follow-up.

A generalization of the model in (4),

$$h(t) = h^*(t)v_0(t)\exp\left(\boldsymbol{z}^T\boldsymbol{\beta}\right),$$

was introduced by Andersen et al. [31]. Here the baseline ratio of the patient hazard to the general population hazard is given by $v_0(t)$ and deviations from this function is modelled by the linear predictor $\boldsymbol{z}^T\boldsymbol{\beta}$. This model can be fitted in the framework of Cox proportional hazards models [24] by allowing for a time-varying offset. Note that the hazard function provides the conditional instantaneous risk of death and so v_0 provides a ratio of two instantaneous risks.

3.2 Additive models

While the multiplicative models resemble the CPH model, another type of model incorporating the general population mortality may be established similarly to the additive Aalen model [32]. The additive hazards model is given as,

$$h(t|\boldsymbol{z}) = h^*(t|\boldsymbol{z}) + \lambda(t|\boldsymbol{z}). \tag{5}$$

Here, $\lambda(t|\boldsymbol{z})$ denotes the additional mortality a patient experiences as compared to the general population also known as the *excess hazard* or *excess mortality*. Although various models can be formulated for $\lambda(t|\boldsymbol{z})$, proportional excess hazard models,

$$\lambda(t|\boldsymbol{z}) = \lambda_0(t)\exp\left(\boldsymbol{z}^T\boldsymbol{\beta}\right), \tag{6}$$

are popular. However, unlike the CPH model, a fitting procedure which does not require specification of the baseline excess hazard, λ_0, is not available and thus λ_0 is commonly specified parametrically. A simple way to so, is to divide the follow-up into K intervals and assume a piece-wise constant baseline, $\lambda_0(t) = \sum_{j=1}^{K}\gamma_j \mathbb{1}[t \in I_j]$, where I_j is the j^{th} interval and γ_j is the piecewise level of the baseline excess hazard [33]. Alternative parametric models for λ_0 may also be employed.

Using the relation between the hazard and the survival function, we obtain

$$\begin{aligned} S(t) &= \exp\left(-\int_0^t h(u)du\right) = \exp\left(-\int_0^t h^*(u) + \lambda(u)du\right) \\ &= \exp\left(-\int_0^t h^*(u)du\right)\exp\left(-\int_0^t \lambda(u)du\right). \end{aligned}$$

Let S^* be the survival function of the general population and let $R(t) = \exp\left(-\int_0^t \lambda(u)du\right)$. Then we obtain

$$R(t) = \frac{S(t)}{S^*(t)}. \tag{7}$$

The function $R(t)$ is termed the *relative survival* since it is the ratio between the patient and general population survival function. Thus, while (6) defines a model for the excess hazard, it also defines a model for the relative survival. Note, the relative survival is not bounded to be below 1 or decreasing and so is not a proper survival function. Examples of the patient, general population, and relative survival are shown in Figure 1.

Assuming a parametric model for $\lambda(t|\boldsymbol{z})$, (5) may be fitted by maximum likelihood estimation. By inserting (5) and (7) into (2) we obtain

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\delta_i \log\left[h^*(T_i|\boldsymbol{z}_i) + \lambda(T_i|\boldsymbol{z}_i,\boldsymbol{\theta})\right] + \log(R(T_i|\boldsymbol{z}_i,\boldsymbol{\theta})) + \log(S^*(T_i|\boldsymbol{z}_i)).$$

Since S^* does not depend on any model parameters, the likelihood reduces

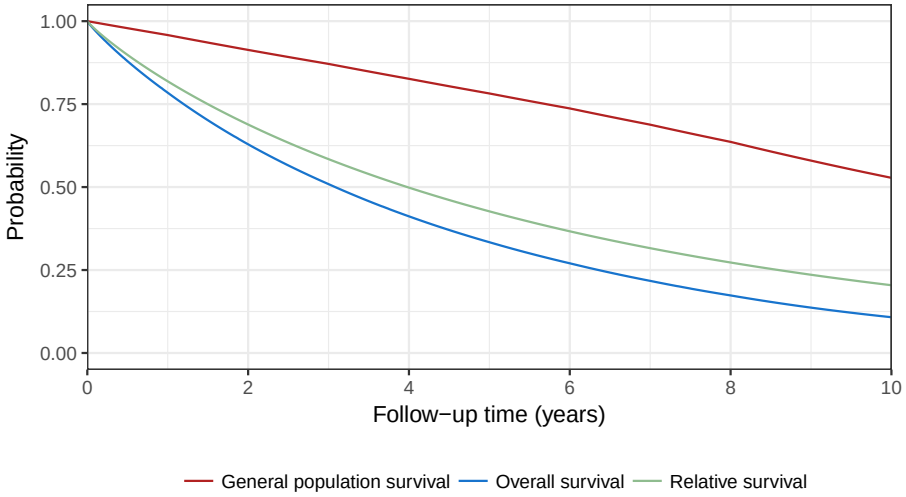

Figure 1: An example of the patient, general population, and relative survival. The general population survival was computed for a Danish man aged 75 years in 2010.

to

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log\left[h^*(T_i|\boldsymbol{z}_i) + \lambda(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})\right] + \log(R(T_i|\boldsymbol{z}_i, \boldsymbol{\theta})). \tag{8}$$

The above likelihood highlights the advantage of using parametric additive excess hazards (or relative survival) models. The only additional information needed to the likelihood is the general population hazard evaluated at each observed event time, which can easily be obtained. Clearly if $h^* = 0$, the usual likelihood function for parametric survival models is obtained. As parametric relative survival models are obtained by simply adding the general population hazard to the usual likelihood function, the Royston-Parmar model can easily be extended to relative survival when dealing with right-censored survival data, which was demonstrated by Nelson et al. [34]. As shown in the Appendix, the same information from the general population is needed when dealing with left-truncated survival data. However, for interval-censored data, S^* cannot be ignored in the likelihood, which makes it slightly more tedious to fit relative survival models based on this data type.

Non-parametric estimators of the relative survival function, including the Ederer I [35], Ederer II [36], and Hakulinen [37] estimator, have also previously been introduced. These three were all established by considering an estimator of the patient and general population cumulative hazard functions, subtracting these, and then transforming to the survival scale to obtain (7). The estimators differ in their formulation of the general population cumu-

lative hazard, but use the same estimator for the patient cumulative hazard, namely the Nelson-Aalen estimator [32].

The relative survival function is often associated with *net survival*, which is a term used about the patient survival in the hypothetical scenario where deaths due to other causes than the considered disease are eliminated. Similarly to Perme et al. [38], we may define two random follow-up times, T_D and T_P, which are the time to disease-related death and death due to other causes, respectively. However, we observe only the smallest follow-up time, i.e., $T = \min(T_D, T_P)$. The net survival for an individual with covariates $\boldsymbol{z}$ is

$$S_D(t|\boldsymbol{z}) = \mathrm{P}(T_D > t|\boldsymbol{z}) = \frac{\mathrm{P}(T_D > t, T_P > t|\boldsymbol{z})}{\mathrm{P}(T_P > t|\boldsymbol{z})} = R(t|\boldsymbol{z}).$$

The above equality builds on two main assumptions: T_D and T_P are conditionally independent given the covariate vector $\boldsymbol{z}$ and the survival associated with T_P is equal to the survival in the general population, S^*. The marginal net survival may be computed by

$$S_D(t) = \frac{1}{n}\sum_{i=1}^{n} S_D(t|\boldsymbol{z}_i).$$

Perme et al. showed that the the three non-parametric relative survival estimators, Ederer I, Ederer II, and the Hakulinen estimator, do not estimate the marginal net survival [38]. In addition, Perme et al. derived an estimator for the marginal net survival, which does not explicitly give a formulation of S^*, but rather estimates the relative survival directly.

The estimator of Perme et al. requires the two aforementioned assumptions in order to give the relative survival a net interpretation. However, these assumptions are generally challenging to verify, which was also argued by Andersen and Keiding in their discouragement of measures for which the interpretation is not in the real world [39].

The excess hazard provides a measure of the instantaneous excess mortality given survival until time t, while the relative survival defines the ratio of the cumulative patient survival to the cumulative general population survival. When conditional evaluations are of interest, the conditional relative survival is commonly used. Given survival until time u, the conditional relative survival is given as

$$R_u(t|\boldsymbol{z}) = \frac{R(t+u|\boldsymbol{z})}{R(u|\boldsymbol{z})} = \frac{S(t+u|\boldsymbol{z})}{S^*(t+u|\boldsymbol{z})}\frac{S^*(u|\boldsymbol{z})}{S(u|\boldsymbol{z})}.$$

Thus, the conditional relative survival is the ratio of the conditional patient survival function to the conditional general population survival function.

3.3 Loss of lifetime

Often in medical studies, the main interest is the mean value of some response, but due to censoring the mean value of the time to some event can typically not be estimated directly. However, techniques can be used to provide estimates of the mean value. We first recall that the mean value of the time to death, T, is related to the survival function by

$$\mathrm{E}[T] = \int_0^\infty S(u)du. \tag{9}$$

This entity is also known as the *mean lifetime* or *mean survival time*. Thus, the mean lifetime can be computed by the area under the the survival function (Figure 2). The conditional mean lifetime given survival until time t is

$$\mathrm{E}[T - t | T > t] = \int_t^\infty \frac{S(u)}{S(t)}du.$$

This function is known as the *mean residual lifetime*. The difference between the patient and general population mean residual lifetime,

$$L(t) = \int_t^\infty \frac{S^*(u)}{S^*(t)}du - \int_t^\infty \frac{S(u)}{S(t)}du, \tag{10}$$

provides a dynamic measure of the severity of the disease. We term this the *loss of lifetime* function (Figure 2), and if $t = 0$, we obtain the number of years lost due to the disease.

When dealing with censored data, these three measure cannot be computed without some form of extrapolation, which makes the use of non-parametric estimators, such as the KM estimator, insufficient. Therefore, extensive research has been devoted to developing methods for providing accurate extrapolations of the survival function. A common approach is to use long-term external data, such as register data or life tables, to provide more stable extrapolations [40]. However, generally the accuracy of extrapolated survival probabilities cannot be assessed and even well-fitting models may extrapolate poorly [41].

To avoid extrapolation, a common practice is to replace ∞ with some finite time point, τ. In terms of the mean lifetime, this corresponds to computing $\mathrm{E}[\min(T, \tau)]$, which has a different interpretation than (9). This measure is the mean lifetime over the following τ years and is commonly known as the *restricted mean lifetime*. The Royston-Parmar model [25] and pseudo values [42] are recommended approaches for estimating the restricted mean lifetime [43]. By replacing ∞ with τ in the loss of lifetime, (10), we obtain the conditional number of years lost due to the disease over the following $\tau - t$ years given survival until time t.

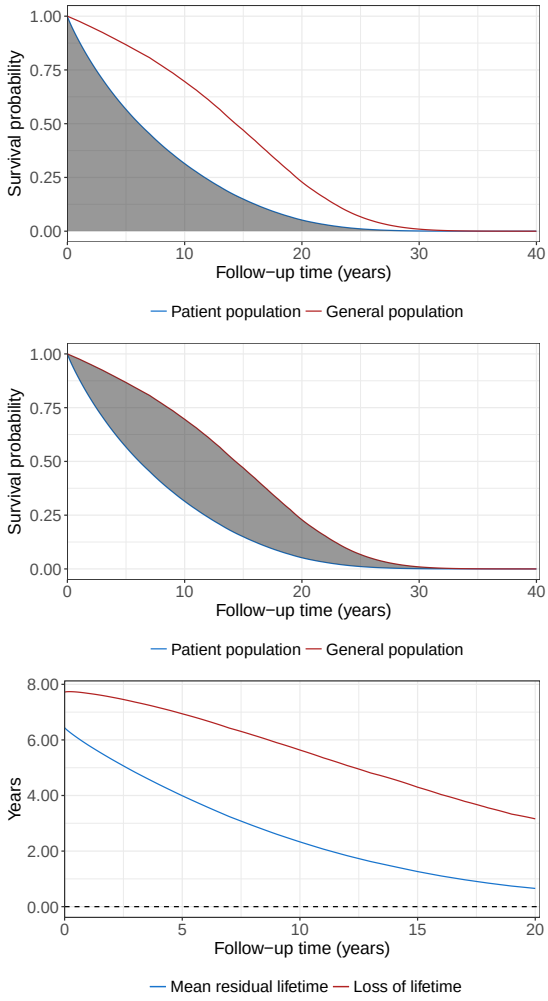

Figure 2: In the top figure, the mean survival of the patients is indicated by the shaded area. In the middle, the shaded area indicates the loss of lifetime at time zero. The bottom figure shows the mean residual lifetime and the loss of lifetime function from time 0 until 20 years.

4 Cure models

Occasionally when dealing with time-to-event data, it is reasonable to assume that some individuals will never experience the event of interest. We consider these individuals cured from the event. For example, some mothers will never have more than one child. So the time between the birth of the first and second child displays a cure pattern [44]. In this context, it may be of interest to estimate the proportion of mothers who will never have a second child.

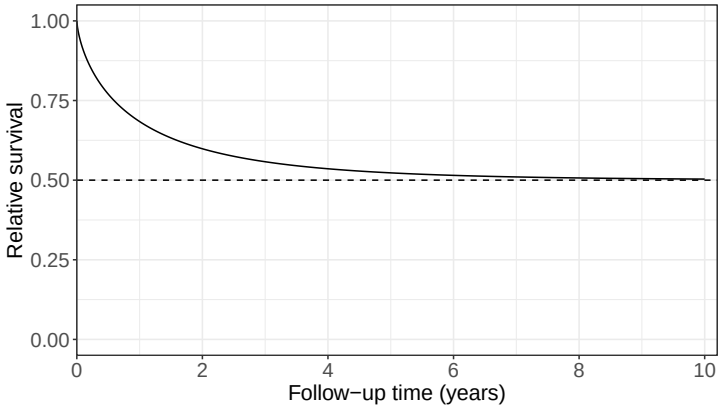

Figure 3: Cure model for the relative survival with $R(t) \to \pi = 0.5$ as $t \to \infty$.

While this is hampered by the occurrence of censoring, a reasonable estimate may still be obtained if the follow-up is sufficiently long. In particular, the cure proportion, π, can be modelled and estimated using cure models, which have a long history in the statistical literature [45], and may be formulated both parametrically and non-parametrically [46, 47].

However, in analyses of time to death, it is not appropriate to assume cure from death. Instead, cure models for the relative survival function may be used if statistical cure can reasonably be assumed for a proportion of the individuals. That is, some individuals experience the same survival as the general population, which is indicated by a plateau in the relative survival [48, 49]. In the following, we will only consider cure models for the relative survival function.

As the patient population may be modelled as a mixture of patients that are statistically cured and patients who are uncured, we may model the relative survival as a mixture model. The relative survival of cured patients is a constant equal to one, and so the total relative survival may be written as

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})]\, S_u(t|\boldsymbol{z}),$$

where $S_u(t|\boldsymbol{z})$ denotes the relative survival of the uncured patients. This model is known as a *mixture cure model* [49]. Often $S_u(t|\boldsymbol{z})$ is assumed to be proper since the ratio of the patient to the general population survival function in uncured patients is expected to approach zero as time approach infinity. Therefore, under this assumption, we also have that $R(t) \to \pi$ as $t \to \infty$ (Figure 3).

Another type of cure models, namely *non-mixture cure models* or promotion time cure models, also enables analysis of the cure proportion [49]. The

model is given as

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})},$$

where $\widetilde{F}(t|\boldsymbol{z})$ is a proper distribution function. The function $\widetilde{F}(t|\boldsymbol{z})$ is not interpreted as the distribution function of those who are uncured, but is merely used to model the time-effect of the relative survival. However, the relative survival of the uncured can still be obtained by a simple transformation,

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})] \frac{\pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})} - \pi(\boldsymbol{z})}{1 - \pi(\boldsymbol{z})},$$

such that,

$$S_u(t|\boldsymbol{z}) = \frac{\pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})} - \pi(\boldsymbol{z})}{1 - \pi(\boldsymbol{z})}.$$

Thus, the two primary entities in cure models, namely the cure proportion and the survival of those who are uncured, can be estimated from both mixture and non-mixture cure models. Although several non-parametric cure models have been introduced before, these are not immediately applicable to relative survival [46, 50]. In contrast, given a parametric formulation of π and S_u (or $\widetilde{F}$), these models can be fitted by optimizing the log-likelihood in (8). A Stata implementation of mixture and non-mixture cure models is available, which enables the use of several distribution functions for S_u (weibull, lognormal, and generalized gamma) and link functions for π (identity, logit, and log-log) [51].

Due to gradual improvements in cancer care, the relative survival of several cancer types has been increasing during the last two decades [52]. In the framework of cure models, relative survival improvements may occur in four different ways. The increase in cancer survival may be seen by an increase in the cure proportion, which may occur if, for instance, a new therapy is introduced. We may also see the mortality improvement by an increase in the survival of the uncured, S_u. A combination of both scenarios may also occur. Finally, an increase in the cure proportion may be observed together with a decrease in the survival of the uncured. This might occur if the introduction of a new drug implies that the uncured group consists of fewer, but more frail patients.

4.1 Flexible parametric cure models

While implementations of mixture and non-mixture cure models utilizing simple distribution functions are readily available, these models suffer from the disadvantage of being too simple to capture the shape of the true survival of the uncured, S_u. Therefore, attempts have been made to provide more flexible parametric cure models. Lambert et al. [53] introduced a mixture

model for S_u to increase the model flexibility. The model for S_u was described as a mixture of a Weibull and an exponential survival function or a mixture of two Weibull models. However, this model is challenged by the variable selection required for each term.

Another type of cure model was introduced by Andersson et al. who altered the RCS of the Royston-Parmar model in order to obtain a constant relative survival after the last knot [7]. The cure model was formulated similarly to (3), but was shown to be a special case of a non-mixture cure model, where $\widetilde{F}(t)$ is 1 beyond the last knot. As opposed to the more simple mixture and non-mixture cure models, cure is assumed at the last knot of the spline. Therefore, the placement of the last knot becomes crucial for the cure proportion estimates [7]. A similar approach was suggested by Bremhorst and Lambert [54], who used cubic P-splines to model the time-effect of a non-mixture cure model in a Bayesian framework. By letting the last parameter of the spline be sufficiently large, $\widetilde{F}(t)$ was forced to be one by the end of the follow-up. The latter approach was suggested for total survival cure models, but is generalizable to relative survival.

5 Prognostic survival models

In medical research, some studies aim to investigate the (causal) effect of some variable on a response of interest. This could be the effect of a new cancer treatment on the patient survival. Other studies aim to establish models that can predict a certain event as accurately as possible in new patients based on a number of covariates. These two types of studies are known as etiological and predictive studies, respectively, with the vast majority of medical studies focusing on etiological conclusions [55]. Since the aims of etiological and predictive studies differ, the preferred statistical approaches also deviate [56]. The bias-variance trade-off is often used to depict the difference between these two types of studies. The expected prediction error (EPE) for a model is the expectation of the prediction error with respect to both the response and covariates, and may be factorized by

$$\text{EPE} = \text{Irreducible error} + \text{Bias}^2 + \text{Variance}.$$

Etiological studies are concerned with minimizing the bias term in order to accurately explain the relationship between the response and the covariates. In predictive studies, the aim is to minimize the entire EPE in order to achieve a model that leads to accurate predictions in future patients. Since the EPE is minimized by reducing the sum of the bias and the variance, choosing a more biased model, which at the same time reduces the variance, may lead to a better model. Therefore, some models may be suitable for inferring

etiological conclusions while others might be more appropriate for predicting new events.

Also, etiological studies tend to focus on estimating relative measures such as hazard ratios, odds ratios, or relative risks, while predictive studies focus on predicting the absolute risk of some event [55]. Validating a predictive model in independent data is an important part of predictive studies. The validation can be conducted in various ways dependent on the desired generalizability of the model and is commonly based upon some type of performance measure.

5.1 Performance measures

Let $\widehat{G}$ be some estimated predictive model for a non-censored continuous response T with covariates $\boldsymbol{z}$. Let $(T_1, \boldsymbol{z}_1), ..., (T_n, \boldsymbol{z}_n)$ be n new observations which were not used to fit the model $\widehat{G}$, and let $\widehat{T}_i = \widehat{G}(\boldsymbol{z}_i)$ be the predicted responses from the model. The accuracy of the predicted responses can then be assessed through various loss functions, e.g., the mean square error, which is computed by $(1/n)\sum_{i=1}^{n}(T_i - \widehat{T}_i)^2$.

When dealing with time-to-event data, the predicted time to event, $\widehat{T}$, may be obtained as the mean or median survival time based on the predicted survival function, $\widehat{S}(t|\boldsymbol{z})$ [57]. However, in prognostic research, it is more often seen that the accuracy of the model is based directly on the predicted survival function, $\widehat{S}(t|\boldsymbol{z})$. Another common approach is to base the model accuracy on some risk predictor, $\mathcal{G}(\boldsymbol{z})$, where larger values indicate a shorter time to event.

The C-index is widely used in the medical literature and provides a measure of the concordance between the order of the predicted risks and the event times for two randomly selected patients,

$$C = \mathrm{P}(\mathcal{G}(\boldsymbol{z}_1) > \mathcal{G}(\boldsymbol{z}_2)|T_2 > T_1).$$

The above C-index was first introduced by Harrell et al. [58], but alternative definitions and several estimators have since been introduced [59].

The time-varying AUC (tAUC) is a generalization of the receiver operating characteristic curve and may be defined in multiple ways [60], which yield slightly different interpretations. For two randomly selected patients, the cumulative dynamic tAUC is given by

$$\mathrm{AUC}(t) = \mathrm{P}(\mathcal{G}(\boldsymbol{z}_1) > \mathcal{G}(\boldsymbol{z}_2)|T_1 \leq t, T_2 > t).$$

Thus, the tAUC matches how well the risk ranking corresponds to the actual ranking of the survival times. In the case of no censoring, the tAUC has a natural estimator [61], but whenever censoring is present, the natural estimator

needs to be modified. In particular, weighting with the inverse probability of censoringis a commonly used strategy [61]. Using this approach, the tAUC cannot be estimated for time points smaller than the first follow-up time.

Another performance measure, which is similar to the mean square error, is the Brier score [62],

$$\mathrm{BS}(t, \widehat{S}) = \mathrm{E}\left[Y(t) - \hat{S}(t|\boldsymbol{z})\right]^2,$$

where $Y(t) = \mathbf{1}[T \geq t]$ [63] and the expectation is with respect to both T and $\boldsymbol{z}$. The integrated Brier score is given as

$$\mathrm{IBS}(\widehat{S}) = \int_0^\tau \mathrm{BS}(u, \widehat{S}) ds,$$

for some upper limit τ. If the data were uncensored, the Brier score could easily be estimated by computing the empirical mean. Similarly to the tAUC, inverse probability weighting has been proposed to obtain consistent estimators of the Brier score [64]. The Brier score measures how well the predicted survival probabilities correspond to the observed survival statuses. Thus it provides a measure of the model *calibration*, while the tAUC provides a measure of the *discrimination*. Therefore, the performance ranking of two models is not guaranteed to be the same when based on these measures. In fact, one model might provide an accurate risk ranking, but at the same time provide inaccurate survival probabilities.

6 Overview of the thesis

In the following, the papers of this thesis are presented. For published articles, the articles as available through the publishers website are included, and the remaining papers are formatted for this thesis. All papers are accompanied by a publication status and a small description of the content.

Paper I deals with the survival of DLBCL patients in first complete remission after R-CHOP(-like) treatment. In particular, the survival was compared to that of the general population by using data from LYFO. The main focus was to confirm a previously published result showing that DLBCL patients who remain event-free for two years after diagnosis experience the same mortality as the general population [65]. In conclusion, the survival was not normalized despite many years in complete remission. However, the difference in restricted mean survival between the patients and the general population was rather small after two years of event-free survival.

Paper II deals with methods for estimating the loss of lifetime function which is given as the difference between the patient and general population mean residual lifetime. Due to censoring, the loss of lifetime function cannot

be directly estimated from the KM estimator, but requires extrapolation of the patient and general population survival functions. We investigated three extrapolation approaches based on relative survival. The approaches were tested in a simulation study and by using cancer survival data from the DCR where complete follow-up was available. For illustrative purposes, the loss of lifetime function was estimated in data from LYFO.

Paper III describes the concept of cure points, which is the time point at which the patient mortality reaches the same level as general population mortality. The estimation of cure points is not appropriately done by conventional hypothesis testing. We introduced an approach for cure point estimation using clinical relevant measures of the mortality difference between the patients and the general population. The accuracy of the cure point estimator was evaluated in a simulation study. Real-world examples were carried out by using data from the Danish Colorectal Cancer Group Database, the Danish Acute Leukemia Registry, and LYFO.

Paper IV deals with cure models and attempts to provide a general and flexible estimation framework for these. The approach expands on the generalized survival models by Liu et al. [26]. We considered two classes of parametric cure models, namely explicit and latent cure models, which differ in their inclusion of parameters for the cure proportion. Through simulations and analysis of real-world data from the Danish Colorectal Cancer Group Database and LYFO, we highlighted the differences between these two classes in terms of identifiability and performance.

Paper V evaluates the performance of commonly used prognostic scores for 11 frequent lymphoma types by utilizing survival data from LYFO. Each prognostic score was tested against a simple model based on age and performance status alone. In addition, a more advanced modelling approach was also tested. The paper illustrated a subpar performance of most prognostic scores, which could easily be improved upon using simple CPH models. Additionally, advantages of using more advanced modelling were only observed for few lymphoma types.

The developed methods in paper II, III, and IV were implemented in the `R`-package `cuRe` (see `https://github.com/LasseHjort/cuRe`). The package contains functions for fitting flexible parametric cure models and associated summary measures. In particular, the loss of lifetime, crude relative survival-based cumulative incidences, and cure points can be computed from `cuRe`.

Appendix

In the following, we consider the likelihood function in the case of not only right-censored survival data, but also left-truncated and interval-censored data. Similarly to Liu et al. [26], we assume that (X, Y_L, T_1, T_2) is observed, where X is the observed, and possibly right-censored, follow-up time, Y_L is the delayed entry time, and T_1 and T_2 are the boundaries for interval-censored observations. For observations without delayed entry $Y_L = 0$. Let D denote the set of patients for whom the exact event time is known, let R be the set of right-censored patients, and let I denote interval-censored samples. For $n = |D| + |R| + |I|$ independent observations and a parametric survival model with parameters $\boldsymbol{\theta}$, the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i \in D} \frac{h(X_i|\boldsymbol{z}_i, \boldsymbol{\theta}) S(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})}{S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta})} \prod_{i \in R} \frac{S(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})}{S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta})} \prod_{i \in I} \frac{S(T_{1i}|\boldsymbol{z}_i, \boldsymbol{\theta}) - S(T_{2i}|\boldsymbol{z}_i, \boldsymbol{\theta})}{S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta})},$$

which yields the log-likelihood,

$$\begin{aligned}
\ell(\boldsymbol{\theta}) = &\sum_{i \in D} \left[\log(h(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) + \log(S(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right] \\
&+ \sum_{i \in R} \left[\log(S(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right] \\
&+ \sum_{i \in I} \left[\log(S(T_{1i}|\boldsymbol{z}_i, \boldsymbol{\theta}) - S(T_{2i}|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(S(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right].
\end{aligned}$$

Assume that only patients in D and R are observed, i.e., no patients are interval-censored. For a relative survival model, we rewrite the likelihood with $h(t|\boldsymbol{z}_i, \boldsymbol{\theta}) = h^*(t|\boldsymbol{z}_i) + \lambda(t|\boldsymbol{z}_i, \boldsymbol{\theta})$ and $S(t|\boldsymbol{z}_i, \boldsymbol{\theta}) = S^*(t|\boldsymbol{z}_i) R(t|\boldsymbol{z}_i, \boldsymbol{\theta})$, and obtain

$$\begin{aligned}
\ell_{\text{LT}}(\boldsymbol{\theta}) = &\sum_{i \in D} \left[\log\left(\lambda(X_i|\boldsymbol{z}_i, \boldsymbol{\theta}) + h^*(X_i|\boldsymbol{z}_i)\right) + \log(R(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) + \log(S^*(X_i|\boldsymbol{z}_i))\right. \\
&\left. - \log(R(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(S^*(Y_{Li}|\boldsymbol{z}_i))\right] \\
&+ \sum_{i \in R} \left[\log(R(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) + \log(S^*(X_i|\boldsymbol{z}_i)) - \log(R(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right. \\
&\left. - \log(S^*(Y_{Li}|\boldsymbol{z}_i))\right] \\
\equiv &\sum_{i \in D} \left[\log\left(\lambda(X_i|\boldsymbol{z}_i, \boldsymbol{\theta}) + h^*(X_i|\boldsymbol{z}_i)\right) + \log(R(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(R(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right] \\
&+ \sum_{i \in R} \left[\log(R(X_i|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(R(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta}))\right].
\end{aligned}$$

If some patients are interval-censored, the corresponding likelihood is

$$\ell_{\text{GEN}}(\boldsymbol{\theta}) = \ell_{\text{LT}}(\boldsymbol{\theta}) + \sum_{i \in I} \log \left[R(T_{1i}|\boldsymbol{z}_i, \boldsymbol{\theta}) S^*(T_{1i}|\boldsymbol{z}_i) - R(T_{2i}|\boldsymbol{z}_i, \boldsymbol{\theta}) S^*(T_{2i}|\boldsymbol{z}_i) \right] - \log(R(Y_{Li}|\boldsymbol{z}_i, \boldsymbol{\theta})) - \log(S^*(Y_{Li}|\boldsymbol{z}_i)).$$

Hence, for left-truncated data, the only required external information for the likelihood is the expected hazard at the exact event times, which is the same in the case of right-censored data. On the contrary, interval-censored data requires both $S^*(T_1|\boldsymbol{z}_i)$ and $S^*(T_2|\boldsymbol{z}_i)$ to be specified in order to fit the model, which makes the fitting procedure of parametric relative survival models for interval-censored data more tedious.

References

[1] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

[2] Cancerregisteret. Nye Kræfttilfælde i Danmark (2016), 2017. URL: `https://sundhedsdatastyrelsen.dk/da/tal-og-analyser/analyser-og-rapporter/sygdomme/cancerregisteret`. Accessed 12 April 2018.

[3] L. S. Elting, C. Cooksley, B. N. Bekele, M. Frumovitz, E. B. C. Avritscher, et al. Generalizability of cancer clinical trial results. *Cancer*, 106(11):2452–2458, 2006.

[4] P. Tralongo, L. Dal Maso, A. Surbone, A. Santoro, U. Tirelli, et al. Use of the word "cured" for cancer patients—implications for patients and physicians: the Siracusa charter. *Current Oncology*, 22(1):e38–e40, 2015.

[5] P. Tralongo, M. S. McCabe, and A. Surbone. Challenge for cancer survivorship: improving care through categorization by risk. *Journal of Clinical Oncology*, 35(30):3516–3517, 2017.

[6] A. Surbone, M. A. Annunziata, A. Santoro, U. Tirelli, and P. Tralongo. Cancer patients and survivors: changing words or changing culture? *Annals of Oncology*, 24(10):2468–2471, 2013.

[7] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, and P. C. Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, 2011.

[8] P. D. Baade, D. R. Youlden, and S. K. Chambers. When do I know I am cured? Using conditional estimates to provide better information about cancer survival prospects. *The Medical Journal of Australia*, 194(2):73–77, 2011.

[9] H. C. van Houwelingen and H. Putter. *Dynamic Prediction in Clinical Survival Analysis*. CRC Press, 2012.

[10] S. H. Swerdlow, E. Campo, S. A. Pileri, N. L. Harris, H. Stein, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, 127(20):2375–2390, 2016.

[11] H. Tilly, M. Gomes da Silva, U. Vitolo, A. Jack, M. Meignan, et al. Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 26(Suppl 5):v116–v125, 2015.

[12] M. Dreyling, M. Ghielmini, R. Marcus, G. Salles, U. Vitolo, et al. Newly diagnosed and relapsed follicular lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 25(Suppl 3):iii76–iii82, 2014.

[13] The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *New England Journal of Medicine*, 329(14):987–994, 1993.

[14] M. L. Gjerstorff. The Danish Cancer Registry. *Scandinavian Journal of Public Health*, 39(Suppl 7):42–45, 2011.

[15] C. B. Pedersen. The Danish Civil Registration System. *Scandinavian Journal of Public Health*, 39(Suppl 7):22–25, 2011.

[16] E. Lynge, J. L. Sandegaard, and M. Rebolj. The Danish National Patient Register. *Scandinavian Journal of Public Health*, 39(Suppl 7):30–33, 2011.

[17] K. Helweg-Larsen. The Danish Register of Causes of Death. *Scandinavian Journal of Public Health*, 39(Suppl 7):26–29, 2011.

[18] P. Ingeholm, I. Gögenur, and L. H. Iversen. Danish Colorectal Cancer Group Database. *Clinical Epidemiology*, 8:465–468, 2016.

[19] L. S. G. Østgård, J. M. Nørgaard, K. K. Raaschou-Jensen, R. S. Pedersen, D. Rønnov-Jessen, et al. The Danish National Acute Leukemia Registry. *Clinical Epidemiology*, 8:553–560, 2016.

[20] B. Arboe, P. Josefsson, J. Jørgensen, J. Haaber, P. Jensen, et al. Danish National Lymphoma Registry. *Clinical Epidemiology*, 8:577–581, 2016.

[21] R. Erichsen, T. L. Lash, S. J. Hamilton-Dutoit, B. Bjerregaard, M. Vyberg, et al. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clinical Epidemiology*, 2:51–56, 2010.

[22] J. P. Klein and M. L. Moeschberger. *Survival analysis : techniques for censored and truncated data*. Springer, second edition, 2003.

[23] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[24] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[25] P. Royston and M. K. B. Parmar. 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(15):2175–2197, 2002.

[26] X.-R. Liu, Y. Pawitan, and M. Clements. Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5):1531–1546, 2018.

[27] S. N. Wood. *Generalized additive models : an introduction with R*. CRC Press, second edition, 2017.

[28] Human mortality database. URL: `www.mortality.org`. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

[29] M. Pohar and J. Stare. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81(3):272–278, 2006.

[30] N. E. Breslow, J. H. Lubin, P. Marek, and B. Langholz. Multiplicative models and cohort analysis. *Journal of the American Statistical Association*, 78(381):1–12, 1983.

[31] P. K. Andersen, K. Borch-Johnsen, T. Deckert, A. Green, P. Hougaard, et al. A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, 41(4):921–932, 1985.

[32] T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data*. Springer New York, 2006.

[33] J. Estève, E. Benhamou, M. Croasdale, and L. Raymond. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9(5):529–538, 1990.

[34] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, 2007.

[35] F. Ederer, L. M. Axtell, and S. J. Cutler. The relative survival rate: a statistical methodology. *National Cancer Institute monograph*, 6:101–121, 1961.

[36] F. Ederer and H. Heise. Instructions to IBM 650 programmers in processing survival computations. *Methodological note No. 10, End Results Evaluation Section. Bethesda:National Cancer Institute*, 1959.

[37] T. Hakulinen. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 38(4):933–942, 1982.

[38] M. P. Perme, J. Stare, and J. Estève. On estimation in relative survival. *Biometrics*, 68(1):113–120, 2012.

[39] P. K. Andersen and N. Keiding. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, 31(11-12):1074–1088, 2012.

[40] C. Jackson, J. Stevens, S. Ren, N. Latimer, L. Bojke, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Medical Decision Making*, 37(4):377–390, 2017.

[41] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, M. Lambe, and P. C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, 2013.

[42] P. K. Andersen, M. G. Hansen, and J. P. Klein. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis*, 10(4):335–350, 2004.

[43] P. Royston and M. K. B. Parmar. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30(19):2409–2421, 2011.

[44] V. Bremhorst, M. Kreyenfeld, and P. Lambert. Fertility progression in Germany: an analysis using flexible nonparametric cure survival models. *Demographic Research*, 35:505–534, 2016.

[45] J. W. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11:15–53, 1949.

[46] J. Zhang and Y. Peng. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*, 26(16):3157–3171, 2007.

[47] R. Sposto. Cure model analysis in cancer: an application to data from the Children's Cancer Group. *Statistics in Medicine*, 21(2):293–312, 2002.

[48] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–454, 1999.

[49] P. C. Lambert, J. R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, 2007.

[50] R. Sposto, H. N. Sather, and S. A. Baker. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, 48(1):87–99, 1992.

[51] P. C. Lambert. Modeling of the cure fraction in survival studies. *Stata Journal*, 7(3):351–375, 2007.

[52] R. De Angelis, M. Sant, M. P. Coleman, S. Francisci, P. Baili, et al. Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5—a population-based study. *The Lancet Oncology*, 15(1):23–34, 2014.

[53] P. C. Lambert, P. W. Dickman, C. L. Weston, and J. R. Thompson. Estimating the cure fraction in population-based cancer studies by using finite mixture models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1):35–55, 2010.

[54] V. Bremhorst and P. Lambert. Flexible estimation in cure survival models using Bayesian P-splines. *Computational Statistics & Data Analysis*, 93:270–284, 2016.

[55] K. G. M. Moons, P. Royston, Y. Vergouwe, D. E. Grobbee, and D. G. Altman. Prognosis and prognostic research: what, why, and how? *BMJ (Clinical research ed.)*, 338:b375, 2009.

[56] G. Shmueli. To explain or to predict? *Statistical Science*, 25(3):289–310, 2010. arXiv: `1101.0891v1`.

[57] J. F. Lawless and Y. Yuan. Estimation of prediction error for survival models. *Statistics in Medicine*, 29(2):262–274, 2009.

[58] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *JAMA*, 247(18):2543–2546, 1982.

[59] H. Uno, T. Cai, M. J. Pencina, R. B. D'Agostino, and L. J. Wei. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30(10):1105–1117, 2011.

[60] P. Blanche, A. Latouche, and V. Viallon. Time-dependent AUC with right-censored data: a survey study, 2012. arXiv: `1210.6805`.

[61] H. Hung and C.-T. Chiang. Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics*, 38(1):8–26, 2010.

[62] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[63] U. B. Mogensen, H. Ishwaran, and T. A. Gerds. Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, 50(11):1–23, 2012.

[64] T. A. Gerds and M. Schumacher. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040, 2006.

[65] M. J. Maurer, H. Ghesquières, J.-P. Jais, T. E. Witzig, C. Haioun, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *Journal of Clinical Oncology*, 32(10):1066–1073, 2014.

Part B

Papers

Paper I

Minimal loss of lifetime for patients with diffuse large B-cell lymphoma in remission and event free 24 months after treatment: A Danish population-based study

Lasse H. Jakobsen[1,2], Martin Bøgsted[1,2], Peter dN. Brown[3], Bente Arboe[3], Judit Jørgensen[4], Thomas S. Larsen[5], Maja B. Juul[5], Lene Schurmann[6], Linda Højberg[7], Olav J. Bergmann[8], Therese Lassen[9], Pär L. Josefsson[10], Paw Jensen[1], Hans E. Johnsen[1,2], and Tarec C. El-Galaly[1,2]

1. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; 2. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; 3. Department of Hematology, Copenhagen University Hospital, Copenhagen, Denmark; 4. Department of Hematology, Aarhus University Hospital, Aarhus, Denmark; 5. Department of Hematology, Odense University Hospital, Odense, Denmark; 6. Department of Medicine, Hospitalsenheden Vest, Holstebro, Denmark; 7. Department of Hematology, Sydvestjysk Sygehus, Esbjerg, Denmark; 8. Department of Hematology, Vejle Sygehus, Vejle, Denmark; 9. Department of Hematology, Roskilde Sygehus, Roskilde, Denmark; 10. Department of Hematology, Herlev Hospital, Herlev, Denmark.

Published in Journal of Clinical Oncology (2017) 35(7):778-784.

Description

In this paper, we compared the survival of Danish diffuse large B-cell lymphoma patients with that of the Danish matched general population by utilizing data from the Danish Lymphoma Registry. In addition, death causes and updated relapses in deceased patients were obtained by reviewing patient reports within each hematology department in Denmark. The aim of the study was to evaluate the patient survival as compared to the general population and determine how this evolves as patients remain event-free after successful treatment with the standard R-CHOP therapy. Thus, the study included only adult patients who achieved complete remission/unconfirmed after first line treatment.

VOLUME 35 · NUMBER 7 · MARCH 1, 2017

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Minimal Loss of Lifetime for Patients With Diffuse Large B-Cell Lymphoma in Remission and Event Free 24 Months After Treatment: A Danish Population-Based Study

Lasse Hjort Jakobsen, Martin Bøgsted, Peter de Nully Brown, Bente Arboe, Judit Jørgensen, Thomas Stauffer Larsen, Maja Bech Juul, Lene Schurmann, Linda Højberg, Olav Jonas Bergmann, Therese Lassen, Pär Lars Josefsson, Paw Jensen, Hans Erik Johnsen, and Tarec Christoffer El-Galaly

Author affiliations and support information (if applicable) appear at the end of this article.

Published at jco.org on January 17, 2017.

Corresponding author: Lasse Hjort Jakobsen, MSc, Department of Hematology, Aalborg University Hospital; e-mail: lasse.j@rn.dk

0732-183X/17/3507w-778w/$20.00

ABSTRACT

Purpose
The general outlook for patients with diffuse large B-cell lymphoma (DLBCL) in first remission is important information for patients and for planning post-treatment follow-up. The purpose of this study was to evaluate the survival of patients with DLBCL in remission compared with a matched general population.

Methods
A total of 1,621 patients from the Danish Lymphoma Registry who were newly diagnosed with DLBCL between 2003 and 2011 were included in this study. All patients were ≥ 16 years of age at diagnosis and had achieved complete remission or complete remission unconfirmed after first-line rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) or R-CHOP–like therapy.

Results
The 5-year post-treatment DLBCL survival was inferior to survival in the matched general population (78%; 95% CI, 76 to 80; *v* 87%; standardized mortality ratio, 1.75; $P < .001$). Excess mortality was present but reduced for patients achieving post-treatment event-free survival for 24 months (pEFS24; standardized mortality ratio, 1.27; $P < .001$). In age-stratified analyses, the survival of patients < 50 years of age was normalized to the general population after achieving pEFS24 ($P = .99$). During the first 8 years after pEFS24, the average loss of lifetime was 0.31 mo/y (95% CI, 0.11 to 0.50 mo/y). Excess mortality diminished when analyzing death from lymphoma as competing event to death from other causes, suggesting that early and late relapse is responsible for increased mortality in patients with DLBCL.

Conclusion
Although this population-based study does not support complete normalization of survival for patients with DLBCL achieving pEFS24, the estimated loss of residual lifetime was low for patients in continuous remission 2 years after ending treatment. Therefore, pEFS24 is an appealing and relevant milestone for patient counseling and could be a surrogate end point in clinical trials.

J Clin Oncol 35:778-784. © 2017 by American Society of Clinical Oncology

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin lymphoma (NHL) in the Western world.[1] The addition of rituximab to cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) has increased DLBCL survival substantially, but 30% to 40% of patients are refractory to treatment or relapse after initial response to therapy.[2-5] Patients with relapsed/refractory DLBCL have dismal outcomes, and only a fraction of patients with early relapse are cured by intensive salvage therapies.[6] Thus, achieving durable remission early is crucial for favorable long-term outcomes in DLBCL. The chance of surviving DLBCL beyond a given time point (eg, 4-year survival) can be assessed from clinical prognostic scores, such as the International Prognostic Index (IPI), revised IPI (R-IPI), and National Comprehensive Cancer Network–IPI.[7-9] Although practical, these models oversimplify individual prognostic information by pooling noncomparable patients into a few groups, making the resulting group-specific

ASSOCIATED CONTENT

Data Supplement
DOI: 10.1200/JCO.2016.70.0765

DOI: 10.1200/JCO.2016.70.0765

survival rates inappropriate for patient counseling. For example, consider two patients with identical IPI scores who are both set to receive R-CHOP. The first patient is an 80-year-old with poor performance status due to comorbidity (Eastern Cooperative Oncology Group 2), elevated lactate dehydrogenase, and stage III disease. The second patient is a 25-year-old with stage IV disease, extensive extranodal involvement, elevated lactate dehydrogenase, and lymphoma-related reduced performance (Eastern Cooperative Oncology Group 2). Both are high risk according to the R-IPI model, with an estimated 4-year overall survival (OS) of 55%.[8] However, the normal 4-year life expectancy for healthy 25-year-olds and 80-year-olds are not comparable, and, without this context, meaningful interpretation of prognostic information becomes difficult for the patients. Normalization to the survival of a matched general population was recently proposed as an alternative to classic time to event analysis in DLBCL. The pivotal study by Maurer et al[10] showed that event-free survival for 24 months (EFS24) is an important milestone for patients with DLBCL, and survival beyond this time point is equivalent to that of the general population. If confirmed, EFS24 will be important, not only for daily patient consultation but also for planning disease surveillance strategies and as a possible surrogate end point in clinical trials. In the current study, a similar post-treatment EFS24 milestone was evaluated in a population-based study of patients with DLBCL in first remission.

METHODS

The Danish Lymphoma Registry and Patient Inclusion

This is a retrospective study based on the Danish National Lymphoma Registry (LYFO). LYFO has been nationwide since 2000 and contains detailed information on patients with lymphoma diagnosed and treated at Danish hematology departments. Local hematologists prospectively report a wide range of clinicopathologic features to LYFO, which also captures information on treatment, the response to treatment, relapse, and vital status. Arboe et al[11] recently validated the data quality in LYFO. The coverage was 94.9%, and the accuracy concerning lymphoma histology and treatment information was 99.9% and 99.3%, respectively. Dates of death for deceased patients are obtained by periodic merging with the Danish Civil Registration System using the unique Danish personal identification numbers.[12] Relapses are also reported to LYFO by local departments. Safeguarding measures against missing registrations included notification if biopsy-confirmed lymphoma relapse was identified from automatized queries to the Danish National Pathology Registry or if chemotherapy and/or radiotherapy prescriptions occurred ≥ 9 months after the first pathologic diagnosis of DLBCL. Information on cause of death is often incomplete in LYFO, and, as a part of this study, we performed a national review of patient records to determine cause of death whenever possible.

The inclusion criteria for the current study were: ≥ 16 years of age at diagnosis, newly diagnosed DLBCL between 2003 and 2011, and complete remission (CR) or complete remission unconfirmed (CRu) after first-line treatment with R-CHOP or equivalently effective regimens. The standardized response criteria for NHL were used if stand-alone computed tomography (CT) was performed for the final response assessment, and the revised response criteria for malignant lymphoma were used if positron emission tomography (PET)/CT was performed for the final response assessment.[13-15] LYFO does not contain information on the imaging modality used for response assessment, but hospitals were reimbursed for PET/CT within the first 120 days of completing therapy in > 50% of patients with DLBCL in CR from 2007 onward.

Statistics and Ethics

Post-treatment OS (pOS) was defined as the time from end of treatment until death or censoring in patients still alive at last follow-up (September 6, 2015). Post-treatment event-free survival (pEFS) was defined as the time from end of treatment until death, relapse, or censoring, whichever came first. The survival of patients with DLBCL at different pEFS milestones (6, 24, 36, 48, and 60 months) was computed using the Kaplan-Meier method.[16] A general population derived from Danish life tables[17] was matched on sex, age, and calendar year using a conditional approach.[18] Excess mortality in patients with DLBCL was calculated as the standardized mortality ratio (SMR)—the ratio of observed to expected mortality.[19] A proportional hazards competing risks regression model was used to assess associations between relapse risk and baseline prognostic factors.[20] The 10-year loss of lifetime for patients with DLBCL relative to the general population was calculated as the area between the patient and the population-specific survival curves from study entry until 10 years of follow-up, at which time many patients were still at risk. Causes of death at different pEFS time points were assessed by cumulative incidences from a competing risks analysis. For survival comparisons from a pEFS milestone (eg, pEFS24), only patients achieving that milestone were analyzed. The survival clock was started at the milestone of interest and a general population rematched to the patient age and calendar time at that milestone. Statistical analyses were conducted in R (version 3.3.1). Double-sided P values ≤ .05 were considered significant. The study was approved by the Danish Data Protection Agency (2008-58-0028) and the Danish Health and Medicines Authority (3-3013-1373/1).

RESULTS

Patients and Outcomes

A total of 1,621 Danish patients with DLBCL met the inclusion criteria (Data Supplement). The clinicopathologic characteristics are listed in Table 1. With a median follow-up of 85 months (reverse Kaplan-Meier method[21]), the 5-year pOS was significantly lower for the patients with DLBCL than for the matched general population (78%; 95% CI, 76 to 80; v 87%). The baseline (end of treatment) SMR was 1.75 (95% CI, 1.60 to 1.91; $P < .001$; Fig 1A), which was reduced to 1.27 (95% CI, 1.12 to 1.44; $P < .001$; Fig 1B) for patients achieving pEFS24 and 1.32 (95% CI, 1.11 to 1.54; $P < .001$; Fig 1C) for patients achieving pEFS48. Age-stratified analyses (< 50 years v ≥ 50 years) revealed early normalization of survival in the younger patients who achieved pEFS24 (5-year pOS, 99% v 99%; SMR, 1.11; 95% CI, 0.22 to 3.25; $P = .99$; Data Supplement). In contrast, patients with DLBCL ≥ 50 years were at increased risk of death even after achieving pEFS60 (SMR, 1.36; 95% CI, 1.12 to 1.63; $P = .001$). In addition, after achieving pEFS6, only eight (4%) of 222 patients < 50 years old died during the follow-up period.

Loss of Residual Lifetime Estimation and Event Decomposition

For all patients with DLBCL in CR or CRu after treatment, residual lifetime was reduced by 1.07 mo/y (95% CI, 0.87 to 1.27 mo/y) within the first 10 years after treatment compared with the expected survival in the general population (Data Supplement). In patients achieving pEFS24, pEFS36, and pEFS48, the estimates were 0.31 mo/y (95% CI, 0.11 to 0.50 mo/y), 0.29 mo/y (95% CI, 0.09 to 0.49 mo/y), and 0.29 mo/y (95% CI, 0.08 to 0.51 mo/y), respectively. In age-stratified analyses (< 50 years, 50 to 60 years,

Table 1. Clinicopathologic and Demographic Characteristics of Patients With DLBCL in CR or CRu Diagnosed Between 2003 and 2011 (n = 1,621)

Characteristic	Summary (n = 1,621)	Missing Information
Median age, years (range)	65 (16-92)	0
Age groups, years		
16-40	88 (5.4)	
40-50	139 (8.6)	
50-75	1,101 (67.9)	
≥ 75	293 (18.1)	
Sex ratio, M/F	1.20	0
Ann Arbor stage > 2	886 (55.1)	14
IPI > 2	526 (34.0)	72
ECOG performance > 1	185 (11.5)	14
B symptoms	663 (41.5)	25
Elevated LDH	743 (47.1)	45
Extranodal	952 (58.7)	0
Radiotherapy	626 (38.6)	0
Chemotherapy		0
CEOP	25 (1.5)	
CHOEP	129 (8.0)	
CHOP	1,467 (90.5)	

NOTE. Data presented as No. (%) unless otherwise noted.
Abbreviations: CEOP, cyclophosphamide, etoposide, vincristine, prednisone; CHOEP, cyclophosphamide, doxorubicin, vincristine, etoposide, prednisone; CHOP, cyclophosphamide, doxorubicin, vincristine, prednisone; CR, complete remission; CRu, complete remission unconfirmed; DLBCL, diffuse large B-cell lymphoma; ECOG, Eastern Cooperative Oncology Group; IPI, International Prognostic Index; LDH, lactate dehydrogenase.

and > 60 years), patients younger than 50 years quickly normalized to the general population, whereas patients age ≥ 50 years had continuously increased mortality (Data Supplement). Loss of lifetime was more pronounced in the high-risk patients but was reduced for all patients achieving pEFS24 regardless of risk group (Fig 2). The absolute difference in loss of lifetime for patients with high-risk IPI versus patients with low-risk IPI disease was also significantly reduced at pEFS24, suggesting that pretreatment risk factors become prognostically less important with long-term remission. No difference in loss of lifetime was observed between male and female patients achieving pEFS24 (Fig 2).

The 5-year cumulative incidence of DLBCL relapse was 18% (95% CI, 16% to 20%), which was reduced to 8% (95% CI, 7% to 10%) in patients achieving pEFS24 (Fig 3). From end of treatment, lymphoma (relapse or treatment toxicity) was the most common cause of death (40%; Data Supplement). The 2-year cumulative risk of death from other cancers or cardiovascular disease was 1.2% (95% CI, 0.8% to 1.9%) and 0.5% (95% CI, 0.2% to 1.0%), respectively. For patients achieving pEFS48, the 2-year cumulative risk of death from other cancers or cardiovascular diseases increased to 2.4% (95% CI, 1.6% to 3.7%) and 1% (95% CI, 0.5% to 2%), respectively. From that point on, the risk of death from other cancers was similar to the risk of experiencing DLBCL relapse. Thirteen (16%) of 83 deaths from other cancers were attributed to myelodysplastic syndrome (n = 5) or acute myeloid leukemia (n = 8).

A competing risks model was used to determine the impact of recurrent DLBCL on post-treatment DLBCL survival (Data Supplement). In this analysis, death from relapsed DLBCL (including immediate treatment complications) was treated as a competing event to death from other causes. Using this approach, the survival of patients with DLBCL was equivalent to the survival of the general population at the end of treatment, suggesting that the observed excess mortality in patients with DLBCL was fully explained by early and late relapse. The 5-year risk of relapse after achieving pEFS24 was 4% (95% CI, 2% to 8%), 7% (95% CI, 5% to 11%), and 10% (95% CI, 8% to 12%) for patients younger than 50 years, 50 to 60 years old, and older than 60 years, respectively (Data Supplement). Age older than 60 years, advanced stage, and IPI > 2 were associated with increased risk of late relapse in a multivariate analysis of patients achieving pEFS24 (Table 2).

DISCUSSION

In this study, pEFS24 was evaluated as a milestone for patients with DLBCL in CR or CRu after first-line treatment with R-CHOP (–like) therapy. We confirmed that patients achieving pEFS24 have favorable outcomes, although the present population-based

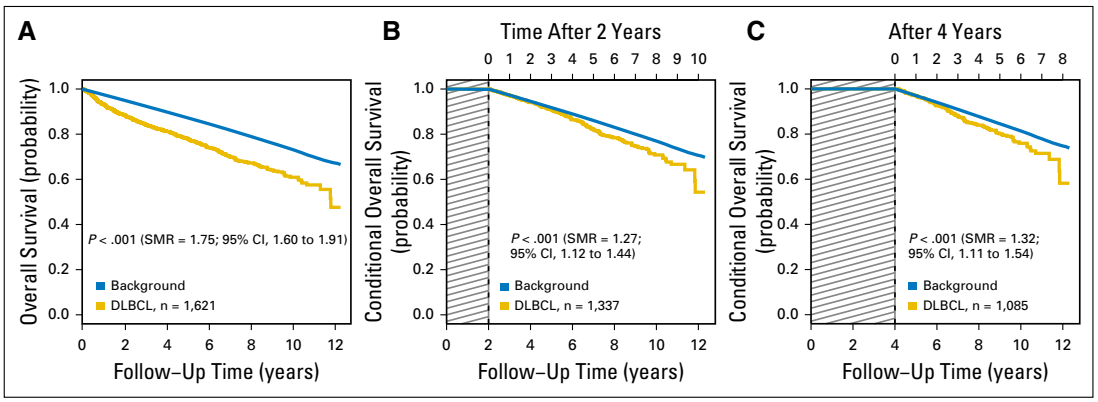

Fig 1. Overall survival of diffuse large B-cell lymphoma (DLBCL) in first complete remission or complete remission unconfirmed versus the expected survival on the basis of the general population for (A) all patients; (B) patients who achieve post-treatment event-free survival for 24 months; and (C) patients achieving post-treatment event-free survival for 48 months. SMR, standardized mortality ratio.

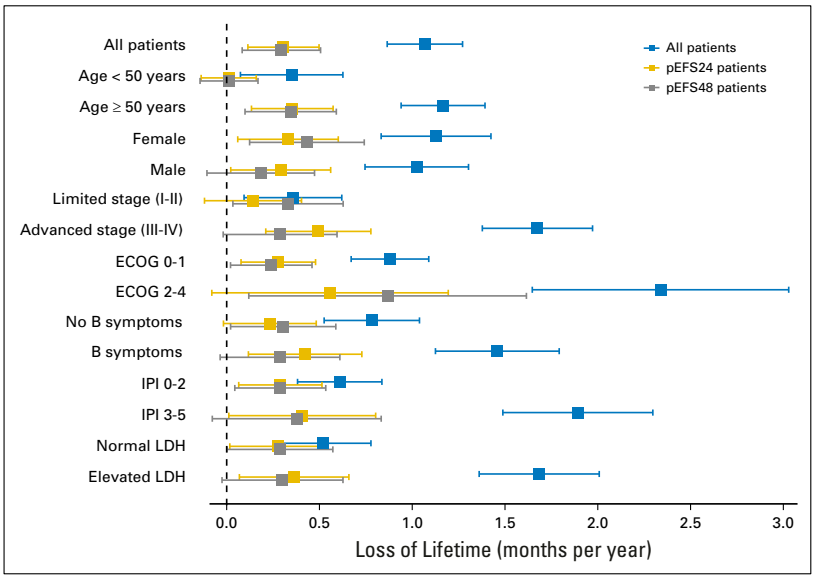

Fig 2. Loss of lifetime in diffuse large B-cell lymphoma subgroups for all patients, patients who achieve post-treatment event-free survival for 24 months (pEFS24), and patients achieving post-treatment event-free survival for 48 months (pEFS48). ECOG, Eastern Cooperative Oncology Group; IPI, International Prognostic Index; LDH, lactate dehydrogenase.

analysis found a slightly reduced survival of patients with DLBCL relative to the general population despite many years in CR or CRu. This finding was age dependent, because the survival of patients younger than 50 years at diagnosis equalized to that of the general population by pEFS24, compared with the continuous increased mortality in middle-age and elderly patients. The excess mortality was mainly driven by relapsed DLBCL; risk factors for DLBCL relapse after pEFS24 were age older than 60 years, advanced stage, and IPI > 2. Thus, predictors of outcome in this situation are similar to well-characterized prognostic factors at baseline.[7,8]

To the best of our knowledge, this is the first population-based study to investigate residual lifetime in patients with DLBCL event free at specified time points after treatment, but it is not the first study of DLBCL survival compared with a matched general population. Maurer et al[10] analyzed patients with DLBCL enrolled in the North Central Cancer Treatment Group NCCTG-N0489 phase II study (n = 87) and the University of Iowa/Mayo Clinic Specialized Programs of Research Excellence Molecular Epidemiology Resource (MER) registry (n = 680). The MER registry included patients treated at the Mayo Clinic or the University of Iowa, or patients referred for a second opinion at the Mayo Clinic or the University of Iowa while treated locally in the community. All patients in the MER registry were enrolled prospectively within 9 months of diagnosis; the majority (80%) of patients were from the upper Midwestern US (MN, IA, IL, WI; M. J. Maurer, personal communication, September 2016). For the entire DLBCL cohort, as well as patient subgroups, the survival of patients achieving the EFS24 milestone normalized to the general US population.[10] Maurer et al[10] successfully reproduced the results in a French DLBCL cohort consisting of patients from clinical trials sponsored by Groupe d'Etude des Lymphomes de l'Adulte (n = 600) and from a hospital-based registry in Lyon (n = 220). Important differences are present in the designs of the study by Maurer et al[10] and the current study. First, Maurer et al[10] used conventional OS and EFS end points (ie, measured from diagnosis), whereas we used pOS and pEFS. The different approach was chosen because our motive was to examine the survival of patients with DLBCL responding with CR or CRu to immunochemotherapy. By definition, this group of patients survived first-line treatment, which could introduce guarantee-time bias if using the time of diagnosis as the entry point when comparing this group to the general population (ie, patients with DLBCL are guaranteed survival for at least 3 months, but this is not the case for the general population). Second, Maurer et al[10] included unplanned treatment as an event in the EFS24 analyses, but the only events of interest in the current study were relapse or death from any cause. However, restricting our analysis to patients with CR or CRu made unplanned treatment in the absence of documented relapse unlikely, and the more conservative EFS definition will not bias toward later normalization of survival. Notably, striking differences were found in the baseline SMRs reported for both US and French patients (2.88 and 4.99, respectively) compared with the 1.75 in our patients. This difference is likely explained by the better outlook for a group of patients defined by CR or CRu.

Less homogeneity between patients with DLBCL and the matched general population may fully or partially explain the conflicting observations between Maurer et al[10] and the current study. Patients treated or seen for a second opinion at tertiary care centers, such as the Mayo Clinic, and patients enrolled in clinical trials may have better overall health than the US general population to which they were compared. This could bias toward earlier normalization of survival for the patients with DLBCL. Given that cancer treatment in Denmark is publicly funded with equal access to all citizens, the patients with DLBCL analyzed in our study may better resemble the general population regarding general health and access to health care. Notably, the 5-year relapse rate in the

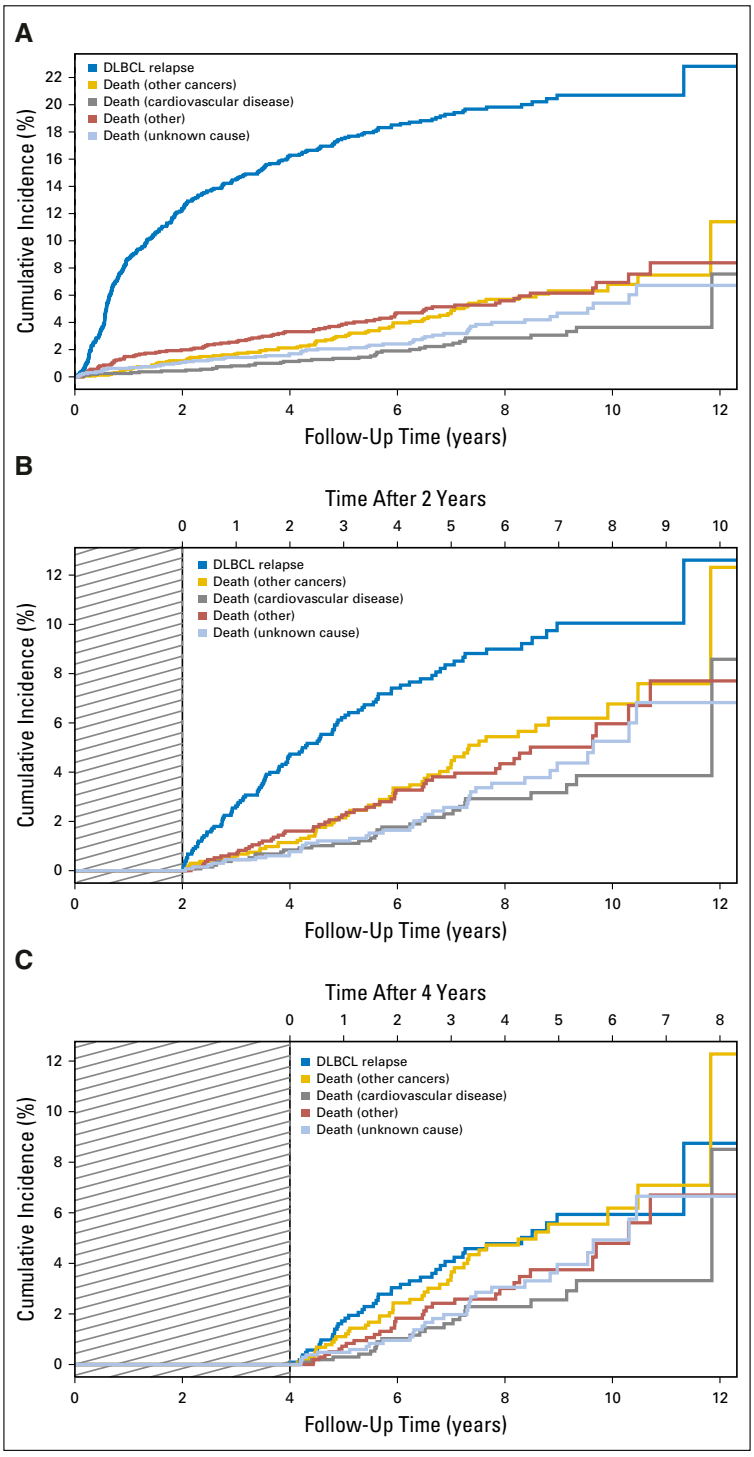

Fig 3. Event-specific cumulative incidences based on a competing risks analysis for all patients with diffuse large B-cell lymphoma (DLBCL) in (A) first complete remission or complete remission unconfirmed; (B) patients achieving post-treatment event-free survival for 24 months; and (C) patients achieving post-treatment event-free survival for 48 months.

Table 2. Time-to-Relapse Determinants From Competing Risk Regression for Patients With DLBCL in First CR or CRu Who Achieved EFS24 (n = 1, 337)

Determinant	Univariate HR (95% CI)	*P*	Multivariate HR (95% CI)	*P*
Age > 60 years	1.73 (1.10 to 2.72)	.02	1.86 (1.18 to 2.91)	.01
Sex	1.08 (0.73 to 1.59)	.70		
Advanced stage	3.41 (2.19 to 5.32)	.00	4.31 (2.61 to 7.12)	.00
IPI > 2	1.58 (1.06 to 2.36)	.02	0.65 (0.42 to 1.02)	.06
ECOG performance > 1	1.29 (0.72 to 2.32)	.39		
B symptoms	1.08 (0.73 to 1.58)	.71		
Elevated LDH	1.11 (0.75 to 1.64)	.60		

NOTE. Significant effects in the univariate analyses were included in the multivariate analysis.
Abbreviations: CR, complete remission; CRu, complete remission unconfirmed; DLBCL, diffuse large B-cell lymphoma; ECOG, Eastern Cooperative Oncology Group; EFS24, event-free survival for 24 months; HR, subdistribution hazard ratio; IPI, International Prognostic Index; LDH, lactate dehydrogenase

current study and reported by Maurer et al[10] in patients achieving pEFS24 and EFS24, respectively, was 8%. This suggests that there is some comparability of the disease characteristics of patients analyzed in the two studies.

Consistent with the observations of the current study, an abstract based on data from the BC Cancer Agency Lymphoid Cancer Database also reported persistent excess mortality for patients achieving conventional EFS24.[22] However, unlike the current study, the excess mortality was explained not only by the continuous risk of relapse but also by increased risk of death from other causes. Interestingly, cell of origin data were available for 601 patients in the Canadian study. Analysis of these patients revealed normalization of survival at EFS24 for patients with germinal center B-cell–like DLBCL but not patients with non–germinal center B-cell–like/activated B-cell–like DLBCL.[22] In a recent study by Smeland et al,[23] the survival of Norwegian patients undergoing high-dose therapy with autologous stem cell transplantation for various NHLs was compared with survival in the general population. In that study, patients with DLBCL surviving the first 5 years after therapy normalized to the general population (SMR, 1.7; 95% CI, 0.9 to 3.1).[23] However, there could be a lack of statistical power due to fewer patients at risk after 5 years.

A similar study of patients with Hodgkin lymphoma (HL) diagnosed between 1989 and 2012 who were treated with doxorubicin, bleomycin, vinblastine, and dacarbazine (or equivalently effective regimes) in British Columbia found increased mortality, even after achieving EFS60, which may seem counterintuitive given the favorable outcome of HL and better salvage opportunities.[24] However, the risk of relapse in patients achieving EFS24 was low (5.6% 5-year cumulative incidence), and the authors concluded that EFS24 could be a potential surrogate end point in clinical trials of HL as proposed for DLBCL by Maurer et al.[10,24] With the low loss of lifetime for patients with DLBCL achieving pEFS24 in our study, milestones such as pEFS24 may be surrogates for OS and progression-free survival in clinical trials involving patients with DLBCL. However, studies that demonstrate sufficient correlation between effect measures on the basis of pEFS24 and classic end points are warranted.[25] In particular, precautions should be taken against using pEFS24 as a surrogate end point if there are concerns of severe long-term toxicity or if survival curves could later cross during follow-up. The recently published long-term results of a study comparing three cycles of CHOP plus radiotherapy to eight cycles of CHOP illustrates this concern.[26] The 5-year progression-free survival and OS were in favor of three cycles of CHOP plus radiotherapy, but corresponding 10-year estimates were similar in the two treatment arms.[26] Acknowledging the shortcomings of EFS24-based effect measures, it is clear that less expensive setups for clinical trials with faster reporting of results are welcome in the early era of personalized medicine and at a time when the therapeutic armamentarium for lymphoid malignancies is rapidly expanding.

pEFS24 or EFS24 may also be useful for planning follow-up programs, and the low relapse risk in young patients with DLBCL achieving pEFS24 indicates that less-intensive disease surveillance strategies may be pursued for this subgroup of patients. In terms of patient counseling, survival relative to age- and sex-matched persons without DLBCL is more relevant than crude 5-year survival probabilities.

The strengths of the current study are the use of population-based data, long follow-up, and complete information on vital status. However, the median follow-up of 7 years is too short to estimate the impact of DLBCL on residual lifetime for younger patients. In particular, the impact of late toxicities is not covered by this study. Another strength of this study is similar regional outcomes of DLBCL in Denmark, permitting the pooling of data into a single study.[27] The increasing use of PET/CT during the inclusion period could have changed the quality of remission assessment and therapy outcome of patients concluded to be in remission. However, a stratified analysis of patients with early (2003 to 2006) and late diagnosis (2007 to 2011) revealed no clinically relevant differences in loss of lifetime in patients achieving pEFS24 (0.07 mo/y; 95% CI, −0.23 to 0.37 mo/y).

In conclusion, we showed that a persistent risk of relapse prevents normalization of survival for patients with DLBCL despite years in remission. Nevertheless, the loss of residual lifetime for patients achieving pEFS24 is minimal and, although significant, may not be clinically relevant to individual patients. However, relying on pEFS24 or EFS24 as end points in clinical trials could potentially miss the impact of late toxicities and other late adverse effects on patient outcomes.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Lasse Hjort Jakobsen, Martin Bøgsted, Peter de Nully Brown, Hans Erik Johnsen, Tarec Christoffer El-Galaly
Collection and assembly of data: Lasse Hjort Jakobsen, Peter de Nully Brown, Bente Arboe, Judit Jørgensen, Thomas Stauffer Larsen, Maja Bech Juul, Lene Schurmann, Linda Højberg, Olav Jonas Bergmann, Therese Lassen, Pär Lars Josefsson, Paw Jensen, Tarec Christoffer El-Galaly
Data analysis and interpretation: Lasse Hjort Jakobsen, Martin Bøgsted, Tarec Christoffer El-Galaly
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Armitage JO, Weisenburger DD: New approach to classifying non-Hodgkin's lymphomas: Clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. J Clin Oncol 16:2780-2795, 1998

2. Ziepert M, Hasenclever D, Kuhnt E, et al: Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. J Clin Oncol 28:2373-2380, 2010

3. Coiffier B, Lepage E, Briere J, et al: CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med 346:235-242, 2002

4. Pfreundschuh M, Trümper L, Osterborg A, et al: CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: A randomised controlled trial by the MabThera International Trial (MInT) Group. Lancet Oncol 7: 379-391, 2006

5. Pfreundschuh M, Schubert J, Ziepert M, et al: Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: A randomised controlled trial (RICOVER-60). Lancet Oncol 9:105-116, 2008

6. Gisselbrecht C, Glass B, Mounier N, et al: Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. J Clin Oncol 28:4184-4190, 2010

7. The International Non-Hodgkin's Lymphoma Prognostic Factors Project: A predictive model for aggressive non-Hodgkin's lymphoma. N Engl J Med 329:987-994, 1993

8. Sehn LH, Berry B, Chhanabhai M, et al: The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood 109:1857-1861, 2007

9. Zhou Z, Sehn LH, Rademaker AW, et al: An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. Blood 123:837-842, 2014

10. Maurer MJ, Ghesquières H, Jais J-P, et al: Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. J Clin Oncol 32:1066-1073, 2014

11. Arboe B, El-Galaly TC, Clausen MR, et al: The Danish National Lymphoma Registry: Coverage and data quality. PLoS One 11:e0157999, 2016

12. Pedersen CB, Gøtzsche H, Møller JO, et al: The Danish Civil Registration System. A cohort of eight million persons. Dan Med Bull 53:441-449, 2006

13. Cheson BD, Horning SJ, Coiffier B, et al: Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. J Clin Oncol 17:1244, 1999

14. Reference deleted.

15. Cheson BD, Pfistner B, Juweid ME, et al: Revised response criteria for malignant lymphoma. J Clin Oncol 25:579-586, 2007

16. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. J Am Stat Assoc 53: 457-481, 1958

17. Statistics Denmark: STATBANK. www.statbank.dk/statbank5a/default.asp?w=1920

18. Verheul HA, Dekker E, Bossuyt P, et al: Background mortality in clinical survival studies. Lancet 341:872-875, 1993

19. Breslow NE, Lubin JH, Marek P, et al: Multiplicative models and cohort analysis. J Am Stat Assoc 78:1-12, 1983

20. Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94:496-509, 1999

21. Shuster JJ: Median follow-up in clinical trials. J Clin Oncol 9:191-192, 1991

22. Srour L, Zheng YY, Gerrie AS, et al: EFS24 as a predictor of outcome in a population-based cohort of patients with DLBCL in British Columbia (BC). J Clin Oncol 34, 2016 (suppl; abstr 7569)

23. Smeland KB, Kiserud CE, Lauritzsen GF, et al: A national study on conditional survival, excess mortality and second cancer after high dose therapy with autologous stem cell transplantation for non-Hodgkin lymphoma. Br J Haematol 173:432-443, 2016

24. Hapgood G, Zheng Y, Sehn LH, et al: Evaluation of the risk of relapse in classical Hodgkin lymphoma at event-free survival time points and survival comparison with the general population in British Columbia. J Clin Oncol 34:2493-2500, 2016

25. Fleming TR, Powers JH: Biomarkers and surrogate endpoints in clinical trials. Stat Med 31: 2973-2984, 2012

26. Stephens DM, Li H, LeBlanc ML, et al: Continued risk of relapse independent of treatment modality in limited-stage diffuse large B-cell lymphoma: Final and long-term analysis of Southwest Oncology Group Study S8736. J Clin Oncol 34: 2997-3004, 2016

27. Brown P, Jørgensen J, Munksgaard L, et al: Annual Report of The Danish Lymphoma Group 2011. www.lymphoma.dk/wp-content/uploads/2016/11/Aarsrapport_2011_ver_1_3.pdf

Affiliations

Lasse Hjort Jakobsen, **Martin Bøgsted**, **Paw Jensen**, **Hans Erik Johnsen,** and **Tarec Christoffer El-Galaly**, Aalborg University Hospital; **Lasse Hjort Jakobsen**, **Martin Bøgsted**, **Hans Erik Johnsen**, and **Tarec Christoffer El-Galaly**, Aalborg University, Aalborg; **Peter de Nully Brown** and **Bente Arboe**, Copenhagen University Hospital, Copenhagen; **Judit Jørgensen**, Aarhus University Hospital, Aarhus; **Thomas Stauffer Larsen** and **Maja Bech Juul**, Odense University Hospital, Odense; **Lene Schurmann**, Hospitalsenheden Vest, Holstebro; **Linda Højberg**, Sydvestjysk Sygehus, Esbjerg; **Olav Jonas Bergmann**, Vejle Sygehus, Vejle; **Therese Lassen**, Roskilde Sygehus, Roskilde; and **Pär Lars Josefsson**, Herlev Hospital, Herlev, Denmark.

Support

Supported by the Danish Lymphoma Group (L.H.J.).

Prior Presentation

Presented in part at the Annual Congress of the European Hematology Association, Copenhagen, Denmark, June 9-12, 2016; the annual meeting of the Danish Society of Hematology, Middlefart, Denmark, March 18-19, 2016; and the annual Nordic Lymphoma Group meeting, Oslo, Norway, November 5-6, 2015.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Minimal Loss of Lifetime for Patients With Diffuse Large B-Cell Lymphoma in Remission and Event Free 24 Months After Treatment: A Danish Population-Based Study

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Lasse Hjort Jakobsen
No relationship to disclose

Martin Bøgsted
No relationship to disclose

Peter de Nully Brown
No relationship to disclose

Bente Arboe
Travel, Accommodations, Expenses: Sobi, Takeda Pharmaceutical Company Ltd, Gilead Sciences

Judit Jørgensen
Travel, Accommodations, Expenses: Roche, Gilead Sciences, Quintiles

Thomas Stauffer Larsen
Consulting or Advisory Role: Roche, Takeda Pharmaceutical Company Ltd
Travel, Accommodations, Expenses: Roche, Sanofi, Takeda Pharmaceutical Company Ltd, Novartis

Maja Bech Juul
No relationship to disclose

Lene Schurmann
No relationship to disclose

Linda Højberg
No relationship to disclose

Olav Jonas Bergmann
No relationship to disclose

Therese Lassen
No relationship to disclose

Pär Lars Josefsson
No relationship to disclose

Paw Jensen
No relationship to disclose

Hans Erik Johnsen
No relationship to disclose

Tarec Christoffer El-Galaly
Travel, Accommodations, Expenses: Roche, Takeda Pharmaceutical Company Ltd, Novartis

Supplementary

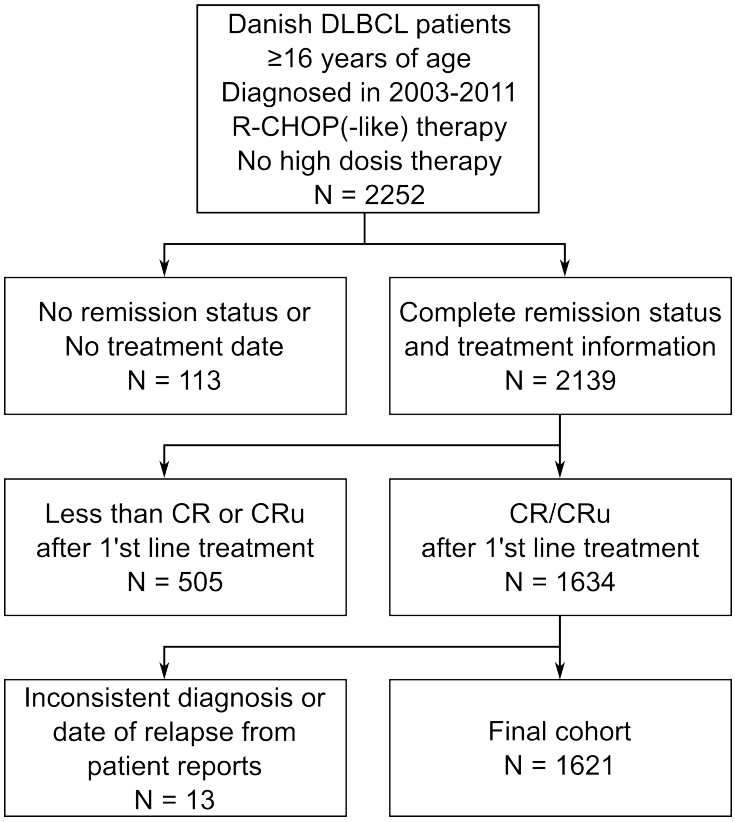

Figure S1: CONSORT diagram of the Danish DLBCL patients.

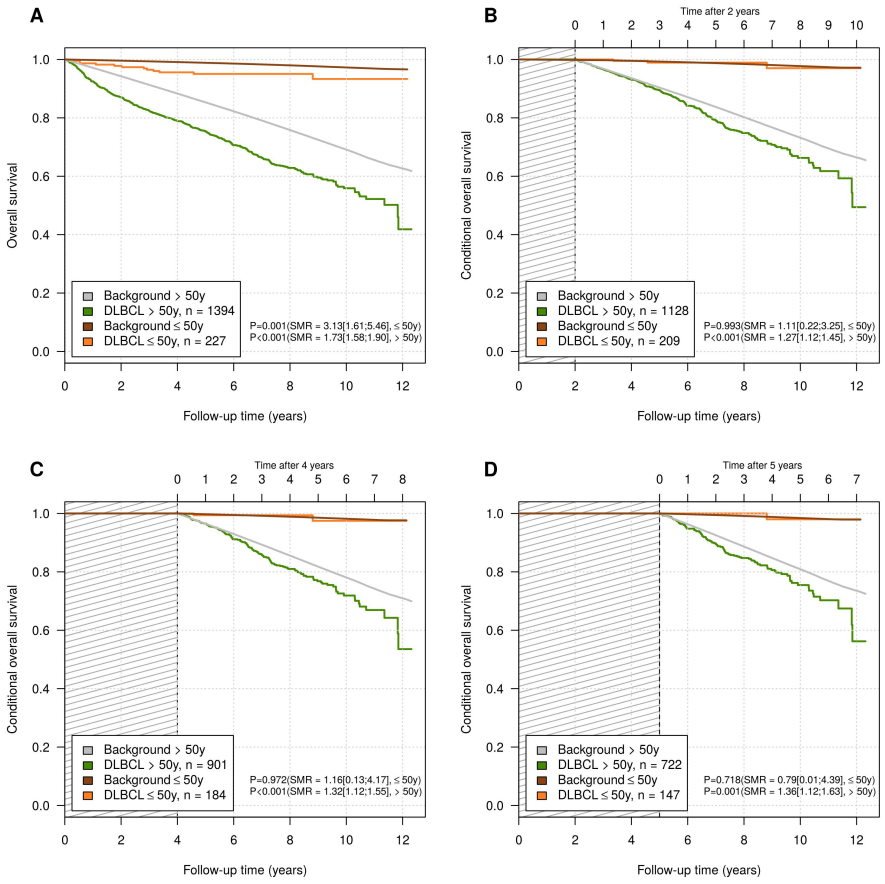

Figure S2: Overall survival of DLBCL in 1st CR/CRu versus expected survival as determined by the general population stratified on age (50< and 50≥) A) for all patients, B) patients that achieve pEFS24, C) patients achieving pEFS48, and D) patient achieving pEFS60.

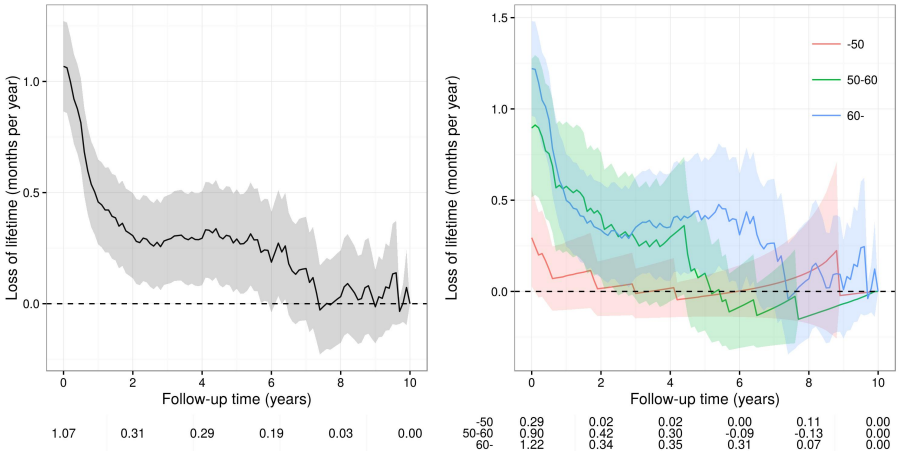

Figure S3: Loss of lifetime (months per year) for the entire DLBCL cohort and age-specific subgroups (-50, 50-60, 60-) computed by the area under the survival curves up to ten years. For a given time point, the loss of lifetime is computed using all patients achieving pEFS at that time point.

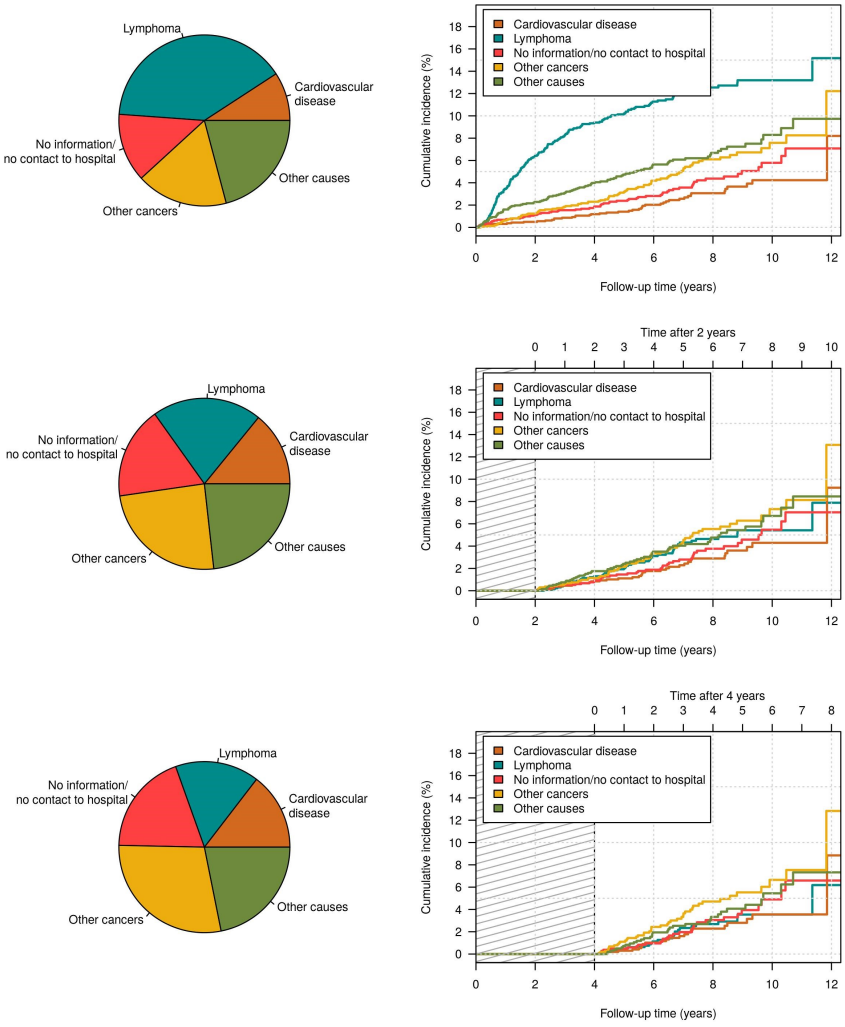

Figure S4: Cause of death pie chart and corresponding cumulative incidences from a competing risks analysis computed for all DLBCL patients in 1st CR/CRu, patients achieving pEFS24, and patients achieving pEFS48.

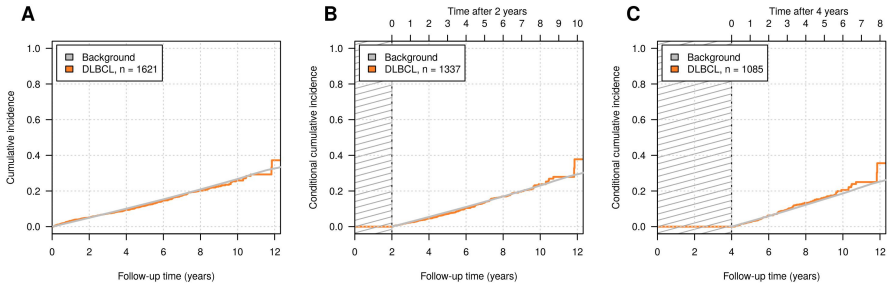

Figure S5: Cumulative incidences of death from a non-lymphoma related cause for DLBCL patients in 1st CR/CRu versus the general population A) for all patients, B) patients that achieve pEFS24, and C) patients achieving pEFS48.

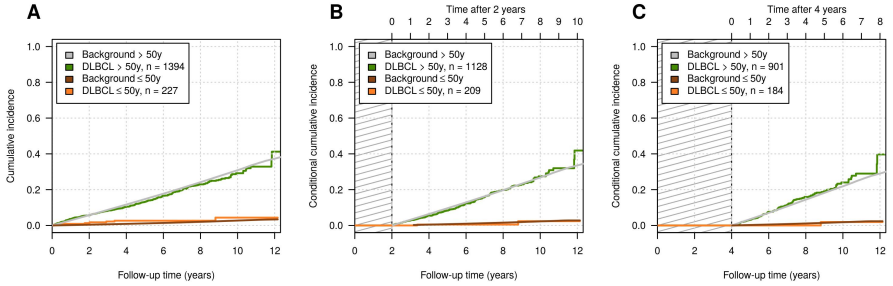

Figure S6: Age stratified cumulative incidences of death from a non-lymphoma related cause for DLBCL patients in 1st CR/CRu versus the general population A) for all patients, B) patients that achieve pEFS24, and C) patients achieving pEFS48.

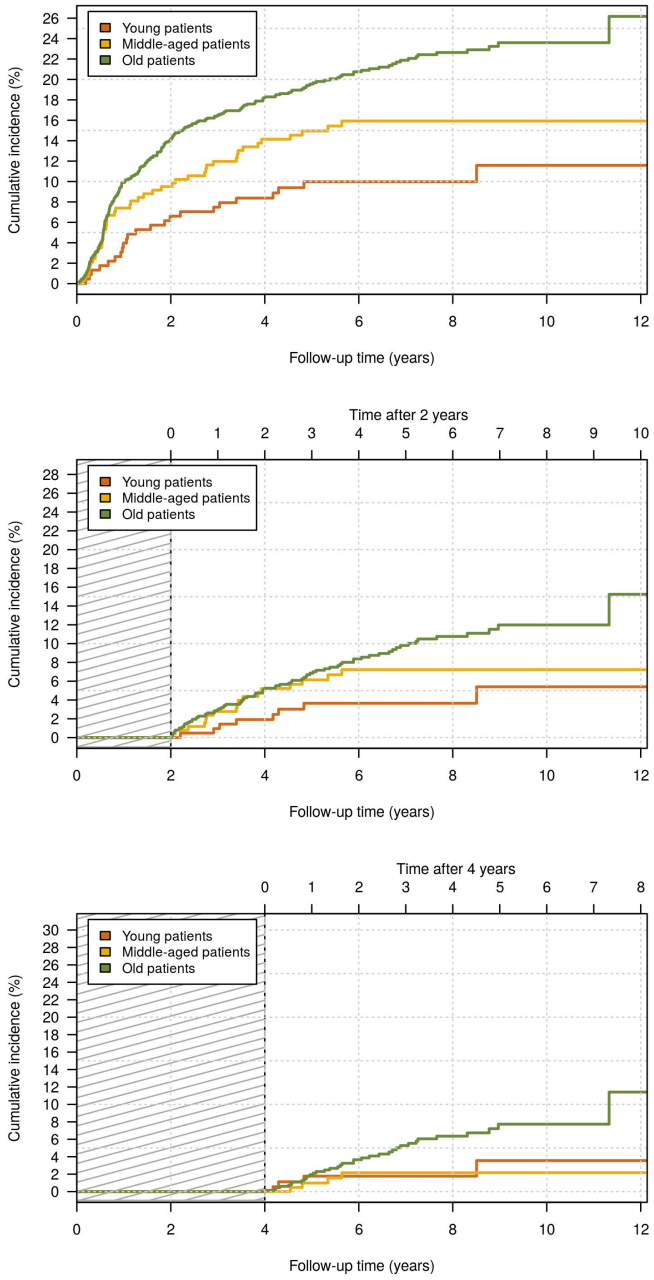

Figure S7: Cumulative incidences of relapse for DLBCL patients in 1st CR/CRu from end of treatment, pEFS24, and pEFS48 stratified a young (<50), a middle-aged (50-60), and an old group ($\geq$60).

Paper II

Estimating the loss of lifetime function using flexible parametric relative survival models

Lasse H. Jakobsen[1,2], Therese M.-L. Andersson[3], Jorne L. Biccler[1,2], Tarec C. El-Galaly[1,2], and Martin Bøgsted[1,2]

1. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; 2. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; 3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

Submitted to *BMC Medical Research Methodology*, 2017.

Description

In this paper, we evaluated the accuracy of various estimators of the loss of lifetime function. The loss of lifetime function measures the dynamic mortality difference between a given patient population and the general population. As the estimation of the loss of lifetime function requires extrapolation of the survival functions, parametric models are particularly useful for this purpose. However, even though a parametric model may fit the data well, it may extrapolate poorly. Therefore it has been proposed to utilize external data from registers or life tables to improve the extrapolations. Relative survival models have previously been used for estimating the loss of lifetime from diagnosis. The aim of this study was to assess the accuracy of the entire loss of lifetime function when employing relative survival models for the extrapolation.

Abstract

Background: Within cancer care, dynamic evaluations of the loss in expectation of life provides useful information to patients as well as physicians. The loss of lifetime function yields the conditional loss in expectation of life given survival up to a specific time point. Due to the inevitable censoring in time-to-event data loss of lifetime estimation requires extrapolation of both the patient and general population survival function. In this context, the accuracy of different extrapolation approaches has not previously been evaluated.
Methods: The loss of lifetime function was computed by decomposing the all-cause survival function using the relative and general population survival function. To allow extrapolation, the relative survival function was fitted using existing parametric relative survival models. In addition, we introduced a novel mixture cure model suited for extrapolation. The accuracy of the estimated loss of lifetime function using various extrapolation approaches was assessed in a simulation study and by data from the Danish Cancer Registry where complete follow-up was available. In addition, we illustrated the proposed methodology by analyzing recent data from the Danish Lymphoma Registry.
Results: No uniformly superior extrapolation method was found, but flexible parametric mixture cure models and flexible relative survival models seemed to be suitable in various scenarios.
Conclusion: Using extrapolation to estimate the loss of lifetime function requires careful consideration of the relative survival function outside the available follow-up period. We propose extensive sensitivity analyses when estimating the loss of lifetime function.

1 Introduction

Dynamic survival prediction is important in cancer care, where prognostic assessments are given numerous times during diagnosis, treatment, and post-treatment follow-up. A popular measure for characterizing the severity of a disease is the expected amount of lifetime lost due to the disease as compared to the general population. This measure is known as the *loss in expectation of life* and may be computed as the difference between the area under the general population and patient survival curves [1]. The loss in expectation of life has previously been used to characterize the disease burden within colon cancer and acute myeloid leukemia [2, 3]. The *loss of lifetime* function generalizes this measure by dynamically evaluating the loss in expectation of life, yielding the conditional number of years lost due to cancer given survival up to specific time points.

Due to the occurrence of censoring, computing the loss of lifetime func-

tion typically requires extrapolation of both the patient and general population survival function. Generally, extrapolation of survival functions estimated from censored time-to-event data is challenging since there is no way to evaluate the extrapolation accuracy and even a well-fitted model may extrapolate poorly.

An extensive literature exists on techniques for extrapolating survival functions. Jackson et al. reviewed methods for incorporating external data, such as register data or national life tables, to extrapolate survival functions [4]. The use of external data requires assumptions about how the survival in the present patient population and the external data differ and how this will continue beyond the follow-up. In particular, extrapolation through the relative survival function has been proposed for both grouped and individual-level data, which has demonstrated improved accuracy in comparison to models for the all-cause survival function [1, 5]. Andersson et al. examined the accuracy of the loss in expectation of life estimates calculated by three types of relative survival models [1]. However, none of these assessments were conducted for the entire loss of lifetime function.

In the following article, we compute the loss of lifetime function using previously introduced extrapolation approaches. In addition, a new flexible parametric relative survival model based on mixture cure models and spline-based proportional hazards models is introduced [6, 7]. We expand the study of Andersson et al. [1] by evaluating the accuracy of the entire loss of lifetime function based on various extrapolation approaches in a simulation study and in data from the Danish Cancer Registry where complete follow-up was available. In addition, as a clinically motivated example, the loss of lifetime function is computed for three lymphoma types using recent data from the Danish Lymphoma Registry.

2 Methods

2.1 Relative survival

The relative survival function is commonly used to describe the disease-specific (net) survival without requiring cause of death information. Given covariate vector $\boldsymbol{z}$, patient population (all-cause) survival function $S(t|\boldsymbol{z})$, and general population survival function, $S^*(t|\boldsymbol{z})$, the relative survival function is given by

$$R(t|\boldsymbol{z}) = \frac{S(t|\boldsymbol{z})}{S^*(t|\boldsymbol{z})}.$$

By using the relation between the hazard function and the survival function, the all cause hazard function corresponding to $S(t|\boldsymbol{z})$ can be written as

$$h(t|\boldsymbol{z}) = h^*(t|\boldsymbol{z}) + \lambda(t|\boldsymbol{z}),$$

where $h^*(t|\boldsymbol{z})$ is the general population hazard function and $\lambda(t|\boldsymbol{z})$ is termed the *excess hazard function* or *excess mortality*. Both $h^*(t|\boldsymbol{z})$ and $S^*(t|\boldsymbol{z})$ are usually computed from publicly available life tables matched on variables such as age, sex, and calendar year. The most popular way to include covariate effects is the proportional excess hazard model with a parametric specification of the baseline excess hazard [8, 9].

2.2 Parametric cure models

In survival analysis, cure models are used to provide useful information, particularly in cancers where the patient hazard function reaches the same level as the general population hazard function after some time [6, 10]. This corresponds to the relative survival reaching a plateau and the patients still alive after this time point are considered statistically cured. The main parameter of interest in cure models is the proportion of patients reaching statistical cure, also known as the cure proportion. Cure models are commonly divided into mixture and non-mixture cure models [6]. In mixture cure models, the patient population is considered a mixture of cure and uncured individuals. The relative survival is a mixture of a relative survival function for the cured and uncured patients, i.e.,

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})]S_u(t|\boldsymbol{z}), \tag{1}$$

where $\pi(\boldsymbol{z})$ is the, potentially covariate dependent, cure proportion and $S_u(t|\boldsymbol{z})$ is the relative survival function of the uncured patients. The cure proportion can be modelled through a link function, e.g., with a logistic, identity, or log-log link function [6]. The function $S_u(t|\boldsymbol{z})$ can conveniently be modelled by regular parametric survival models, such as a Weibull model, a log-normal model, or more flexible alternatives such as a Weibull-Weibull mixture model [11]. The model is estimated by maximum likelihood where the only external information needed is the general population hazard at the observed event times (see Lambert et al. [6] for the likelihood function).

Non-mixture cure models are of a less intuitive form:

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z})^{1-\widetilde{S}(t|\boldsymbol{z})},$$

where the function $\widetilde{S}(t|\boldsymbol{z})$ is a proper survival function which does not have an intuitive interpretation like $S_u(t|\boldsymbol{z})$. By rewriting the non-mixture cure model, it can be formulated as a mixture cure model, with $(\pi(\boldsymbol{z})^{1-\widetilde{S}(t|\boldsymbol{z})} -$

$\pi(z))/(1-\pi(z))$ as the relative survival function of the uncured patients [6]. Thus, estimation of the non-mixture cure model can be carried out similarly to that of mixture cure models.

2.3 Flexible parametric cure models

Royston and Parmar introduced a flexible parametric proportional hazards model by using restricted cubic splines to model the baseline hazard function (on the log cumulative hazard scale) [7]. This approach was applied to relative survival by Nelson et al. where the log-cumulative excess hazard was modelled by restricted cubic splines [9]. Including covariate effects, the relative survival by Nelson et al. is given by

$$\log(-\log(R(t|\boldsymbol{z})) = s_0(x;\boldsymbol{\gamma}_0) + \boldsymbol{z}^T\boldsymbol{\beta} + \sum_{i=1}^{p} s_i(x;\boldsymbol{\gamma}_i)z_i, \tag{2}$$

where $x = \log(t)$, p is the number of time-varying covariate effects, $s_0(x;\boldsymbol{\gamma}_0)$ is a baseline restricted cubic spline, $\boldsymbol{\beta}$ is a vector of regression coefficients, and $s_i(x;\boldsymbol{\gamma}_i)$ is a spline corresponding to the i^{th} covariate, providing a time varying coefficient. For the i^{th} spline, K_i knots, $k_{i1} < k_{i2} < .. < k_{iK_i}$, are selected on the log-time scale. The spline is then given as a linear combination of base functions defined through the chosen knots, i.e., $s_i(x;\boldsymbol{\gamma}_i) = \sum_{j=0}^{K_i-1} v_{ij}(x)\gamma_{ij}$, where $\boldsymbol{\gamma}_i$ are model parameters. The base functions are given by $v_{i0}(x) = 1$, $v_{i1}(x) = x$, and

$$v_{ij}(x) = (x-k_{ij})_+^3 - \lambda_{ij}(x-k_{i1})_+^3 - (1-\lambda_{ij})(x-k_{iK_i})_+^3, \tag{3}$$

for $j = 2, ..., K_i - 1$, where $\lambda_{ij} = \frac{k_{iK_i}-k_{ij}}{k_{iK_i}-k_{i1}}$ and $x_+ = \max(x, 0)$. Generally, the number and placement of the knots in the different spline functions do not need to be the same.

Andersson et al. used (2) to establish a flexible parametric cure model [12]. This model is formulated similarly to (2), but the basis functions of the splines are adjusted to ensure that the relative survival has zero slope after a preselected time point which is used as last knot in all spline functions, i.e., $k_K = k_{0K_0} = k_{1K_1} = \cdots = k_{pK_p}$. The cure proportion is then estimated by $R(k_K)$. Rewriting (2) we obtain

$$R(t|\boldsymbol{z}) = \exp\left(-\exp\left(\gamma_{00} + \boldsymbol{z}^T\boldsymbol{\beta}\right)\exp\left(\sum_{i=1}^{K_0-1} v_i(x)\gamma_i + \sum_{i=1}^{p} s_i(x;\boldsymbol{\gamma}_i)z_i\right)\right).$$

Hence, the model by Andersson et al. can be viewed as a non-mixture cure model where the cure proportion is modelled through the baseline spline

parameter, γ_{00}, and the fixed covariate effects, $\boldsymbol{z}^T\boldsymbol{\beta}$, while the remaining parameters are used to model $1 - \widetilde{S}(t)$ [12]. While this model provides a flexible framework for estimating the cure proportion in cancer studies, the assumption of statistical cure after the last knot is strong. Therefore, we introduce a new flexible parametric cure model which combines regular mixture cure models with flexible parametric survival models. The model is specified by (1) with

$$S_u(t|\boldsymbol{z}) = \exp\left(-\exp\left(s_0(x;\gamma_0) + \boldsymbol{z}^T\boldsymbol{\beta} + \sum_{i=1}^{p} s_i(x;\gamma_i)z_i\right)\right). \tag{4}$$

Similarly to the more simple cure models presented in Lambert et al. [6], $\pi(\boldsymbol{z})$ can be modelled by various link functions and the relative survival cannot fall below $\pi(\boldsymbol{z})$, thus ensuring statistical cure. The model is fitted by maximum likelihood using the likelihood of the mixture cure model. This cure model enables flexible modelling of the relative survival without the strong assumption of cure after the last knot while providing the more intuitive interpretation of a mixture cure model. Also, in this model, the modelling of the cure proportion becomes more clearly separated from the modelling of $S_u(t)$.

2.4 The loss of lifetime function

The conditional expected residual lifetime given survival until a time point t for patients with covariate vector $\boldsymbol{z}$ can be computed by $\int_t^\infty S(u|\boldsymbol{z})du / S(t|\boldsymbol{z})$. Based on this property, the loss of lifetime function can be computed by

$$L(t|\boldsymbol{z}) = \frac{\int_t^\infty S^*(u|\boldsymbol{z})du}{S^*(t|\boldsymbol{z})} - \frac{\int_t^\infty S(u|\boldsymbol{z})du}{S(t|\boldsymbol{z})}, \tag{5}$$

which is the difference in expected residual lifetime after time point t between the general population and the patients.

Extrapolation of both $S^*(\cdot|\boldsymbol{z})$ and $S(\cdot|\boldsymbol{z})$ is required to compute (5) since the survival distributions typically cannot be fully estimated due to censoring. Similarly to Andersson et al. [1], the extrapolation of the expected survival, $S^*(\cdot)$, can be accomplished by using the method of Ederer et al. [13] (Ederer I) and by making assumptions about the future population mortality rates. The latter can be carried out by using mortality rates from the last available time point or, if available, by using predicted future mortality rates.

For the patient survival, we apply the relative survival factorization, i.e., $S(t) = S^*(t)R(t)$, such that the extrapolation is based on the relative survival and the general population survival. Extrapolation of $R(\cdot)$ can be enabled by fitting a parametric relative survival model [1]. Since some cancer patient

groups experience statistical cure after some time while others experience persistent excess mortality, several assumptions on the relative survival can be applied. We consider three flexible parametric relative survival models, which mainly differ in tail:

1) the Nelson et al. [9] relative survival (NRS) model, which is linear on the log cumulative excess hazard scale after the last knot,

2) the Andersson et al. [12] relative survival (ARS) model, which is constant on the log cumulative excess hazard scale after the last knot and thereby incorporates statistical cure, and

3) the flexible mixture cure (FMC) model in (4), which incorporates statistical cure, but is not restricted to a constant log cumulative excess hazard after the last knot.

Due to their flexibility, the three models typically behave similarly within the first part of the follow-up, but may produce different survival trajectories beyond the available follow-up. In cure models, the relative survival cannot fall below π, and thus these models have a parameter to control the asymptote of the relative survival. Therefore, in cases where statistical cure occurs, cure models may improve extrapolation as compared to non-cure models. In cases where statistical cure does not occur, cure models may provide too optimistic extrapolations and hence may not be appropriate. However, in such cases, the introduced FMC model is expected to estimate π close to zero such that the fit is mainly based on the flexible survival function, $S_u(t)$. In the ARS model, letting $\pi = 0$, substantially affects the survival function since this forces $R(k_K) = 0$. Therefore, we consider the FMC model a hybrid between the NRS and ARS models.

2.5 Implementation

Initial values for the optimization procedure for the FMC model were chosen by first fitting a Weibull parametric cure model using only fixed covariate effects, i.e., fitting model (1) with a Weibull formulation of $S_u(t)$ and a logistic link for π. For the cure proportion, initial values were found by fitting a linear model with the predicted cure proportions scaled by the chosen link function as response and the cure proportion covariates as explanatory variables. For the relative survival of the uncured, initial values were found by fitting a linear model with the log-log transformed predicted relative survival of the uncured at the observed event times as response and the splines and covariates of $S_u(t)$ as explanatory variables. The splines do not guarantee that $S_u(t)$ is proper, but this can be obtained by adding a penalty for negative values of $h_u(t) = -d/dt \log S_u(t)$ similarly to Liu et al. [14]. In particular, the

term

$$\frac{\kappa}{2}\sum_{j=1}^{n} h_u(t_j|\boldsymbol{z}_j)^2\mathbf{1}[h_u(t_j|\boldsymbol{z}_j) < 0]$$

is subtracted from the log-likelihood, where t_j and $\boldsymbol{z}_j$ are the observed follow-up time and covariate vector of patient j. Initially, κ is 1, but doubles until no negative values of h_u are obtained. Orthogonalization of the base functions of the restricted cubic splines has previously been recommended due to the potential correlation between their base functions [15]. We employed a QR-decomposition approach to carry out the orthogonalization.

Choosing the number and location of the knots is a key issue in spline-based models. Similarly to Royston and Parmar, the knots of the FMC model were selected according to the quantiles of the uncensored event times [7]. In a simulation study, Rutherford et al. [15] concluded that complex hazard shapes can adequately be captured by the spline-based model of Royston and Parmar [7] provided that a sufficient number of knots is selected. In particular, the survival model was rather insensitive to the number of knots and it was argued that the results should also be valid in relative survival and cure models.

All analyses were performed in the statistical programming language `R`. For the purpose of this article, the NRS and ARS models were fitted using the package `rstpm2` [16]. Functions for estimating the presented FMC model and computing the loss of lifetime function were assembled in the `R`-package `cuRe` (see `https://github.com/LasseHjort/cuRe`). The package also enables estimation of the expected residual lifetime, restricted expected residual lifetime, and restricted loss of lifetime using any of the models considered here. The integrals of the loss of lifetime function are computed numerically by Gauss-Legendre quadrature, while the point-wise variance of the loss of lifetime function is estimated using the delta method and numerical differentiation.

3 Results

3.1 Simulation study

Simulation design

We simulated data according to selected relative survival scenarios by using the independence assumption of the relative survival and general population survival times. Similarly to Rutherford et al. [17], we used the following simulation scheme:

1. Draw a general population survival time T_E from S^*.
2. Draw a relative survival time, T_R from R.

3. Draw a censoring time T_C from C.

4. The observed follow-up time is given by $T = \min(T_R, T_E, T_C)$ and the event indicator is $\delta = \mathbf{1}[\min(T_R, T_E) \leq T_C]$.

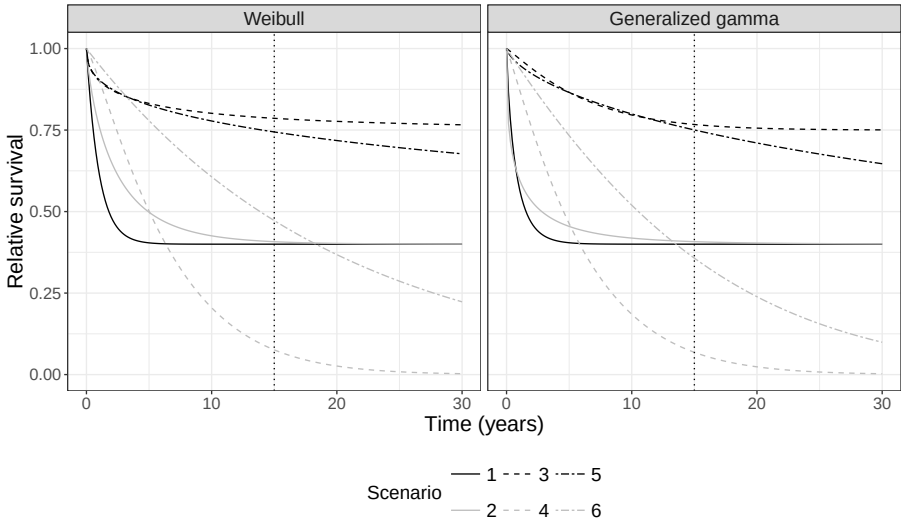

Figure 1: Relative survival functions used to simulate net survival times. In scenario 1, 2, and 3, follow-up times were simulated from a Weibull (generalized gamma) cure model with varying cure proportions, and in scenario 4, 5, and 6, the follow-up times were simulated from a Weibull (generalized gamma) relative survival model.

The general population survival distribution, $S^*(t)$, was chosen corresponding to 50-, 60-, and 70-year-old female patients diagnosed in 1980. For this purpose, we used the Danish general population mortality published by the Human Mortality Database [18]. The relative survival, $R(t)$, was determined by a Weibull mixture cure model according to the scenarios in Figure 1. In scenario 1, 2, and 3, the cure proportion was 40%, 40%, and 75% and cure occurred within the available follow-up, just outside the follow-up, and many years after the last follow-up time, respectively. In scenarios 4, 5, and 6, the cure proportion was zero and therefore the relative survival function corresponded to a regular Weibull model. Scenario 5 was similar to 3 within the follow-up, but differed beyond the follow-up. In scenario 4, most patients died within the follow-up and scenario 6 was included as an example of a clear absence of cure within the follow-up. In scenarios where $R(t)$ had a cure proportion, follow-up times were set to ∞, if there was no solution to the equation $R(t) = U$, where U is uniformly distributed between 0 and 1. To examine the extrapolation performances under different trajectories,

we repeated the simulations after replacing the Weibull distribution with the generalized gamma distribution.

To mimic typical register data, the censoring times were simulated from a uniform distribution, C, between 0 and 15 years. Using S^* and R, the true loss of lifetime function was obtained by inserting into (5). All scenarios were simulated 500 times with a sample size of 1000.

Model	Model	Nr. knots	Knot locations
A	NRS	6	0%, 20%, 40%, 60%, 80%, and 100% quantiles of the uncensored event times.
B	ARS	7	0%, 20%, 40%, 60%, 80%, and 100% quantiles of the uncensored event times with an additional knot placed at 10 years.
C	ARS	7	0%, 20%, 40%, 60%, and 80% quantiles of the uncensored event times. The last knot is placed at 80 years and an additional knot is placed at 10 years.
D	FMC	5	0%, 25%, 50%, 75%, and 100% quantiles of the uncensored event times.
E	FMC	5	First uncensored event time, 0.5, 1, 2, and 5 years.

Table 1: Specification of models used to estimate the loss of lifetime function. NRS: Nelson et al. [9] relative survival model, ARS: Andersson et al. [12] relative survival model, FMC: Flexible mixture cure model.

For estimation of the loss of lifetime function, we considered five models (Table 1). In order to obtain the same number of parameters in each model, an additional knot was required for models B and C, which was placed late in the follow-up, while for model D and E the number of knots was decreased by one since these contain an explicit parameter for the cure proportion. Extrapolation using model A and B were considered by Andersson et al. [1]. We considered a special case of the latter model, where the last knot was placed beyond the available follow-up. We also considered two instances of the FMC model, i.e., D with conventional knot placement and E where the knots were placed in the beginning of the follow-up.

For each model, the loss of lifetime function was computed and the bias was measured by $D(t) = \widehat{L}(t) - L(t)$. The integral, $\int_0^{15} |D(u)|du$, was used to measure the bias of the loss of lifetime estimate during the entire follow-up period.

Simulation results

In scenarios with statistical cure (scenario 1, 2 and 3), all models had comparable performances at time zero for 50-year-old patients (Figure 2). In scenarios 1 and 3, the bias was fairly low for all models at all time points, but in scenario 2, the non-cure model, A, yielded increasingly upward biased estimates. In scenarios without statistical cure (scenario 4, 5, and 6), the diversity

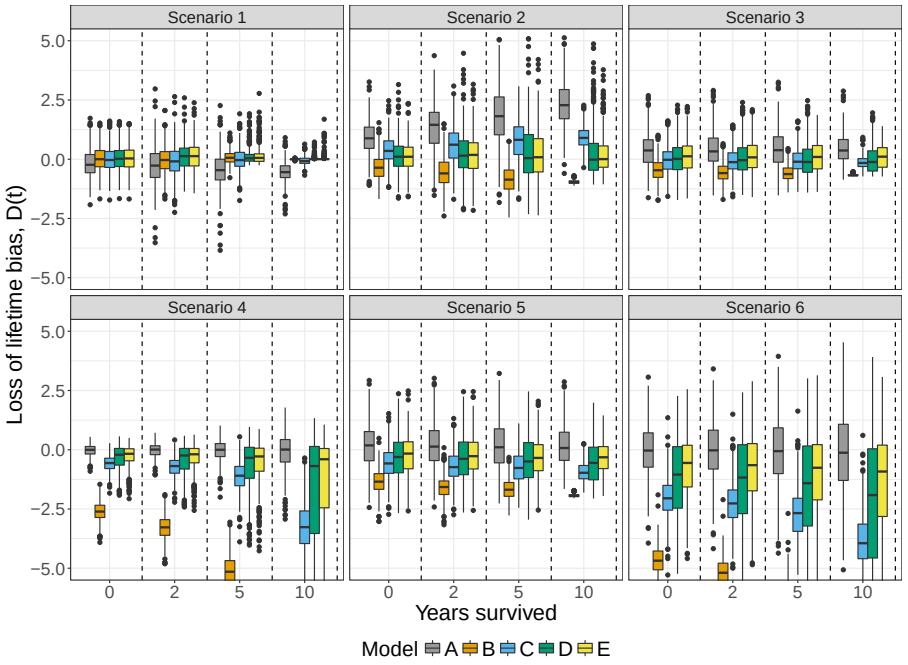

Figure 2: Loss of lifetime bias, $D(t)$, of the models in Table 1 at time 0, 2, 5, and 10 years in 50-year-old patients following six Weibull relative survival scenarios.

between the models became larger. In these scenarios, the non-mixture cure models, B and C, underestimated the loss of lifetime, most markedly seen in model B which assumes cure within the follow-up period.

Generally, the FMC models, D and E, showed good performance both in scenarios with statistical cure occurring within and beyond the available follow-up. In scenarios where statistical cure did not occur, the performance of the FMC models was comparable to model A, but the biases were more dispersed for later time point, especially in scenario 4 and 6. At ten years, the biases of model E were slightly less dispersed compared to model D.

Table 2 shows the integrated loss of lifetime biases for 50-, 60-, and 70-year-old patients. In general, the integrated overall biases were consistent with Figure 2 where model A, D, and E performed well across the six scenarios. In comparison to model D, model E was largely producing less biased estimates, while only being slightly worse than model A in scenario 4 and 6. Generally, the loss of lifetime bias decreased with increasing age and hence reduced the differences between the models. Despite the bias reduction in 70-year-olds, model B still resulted in a relatively large bias in scenario 4 and 6. The results were similar in the generalized gamma case (Figure 3 and Table 3). In particular, the models A and E showed satisfactory performance in all

Age	Scenario	pi	Model A	Model B	Model C	Model D	Model E
50	1	0.40	8.6(0.9-39.2)	2.4(0.5-10.7)	4.1(0.9-19.5)	2.4(0.2-18.5)	2.6(0.3-28.6)
	2	0.40	28.9(5.5-66.9)	11.7(6.0-23.1)	12.7(3.7-35.3)	11.1(0.6-68.0)	9.3(0.7-52.1)
	3	0.00	8.9(0.7-42.7)	8.9(4.3-15.6)	4.8(0.9-16.4)	7.5(0.5-31.5)	7.2(0.2-23.9)
	4	0.75	6.5(0.3-34.5)	144.8(123.4-171.2)	37.2(8.9-82.1)	24.9(0.3-111.8)	18.1(0.2-104.6)
	5	0.00	11.0(0.2-43.3)	25.4(14.8-35.4)	13.0(3.2-31.0)	12.8(0.4-33.8)	9.6(0.3-30.9)
	6	0.00	18.1(0.5-65.4)	106.6(71.6-127.7)	49.2(7.9-92.4)	36.2(0.6-103.7)	23.9(0.2-89.4)
60	1	0.40	6.0(1.3-19.4)	1.9(0.4-8.0)	3.1(0.6-13.5)	2.1(0.2-14.7)	2.4(0.2-17.6)
	2	0.40	14.9(2.6-45.4)	6.4(3.4-14.9)	7.7(2.0-26.2)	7.7(0.6-40.2)	6.6(0.2-42.5)
	3	0.00	7.2(0.4-39.6)	4.2(1.9-10.0)	4.0(0.4-22.7)	5.4(0.3-28.2)	4.7(0.3-19.6)
	4	0.75	5.7(0.3-24.1)	79.5(64.6-93.2)	21.0(6.1-44.4)	14.2(0.3-62.4)	10.2(0.1-49.8)
	5	0.00	7.5(0.3-33.4)	10.7(5.1-18.1)	5.6(1.5-18.6)	7.2(0.5-26.0)	5.0(0.2-17.9)
	6	0.00	10.9(0.6-37.3)	48.2(36.4-61.2)	18.5(4.1-42.5)	16.8(1.2-50.9)	11.2(0.3-45.3)
70	1	0.40	3.6(0.9-12.4)	1.5(0.2-4.8)	2.2(0.4-7.3)	1.7(0.1-8.8)	2.0(0.1-12.9)
	2	0.40	6.2(1.2-20.8)	3.4(1.5-8.8)	4.0(0.9-14.2)	4.7(0.3-19.4)	4.3(0.3-19.2)
	3	0.00	4.9(0.3-20.6)	2.4(0.8-7.0)	3.3(0.3-12.9)	3.6(0.2-19.0)	2.8(0.2-11.2)
	4	0.75	4.3(0.3-16.1)	34.9(26.9-44.4)	9.6(3.5-23.3)	7.3(0.2-31.3)	5.7(0.1-27.5)
	5	0.00	5.3(0.4-25.2)	3.9(1.7-8.9)	3.5(0.6-14.8)	4.3(0.2-18.5)	2.9(0.1-10.5)
	6	0.00	6.0(0.3-21.7)	16.5(9.9-23.8)	6.2(1.8-16.6)	6.9(0.2-23.7)	5.1(0.2-17.4)

Table 2: The integrated loss of lifetime bias in the Weibull scenario, computed by integrating $|D(t)|$ from 0 to 15 years. The loss of lifetime was computed for 50-, 60-, and 70-year-old patients. The mean and range from the 500 simulations are provided.

scenarios while model D was more biased in scenario 6. Also in the generalized gamma case, model E had slightly lower integrated bias compared to model D in scenario 4, 5, and 6.

Age	Scenario	pi	Model A	Model B	Model C	Model D	Model E
50	1	0.40	6.8(0.6-30.1)	2.4(0.4-11.3)	4.0(0.5-21.0)	2.6(0.3-29.8)	3.1(0.3-30.5)
	2	0.40	23.1(5.0-48.1)	10.2(5.7-18.3)	7.1(1.8-22.8)	8.5(0.6-39.3)	10.3(0.7-43.8)
	3	0.00	17.4(1.5-74.8)	10.2(4.2-19.8)	7.6(1.0-31.0)	11.2(0.8-63.1)	10.0(0.3-56.2)
	4	0.75	6.7(0.2-32.6)	146.5(121.6-166.8)	39.2(10.4-75.0)	22.8(0.2-123.1)	15.4(0.2-95.0)
	5	0.00	14.8(0.6-88.3)	35.9(22.7-45.6)	18.0(3.2-39.4)	18.4(0.6-77.5)	13.7(0.3-40.3)
	6	0.00	16.1(0.6-72.9)	130.5(108.5-153.0)	56.1(9.8-99.7)	36.4(0.5-117.8)	21.3(0.8-100.1)
60	1	0.40	4.7(0.9-16.5)	1.9(0.2-7.5)	3.0(0.3-12.1)	2.2(0.2-14.3)	2.8(0.1-21.5)
	2	0.40	12.0(2.7-35.4)	5.5(3.1-10.9)	4.6(1.0-19.4)	5.7(0.4-32.1)	7.2(0.5-35.0)
	3	0.00	11.1(1.0-53.0)	5.2(2.2-11.4)	6.1(0.7-31.7)	7.4(0.8-37.6)	7.5(0.4-31.0)
	4	0.75	5.8(0.2-22.2)	80.3(65.1-96.0)	21.8(7.3-42.4)	13.9(0.3-66.4)	9.5(0.2-49.4)
	5	0.00	10.2(0.7-55.2)	14.8(6.9-22.4)	7.4(1.6-29.4)	9.6(0.3-33.6)	6.8(0.2-22.0)
	6	0.00	9.8(0.3-37.7)	62.2(44.7-77.0)	24.1(5.9-51.5)	17.1(0.7-64.5)	11.4(0.6-54.9)
70	1	0.40	3.1(0.6-11.0)	1.4(0.2-4.8)	2.2(0.2-7.4)	1.8(0.2-8.8)	2.3(0.1-12.7)
	2	0.40	5.5(1.0-16.7)	2.6(1.4-6.0)	2.9(0.4-11.1)	3.2(0.2-13.6)	4.2(0.3-14.6)
	3	0.00	6.4(0.4-30.7)	2.8(1.0-8.8)	4.1(0.3-20.0)	4.6(0.4-24.6)	4.5(0.4-17.6)
	4	0.75	4.2(0.3-14.9)	35.4(26.7-43.6)	10.2(3.9-23.9)	7.4(0.3-30.9)	5.4(0.2-24.0)
	5	0.00	6.1(0.2-26.8)	4.9(2.3-10.5)	3.9(0.6-17.2)	4.8(0.4-26.0)	3.6(0.1-13.1)
	6	0.00	6.1(0.4-23.2)	22.4(14.5-30.4)	8.0(2.4-22.7)	7.2(0.7-29.7)	5.2(0.4-25.1)

Table 3: The integrated loss of lifetime bias in the generalized gamma scenario, computed by integrating $|D(t)|$ from 0 to 15 years. The loss of lifetime was simulated for 50-, 60-, and 70-year-old patients. The mean and range from the 500 simulations are provided.

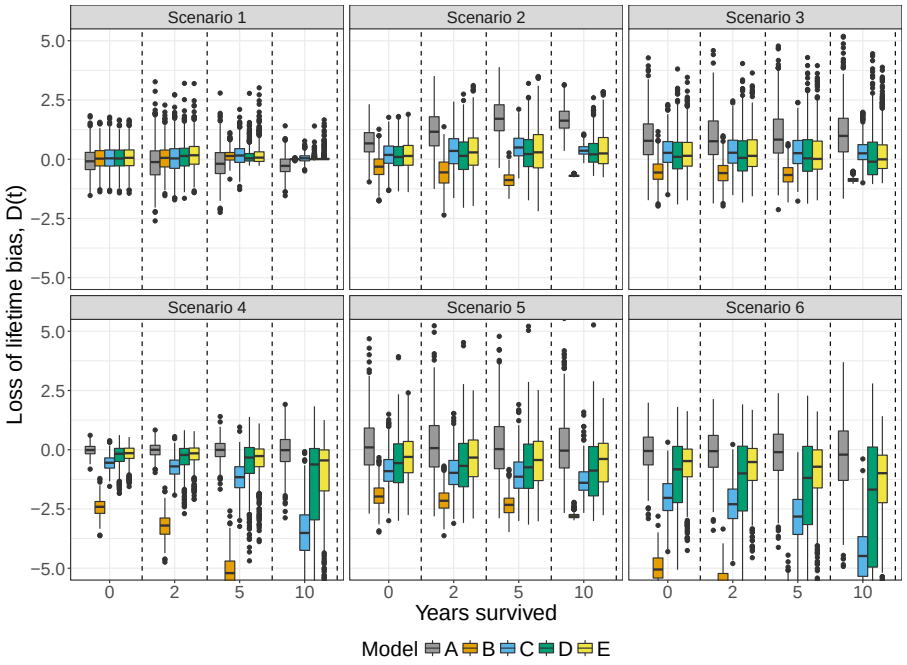

Figure 3: Loss of lifetime bias, $D(t)$, of the models in Table 1 at time 0, 2, 5, and 10 years in 50-year-old patients following six generalized gamma relative survival scenarios.

3.2 Analysis of Danish cancer registry data

Data description

To investigate the performance of the models in Table 1 in cancer survival data, we analyzed data from the Danish Cancer Registry [19] on patients with colon cancer ($n = 4,558$), breast cancer ($n = 21,731$), bladder cancer ($n = 11,738$) and malignant melanoma ($n = 2404$). To achieve (almost) complete follow-up, we included patients diagnosed in the period 1960-1975, who were older than 50 years at diagnosis. The diseases were chosen based on the same considerations as in Andersson et al. [1], i.e., colon cancer typically displays statistical cure, bladder cancer a constant excess hazard, melanoma a rather high survival rate, and breast cancer is seen in both young and old patients. Patients were followed until the end of 2016, where alive patients were censored and follow-up was measured from diagnosis until death or censoring. For the purpose of investigating the extrapolation performance, we restricted the follow-up to 16 years by censoring patients alive in January 1976 and divided patients into age groups; 50-59, 60-69, 70-79, 80+. The true loss of lifetime was calculated by inserting the Kaplan-Meier estimate into (5), and the bias was computed by $D(t)$. For both the true and estimated loss

of lifetime, the upper limit of the integrals in (5) was set to 40 years at which time the true survival was close to zero.

Results

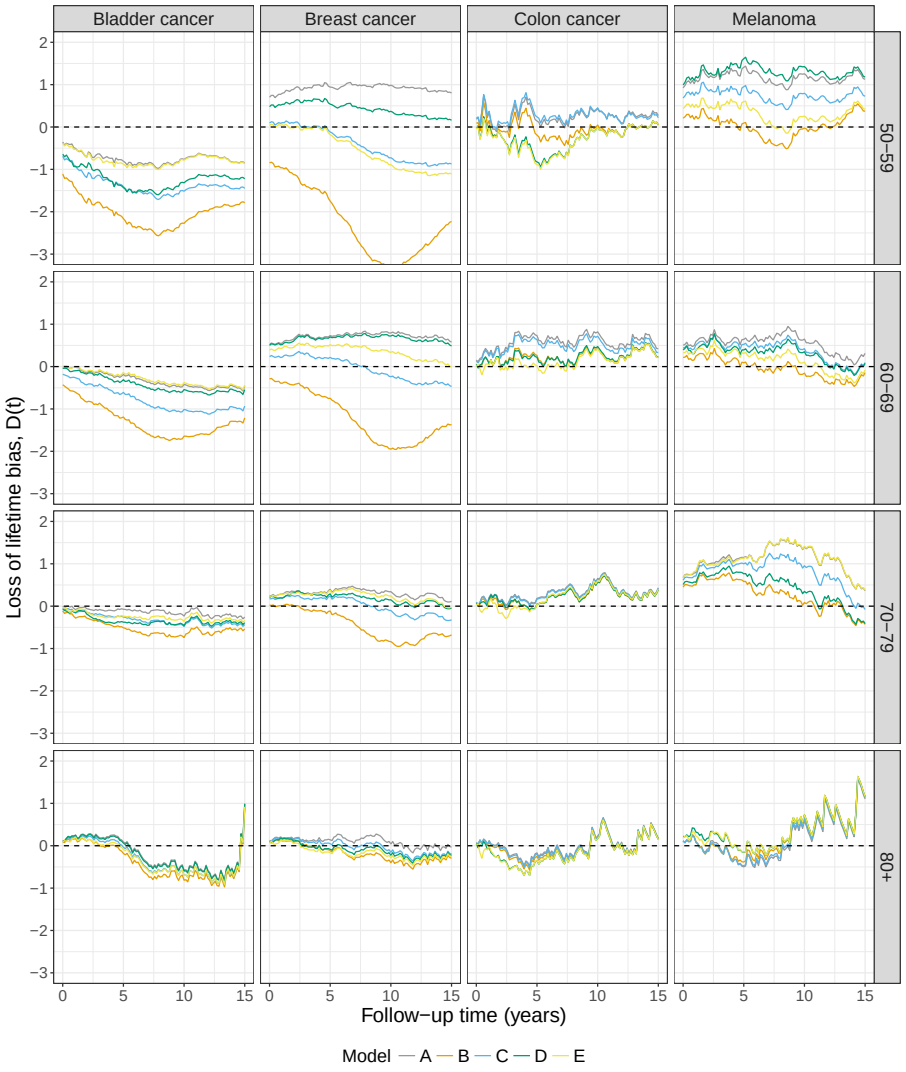

Figure 4: Time-varying loss of lifetime bias using the models in Table 1 for extrapolation in bladder cancer, breast cancer, colon cancer, and melanoma patients registered in the Danish Cancer Registry.

Figure 4 shows the bias function for each disease and each age group

using the five models in Table 1. The corresponding survival curves can be found in Figure S1-S4. The models displayed varying performance across the cancer types and age groups, but biases were commonly decreasing with increasing age. The extrapolation performance within bladder cancer was rather poor; in the age groups 50-59 and 60-69, the models consistently underestimated the loss of lifetime function with model B being the worst. Also for breast cancer, model B underestimated the loss of lifetime function while model C, which assumes statistical cure beyond the follow-up, provided improved results. In breast cancer, the two FMC models resulted in rather different loss of lifetime biases, but the bias was not consistently better in one model. For colon cancer where statistical cure is typically displayed, all models performed fairly well in all age groups and among the melanoma patients, model B had the best performance.

Overall, no model was consistently superior to the others, but in scenarios of statistical cure, there was a slight advantage of using cure models. However, in scenarios without statistical cure, models B and C were substantially biased.

3.3 Analysis of Danish lymphoma register data

Data description

To illustrate a potential clinical application of the proposed extrapolation techniques, we analyzed patient data from the Danish Lymphoma Registry, which covers 94.9% of all lymphoma cases in Denmark [20]. We included adult patients ($\geq$18 years of age) diagnosed with diffuse large B-cell lymphoma (DLBCL, n = 6639), follicular lymphoma (FL, n = 3204), or mantle cell lymphoma (ML, n = 980) in the period from 2000 to 2016. The follow-up period was terminated in June 2017 and the follow-up time was measured from time of diagnostic biopsy to death or censoring.

Population-based loss of lifetime

For each disease, three models were fitted, namely the NRS model with 6 knots, the ARS model with 7 knots, and the FMC model with 5 knots (corresponding to model A, B, and D in Table 1), resulting in the same number of parameters. Figure S5 displays the relative survival of each disease and disease-specific summary measures are shown in Table 4.

The estimated loss of lifetime function based on the three models is shown for each disease in Figure 5. DLBCL and ML patients had a high loss of lifetime at diagnosis with a rapid decrease, while FL patients displayed a fairly low initial loss of lifetime with a slow improvement.

Clearly, the three models, despite being similar in the beginning of the

	Model	DLBCL	FL	ML
Median age (range)		68(18-101)	63(18-97)	70(28-99)
5-year RS (95% CI)	NRS	0.66(0.65-0.68)	0.9(0.88-0.91)	0.61(0.57-0.65)
	ARS	0.66(0.65-0.67)	0.9(0.88-0.91)	0.61(0.57-0.65)
	FMC	0.66(0.64-0.67)	0.9(0.88-0.91)	0.61(0.58-0.65)
Loss of lifetime (95% CI)	NRS	7.43(7.06-7.80)	4.58(3.73-5.42)	7.66(6.86-8.46)
	ARS	6.70(6.42-6.98)	3.57(3.13-4.02)	6.92(6.26-7.59)
	FMC	7.21(6.86-7.55)	3.97(3.24-4.70)	7.74(6.95-8.53)

Table 4: Median age, 5-year relative survival (RS), and loss of lifetime estimates at time zero in Danish diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and mantle cell lymphoma (ML) patients.

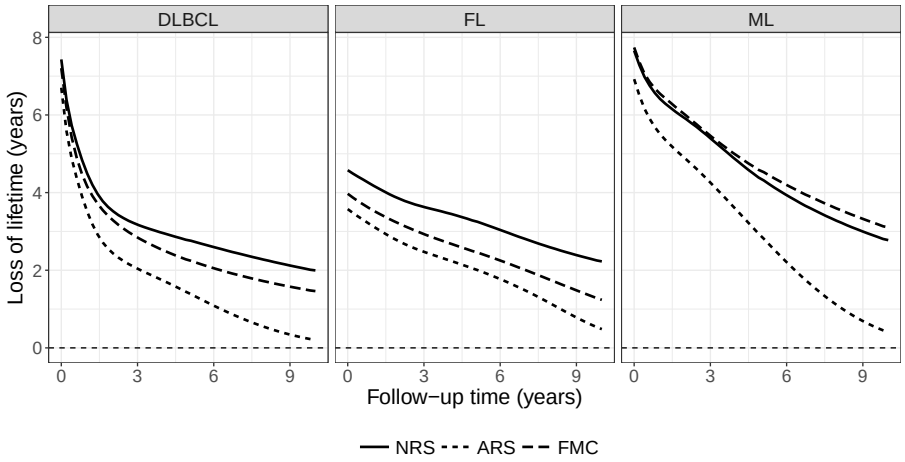

Figure 5: The loss of lifetime function in Danish diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and mantle cell lymphoma (ML) patients.

follow-up, produce rather different conditional loss of lifetime estimates. At time zero, the maximal difference between the models is seen to be around 1 year for FL, for which the assumption of statistical cure is typically not reasonable. The model differences increased as time progressed, with the largest difference seen in ML patients. The presented FMC model yielded a compromise between the NRS and ARS models which was seen by an intermediate loss of lifetime function. However, for the ML patients this model resembled the NRS model and even provided slightly higher loss of lifetime estimates.

Age-dependent loss of lifetime

The patient age at diagnosis plays a crucial role for the individual expected residual lifetimes and thus also the loss of lifetime function. For the NRS

model, a time-dependent age effect was specified, i.e.,

$$R(t|a) = \exp\left(-\exp\left(s_0(x) + s_a(a)s_1(x)\right)\right),$$

where a is the patient age at diagnosis, $s_a(a)$ is a spline-based age effect and $s_1(x)$ is the corresponding time-effect. For the FMC model, (4), the same model was used for $S_u(t|\boldsymbol{z})$ and for $\pi(\boldsymbol{z})$ an age dependent spline-based logistic model,

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + s_a(a).$$

was chosen. Since none of the diseases showed a clear statistical cure trajectory, we did not consider the ARS model here. The number and location of the knots for the baseline spline function, $s_0(x)$, remained unchanged from Section 3.3. For $s_a(a)$, 4 knots placed at the 0%, 33%, 66%, and 100% quantiles of the patient ages were selected and the intercept was removed since this is already modelled by the baseline splines and β_0. For $s_1(x)$, the number of knots was chosen to be 3 and 2 for the NRS model and the FMC model, respectively, yielding the same total number of parameters.

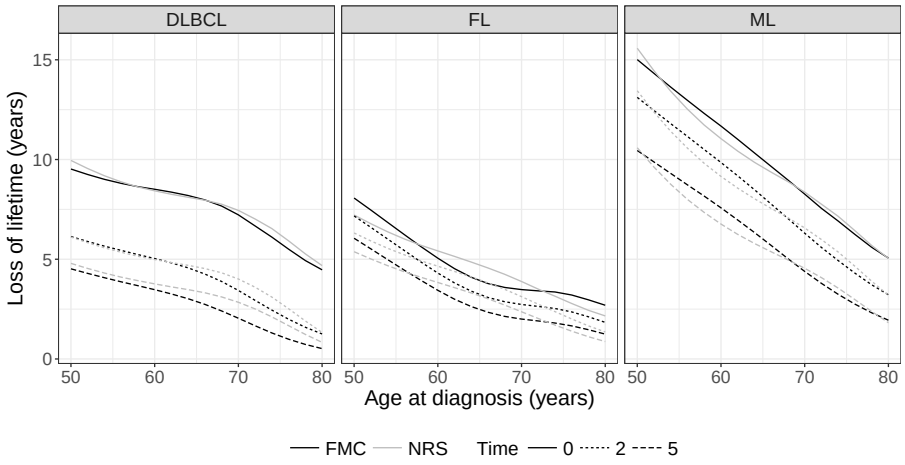

Figure 6: The loss of lifetime conditional on 0, 2, and 5 years of survival for female diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and mantle cell lymphoma (ML) patients diagnosed in 2010 at varying ages.

The loss of lifetime conditional on 0, 2, and 5 years of survival for female patients diagnosed in 2010 is shown in Figure 6 for varying patient ages. In all three cancer types, the loss of lifetime decreased with increasing diagnostic age.

For DLBCL and ML, the two models seem to be in agreement across patient age. However, the agreement between the two models for 60-70 year old

FL patients was poor, likely due to the different model assumptions. Notably, the models did not differ substantially for young patients even though more extrapolation is needed to compute the loss of lifetime for these patients.

4 Discussion

In (4) we introduced a novel model, which incorporates statistical cure by combining regular mixture cure models with spline-based survival models. This model was compared to the NRS model, which has a linear effect in the spline function after the last knot and the ARS model, which is constant after the last knot and thereby incorporates statistical cure. The simulations demonstrated a consistently good performance of the NRS model and the FMC model. The analysis of data from the Danish Cancer Registry did not show consistently satisfactory performance of any model, but in general assuming statistical cure at the end of the follow-up can lead to substantial biases in cases where this assumption is violated, while yielding good estimates when cure is reached.

The present article expanded on the study of Andersson et al. [1] by evaluating the accuracy of the entire loss of lifetime function using three extrapolation approaches. While the loss of lifetime estimates at time zero in Figure 4 seemed to be in agreement with the results reported by Andersson et al. where only 10 years of follow-up were used, the biases were not constant over time.

The general population survival probabilities for young patients are high and precise extrapolation of the relative survival is required to avoid a biased loss of lifetime function for these patients. Confirming this, we observed a higher bias among young patients which should be kept in mind when reporting loss of lifetime results. With longer follow-up and higher age, the bias will decrease and in future studies it would be of interest to estimate for a fixed age distribution, the amount of follow-up needed to provide sufficiently unbiased loss of lifetime estimates.

The loss of lifetime measure provides a crude measure of the cancer-related mortality. In net measures, such as relative survival, it is often seen that elderly patients have an increased mortality since deaths from other causes are not taken into account. For young patients, even a small excess mortality may have a have a large impact on the loss of lifetime function as the expected lifetime without cancer is long. Therefore, it is often seen that young patients have a higher loss of lifetime than elderly patients.

An alternative to the unrestricted loss of lifetime, where extrapolation is avoided, can be obtained by replacing the upper limit of the integrals in (5) by a fixed time point τ. In this setting, pseudo-values and flexible parametric survival models have previously been recommended for computing the mean

survival time [21] and could also be used for estimating the loss of lifetime function. Using the three models to estimate the restricted loss of lifetime would likely yield fairly similar estimates due to the model similarities in the first part of the follow-up (Figure S5). However, interpretation of the restricted loss of lifetime is not straightforward and the measure does not capture the full disease burden.

5 Conclusion

Since there is no way of assessing the performance of extrapolations applied to data with limited follow-up, the inconsistencies between the simulation results and the full follow-up data analysis emphasize the need for sensitivity analyses. We therefore recommend that extensive sensitivity analyses are performed both with respect to the assumptions of the relative survival model as well as the number and location of the knots of the splines as recommended previously [9, 12].

Ethics approval and consent to participate
The study was approved by the Danish Data Protection Agency (2008-58-0028).

Consent to publish
Not applicable.

Availability of data and materials
Data used to generate the findings of the study were obtained from the Danish Clinical Registries (Danish Lymphoma Registry) and the Danish Cancer Registry after approval of our study plan by both registries and the Danish Data Protection Agency. The registries contain patient identifiable information and therefore sharing of these data is not allowed per the terms of the agreement with the registries. However, from the registries, access to the data is granted on a case to case basis after submission and approval of an appropriate study plan and reasonable data request. Data from the Danish Cancer Registry can be applied for at
https://sundhedsdatastyrelsen.dk/da/forskerservice,
and data from the Danish Lymphoma Registry can be applied for at
http://www.rkkp.dk/forskning.

Competing interests
The authors declare that they have no competing interests.

Funding
No funding was received for the study.

Authors' contributions
The idea was conceived by LHJ, MB, and TMLA. LHJ performed all analyses and wrote the first draft of the manuscript. All authors interpreted data and discussed the methodology, as well as read and approved the final manuscript.

Acknowledgements
Not applicable.

References

[1] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, M. Lambe, and P. C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, 2013.

[2] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, A. Sjövall, M. Lambe, et al. The loss in expectation of life after colon cancer: a population-based study. *BMC Cancer*, 15(1):412, 2015.

[3] H. Bower, T. M.-L. Andersson, M. Björkholm, P. W. Dickman, P. C. Lambert, et al. Continued improvement in survival of acute myeloid leukemia patients: an application of the loss in expectation of life. *Blood Cancer Journal*, 6(2):e390, 2016.

[4] C. Jackson, J. Stevens, S. Ren, N. Latimer, L. Bojke, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Medical Decision Making*, 37(4):377–390, 2017.

[5] M. Hakama and T. Hakulinen. Estimating the expectation of life in cancer survival studies with incomplete follow-up information. *Journal of Chronic Diseases*, 30(9):585–597, 1977.

[6] P. C. Lambert, J. R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, 2007.

[7] P. Royston and M. K. B. Parmar. 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(15):2175–2197, 2002.

[8] P. W. Dickman, A. Sloggett, M. Hills, and T. Hakulinen. Regression models for relative survival. *Statistics in Medicine*, 23(1):51–64, 2004.

[9] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, 2007.

[10] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–454, 1999.

[11] P. C. Lambert, P. W. Dickman, C. L. Weston, and J. R. Thompson. Estimating the cure fraction in population-based cancer studies by using finite mixture models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1):35–55, 2010.

[12] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, and P. C. Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, 2011.

[13] F. Ederer, L. M. Axtell, and S. J. Cutler. The relative survival rate: a statistical methodology. *National Cancer Institute monograph*, 6:101–121, 1961.

[14] X.-R. Liu, Y. Pawitan, and M. Clements. Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5):1531–1546, 2018.

[15] M. J. Rutherford, M. J. Crowther, and P. C. Lambert. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, 85(4):777–793, 2015.

[16] M. Clements and X.-R. Liu. *rstpm2: Generalized Survival Models*. R package version 1.4.3. 2018. URL: `http://github.com/mclements/rstpm2`.

[17] M. J. Rutherford, P. W. Dickman, and P. C. Lambert. Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology*, 36(1):16–21, 2012.

[18] Human mortality database. URL: `www.mortality.org`. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Accessed 15 September 2017.

[19] M. L. Gjerstorff. The Danish Cancer Registry. *Scandinavian Journal of Public Health*, 39(Suppl 7):42–45, 2011.

[20] B. Arboe, T. C. El-Galaly, M. R. Clausen, P. S. Munksgaard, D. Stoltenberg, et al. The Danish National Lymphoma Registry: coverage and data quality. *PloS One*, 11(6):e0157999, 2016.

[21] P. Royston and M. K. B. Parmar. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30(19):2409–2421, 2011.

Supplementary

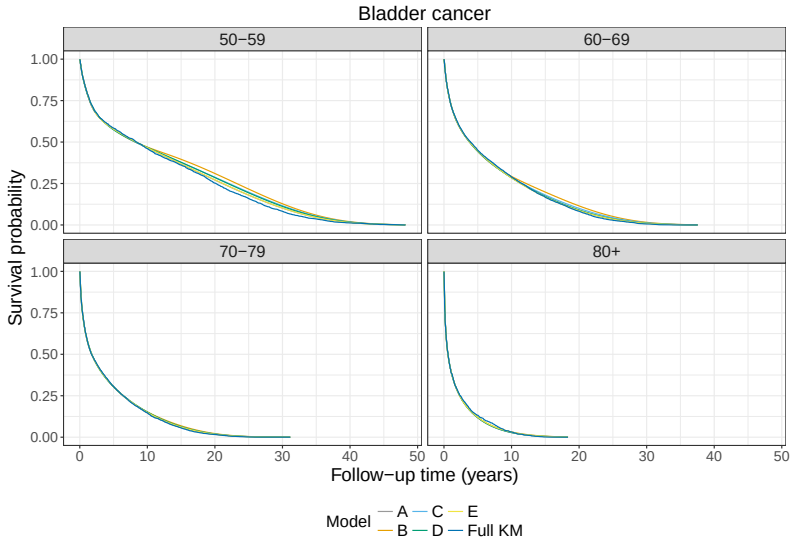

Figure S1: The extrapolated survival function of bladder cancer patients based on 5 relative survival models.

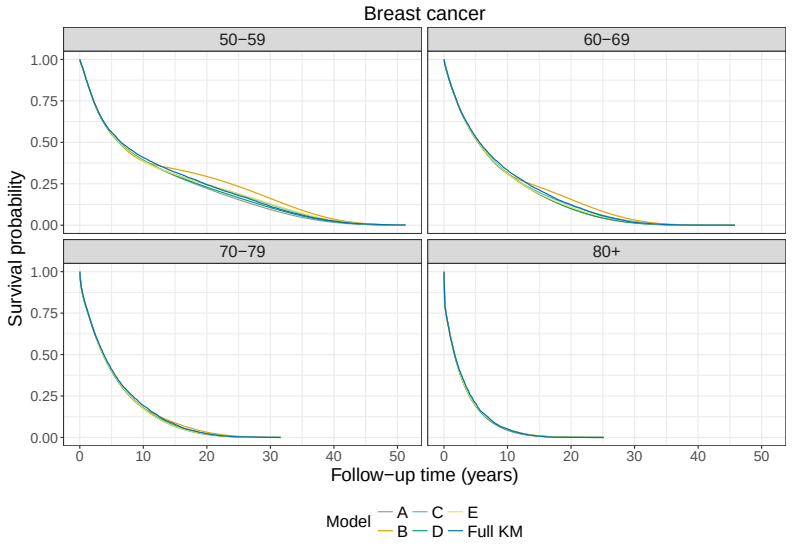

Figure S2: The extrapolated survival function of breast cancer patients based on 5 relative survival models.

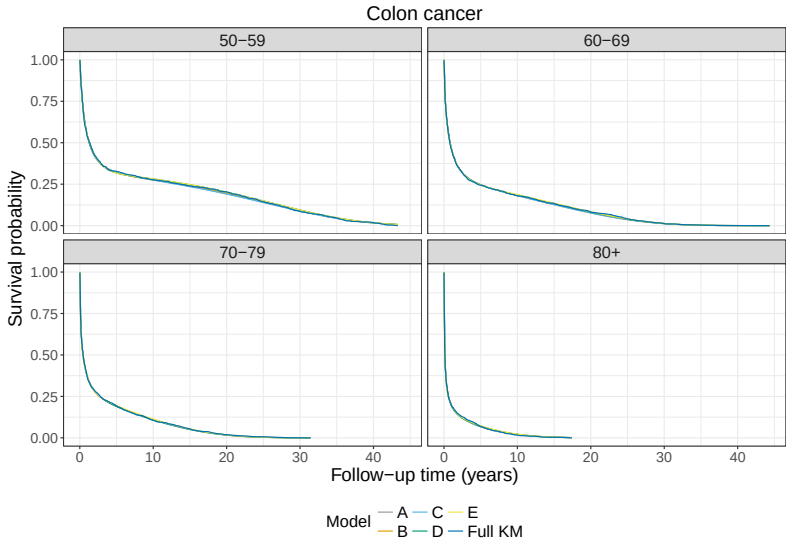

Figure S3: The extrapolated survival function of colon cancer patients based on 5 relative survival models.

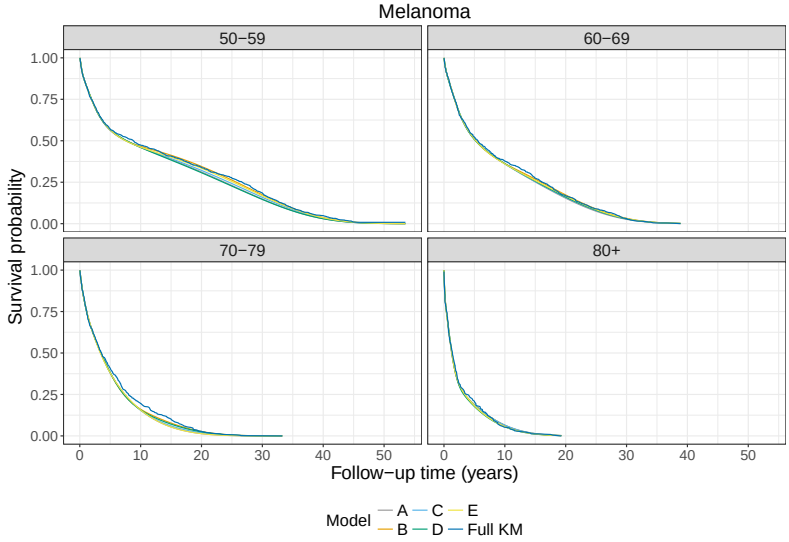

Figure S4: The extrapolated survival function of melanoma patients based on 5 relative survival models.

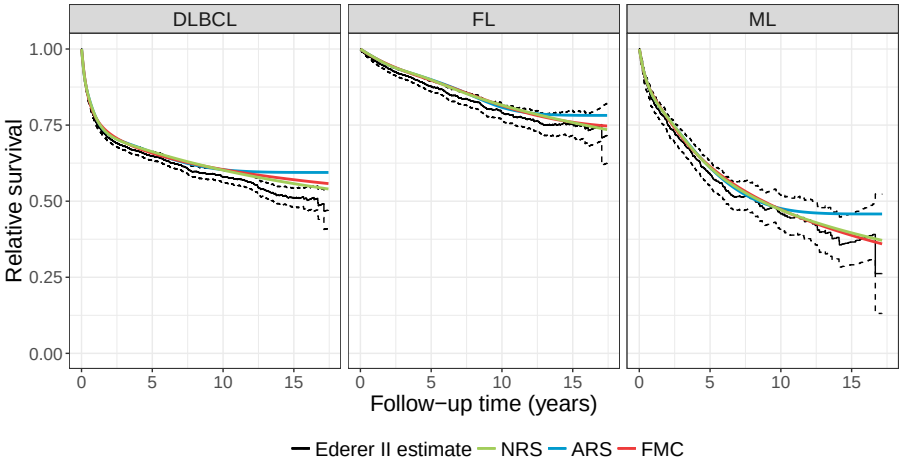

Figure S5: The relative survival of Danish diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and mantle cell lymphoma (ML) patients calculated by the Ederer II method including confidence intervals (dashed lines), the NRS model, the ARS model, and the FMC model.

Paper III

Estimating the time to statistical cure

Lasse H. Jakobsen[1,2], Therese M.-L. Andersson[3], Jorne L. Biccler[1,2], Laurids Ø. Poulsen[4], Marinne T. Severinsen[2], Tarec C. El-Galaly[1,2], and Martin Bøgsted[1,2]

1. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; 2. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; 3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4. Department of Oncology, Aalborg University Hospital, Aalborg, Denmark.

Submitted to *Statistics in Medicine*, 2018.

Description

As patients with cancer, remain alive after diagnosis, their survival will in some cases approach that of the general population. If these eventually become the same, the patients still alive are considered *statistically cured*. While estimation of measures related to statistical cure, such as the proportion of cured patients and the survival of those who are uncured, has been described before, estimation of the time from which there is no longer any excess mortality, termed the *cure point*, has received limited attention. The aim of this paper was to discuss approaches and develop an appropriate method for estimating cure points.

Abstract

In cancer studies, the term to be statistically cured is used to describe patients who over time obtain the same expected mortality risk as the matched general population. The time to statistical cure, often coined the cure point, is of great interest to patients, clinicians, and health care planners. As mortality risks can be quantified in various ways, cure point estimation has been handled in an ad hoc fashion in previous scientific literature, often without considerations about margins of clinical relevance. In this article, we review existing cure point estimators and suggest new clinically relevant estimators based on the loss of lifetime and the conditional probability of cancer-related death. The performance of the proposed method is assessed in a simulation study and the method is illustrated on survival data from Danish colon cancer, acute myeloid leukemia, and diffuse large B-cell lymphoma patients. Despite the usefulness of cure points, one has to be cautious as the estimated cure point can be very sensitive towards the choice of the clinical relevant margin.

1 Introduction

One of the most important aims of cancer patients is to become cured, which in general is different from reaching complete remission due to the risk of relapse, lethal side effects, and late toxicities. However, for some cancers the patient mortality risk reaches the same level as the general population mortality risk. This suggests a possible formulation of cure, namely *statistical cure*, which is achieved if the patients survive until the point at which the patient and general population mortality risks become similar [1]. The time point at which the patients become statistically cured is termed the *cure point* and its estimation is the main focus of this paper.

The clinical problem of estimating cure points may be generalized by considering when the risk of a certain event in exposed individuals equals that of unexposed individuals. One example is smoking cessation, which is associated with a short term increased risk of developing type 2 diabetes, that over time gradually approaches that of non-smokers [2]. A natural question for the ex-smoker is when this risk is normalized. Another example involves the fertility of women who have terminated the use of oral contraceptives, which is typically inferior to that of women who exclusively used diaphragms [3, 4]. However, as time progresses, the fertility returns to the same level as previous diaphragm users. In this case the time at which the fertility level equals that of the control group may be of interest. In the present article, we only consider estimation of survival-related cure points for cancer patients, although the statistical problem applies to other diseases and end-points.

In the applied setting, cure point estimation has implicitly been the focus of a number of recent cancer studies [5–10]. Although researchers often apply ad hoc approaches to compute cure points, methods for estimating cure points have been introduced previously. Rabinowitz and Ryan proposed a time-varying test-statistic for establishing a lower confidence bound for the cure point [11]. However, the lower confidence bound is difficult to interpret in a clinical context and it should be used with caution since it cannot be determined whether statistical cure has occurred or if there is simply not enough data to show a significant excess risk. More recently, methods based on cure models have been proposed. Lambert et al. suggested to compute the cure point as the time at which the relative survival of the uncured patients reaches some clinical relevant margin [12]. However, since this function is interpreted in a rather hypothetical setting where deaths from other causes than cancer cannot occur, deciding on such a margin may be difficult. Boussari et al. proposed to compute the cure point as the time at which the conditional probability of being statistically cured exceeds a specific level, e.g., 95% [13]. However, this method relies fully on cure models, where an explicit definition of the cure point is needed, which seems counter-intuitive.

In this article, we review existing methods and introduce a general approach for cure point estimation from cancer survival data. The performance of the proposed method is evaluated in a simulation study and we illustrate the method on survival data from Danish colon cancer (CC), acute myeloid leukemia (AML), and diffuse large B-cell lymphoma (DLBCL) patients.

2 Methods

Let T be the random survival time of each patient and D the random cause of death variable, which is either *cancer* or *other causes*. Furthermore, let Y denote the unobserved random variable denoting whether the patients are statistically cured ($Y = 1$) or uncured ($Y = 0$). Given a covariate vector $\boldsymbol{z}$, we denote the hazard and survival functions of the patient population at time t by $h(t|\boldsymbol{z})$ and $S(t|\boldsymbol{z})$, respectively and those of the general population are denoted by $h^*(t|\boldsymbol{z})$ and $S^*(t|\boldsymbol{z})$, respectively. The relative survival function, $R(t|\boldsymbol{z})$, is the ratio of the all-cause survival to the expected survival, i.e., $R(t|\boldsymbol{z}) = S(t|\boldsymbol{z})/S^*(t|\boldsymbol{z})$.

2.1 Cure point definition

Excess hazard and conditional relative survival

Following Rabinowitz and Ryan [11], the excess hazard, $\lambda(t|\boldsymbol{z}) = h(t|\boldsymbol{z}) - h^*(t|\boldsymbol{z})$, i.e., the difference between the patient and general population haz-

ard, may be used to define the cure point as

$$t_\epsilon(\boldsymbol{z}) = \inf\{t| \text{ for all } s \geq t, \lambda(s|\boldsymbol{z}) \leq \epsilon\}.$$

This corresponds to the time point at which the excess hazard, $\lambda(t|\boldsymbol{z})$, becomes lower than some prespecified clinical relevant margin ϵ. Rabinowitz and Ryan suggested a time-varying test statistic for the lower confidence bound of t_ϵ with $\epsilon = 0$ and argued that the lower bound should not be interpreted as the cure point, but only implies that for all larger time points there is no significant excess mortality [11].

Dal Maso et al. [7] estimated the cure point as the time point at which the conditional relative survival,

$$\frac{R(t+u|\boldsymbol{z})}{R(t|\boldsymbol{z})},$$

was sufficiently high (e.g., >95%) for a certain time point, e.g., $u = 5$ years.

However, determining a margin of clinical relevance for the excess hazard or the conditional relative survival requires in depth understanding of their meaning and scale, which can be challenging.

Cure models

While the excess hazard and the conditional relative survival may not be appropriate for cure point estimation due to their problematic interpretation, alternatives based on functional parts of cure models have previously been proposed for this purpose. Cure models assume that a proportion, π, of the patients are statistically cured, i.e., $\pi(\boldsymbol{z}) = \mathrm{P}(Y = 1|\boldsymbol{z})$, while the remaining are uncured [12]. However, Y is unobserved and thus $\mathrm{P}(Y = 1|\boldsymbol{z})$ cannot be estimated directly. Instead $\pi(\boldsymbol{z})$ can be estimated from a mixture cure model. Using that the relative survival of cured patients is one, the total relative survival can be formulated as a mixture model,

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})]S_u(t|\boldsymbol{z}), \tag{1}$$

where $S_u(t|\boldsymbol{z})$ is the relative survival of the uncured patients, with $S_u(0|\boldsymbol{z}) = 1$ and $\lim_{t\to\infty} S_u(t|\boldsymbol{z}) = 0$, and we have that $\lim_{t\to\infty} R(t|\boldsymbol{z}) = \pi(\boldsymbol{z})$. Therefore, Lambert et al. proposed to compute the cure point as the time at which $S_u(t|\boldsymbol{z})$ becomes sufficiently low, e.g., 10% or 5% [12]. This approach was also used by Chauvenet et al. [5]. However, $S_u(t|\boldsymbol{z})$ is usually interpreted in a setting where deaths from other causes than cancer cannot occur, which is highly hypothetical for most cancer patients. Therefore, deciding on a clinical relevant margin for this measure may be challenging.

Another cure model-based measure is the conditional probability of cure

given survival until time t,

$$\mathrm{P}(Y = 1|T > t, \boldsymbol{z}) = \frac{\mathrm{P}(T > t|Y = 1, \boldsymbol{z})\mathrm{P}(Y = 1|\boldsymbol{z})}{\mathrm{P}(T > t|\boldsymbol{z})} = \frac{\pi(\boldsymbol{z})}{R(t|\boldsymbol{z})}, \tag{2}$$

which was proposed for cure point estimation by Boussari et al. [13]. The cure point is estimated as the time point at which the probability of cure is sufficiently close to one, e.g., exceeding 95%. This measure provides a more intuitive interpretation than $S_u(t)$, but relies on accurate estimation of $\pi(\boldsymbol{z})$, which is challenging in scenarios where this is based on extrapolating beyond the available follow-up [12].

Conditional probability of cancer-related death

The severity of a cancer can be assessed through the cancer-specific cumulative incidence which can be derived from either cause of death information or relative survival. By using the cumulative incidence, we derive the probability,

$$\begin{aligned}\mathrm{P}(D = \text{cancer}|T > t, \boldsymbol{z}) &= \frac{\mathrm{P}(D = \text{cancer}, T > t|\boldsymbol{z})}{\mathrm{P}(T > t|\boldsymbol{z})} \\ &= \frac{\mathrm{P}(T < \infty, D = \text{cancer}|\boldsymbol{z}) - \mathrm{P}(T \leq t, D = \text{cancer}|\boldsymbol{z})}{\mathrm{P}(T > t|\boldsymbol{z})},\end{aligned} \tag{3}$$

where $\mathrm{P}(T < \infty, D = \text{cancer}|\boldsymbol{z}) = \mathrm{P}(D = \text{cancer}|\boldsymbol{z})$ is the probability of dying due to cancer given covariates $\boldsymbol{z}$. This function provides the conditional probability of eventually dying from cancer given survival until time t. Since all patients are bound to die at some point, computing $1 - \mathrm{P}(D = \text{cancer}|T > t|\boldsymbol{z})$ yields the conditional probability of eventually dying from other causes than cancer. For a given clinical relevant margin, the cure point can then be estimated as the time at which the probability of cancer-related death falls below the margin. Eloranta et al. [14] originally introduced this probability measure by considering specific models such that $\mathrm{P}(D = \text{cancer}) = \mathrm{P}(T \leq t_c, D = \text{cancer})$ for a finite time point t_c, but we do not employ this restriction here.

Loss of lifetime

Due to its straightforward interpretation, the mean residual lifetime, which can be computed by $\int_t^\infty S(u|\boldsymbol{z})du/S(t|\boldsymbol{z})$, is occasionally used as an alternative to conventional effect measures in survival analysis. In addition to the already introduced measures, we consider the *loss of lifetime* function, given as the difference between the mean residual lifetime of the general population

and the patient population, i.e.,

$$L(t|z) = \frac{\int_t^\infty S^*(u|z)du}{S^*(t|z)} - \frac{\int_t^\infty S(u|z)du}{S(t|z)}. \tag{4}$$

This function yields the conditional number of years lost due to the cancer given survival until time t after the diagnosis. Similarly to the probability of cancer-related death, the cure point can be defined as the time at which the loss of lifetime becomes sufficiently low.

2.2 Estimation

For convenience, we introduce a general *comparison measure*, $G(h, h^*)(t)$, quantifying the difference between the patient and the general population hazard. For notational convenience we will write $G(t|z) = G(h, h^*)(t)$. An estimate of G, $\hat{G}$, is obtained by plugging in the estimated hazard function, $\hat{h}(t, z)$, and the known general population hazard function, $h^*(t, z)$. For the purpose of cure point estimation, we focus on the probability of cure (2), the probability of cancer-related death (3), and the loss of lifetime function (4).

Extrapolation

In order to compute the loss of lifetime and probability of cancer-related death from right-censored follow-up data, extrapolation of the involved survival functions and cause-specific hazard functions beyond the available follow-up period is required. For the loss of lifetime function, the expected survival, $S^*(t|z)$, can be extrapolated using the method of Ederer et al. [15] (Ederer I) and by making assumptions about the future population mortality rates [16]. For extrapolating the patient survival, $S(t|z)$, a common strategy is to incorporate external data to provide accurate extrapolations [17]. In particular, flexible parametric relative survival models enable extrapolation directly and has previously demonstrated good performance in long term survival data and simulations [18].

For the conditional probability of cancer-related death, the cancer-specific cumulative incidence is computed by

$$\mathrm{P}(T \leq t, D = \text{cancer}|z) = \int_0^t S(u|z)\lambda^{\text{cs}}(u|z)du, \tag{5}$$

where λ^{cs} is the cancer-specific hazard. Cause of death information is required to compute (5), but these are often incomplete and difficult to determine. Instead of relying on exact cause of death information, flexible parametric relative survival models can be used to compute the cancer-specific hazard by $\lambda^{\text{cs}}(t|z) = \lambda(t|z)$ and the overall survival by $S(t|z) = S^*(t|z)R(t|z)$,

as well as extrapolate both functions [19]. Again, $S^*(t|\boldsymbol{z})$ can be extrapolated using the Ederer I method.

Extrapolation is not require for the probability of cure (2) unless the cure point is outside the available follow-up in which case the probability can be directly extrapolated by using a parametric cure model.

Flexible relative survival models

Royston and Parmar introduced a fully parametric proportional hazards model with the log cumulative baseline hazard modelled by restricted cubic splines [20]. This model was extended to relative survival by Nelson et al. who modelled the log cumulative baseline excess hazard by restricted cubic splines [21]. That is, the relative survival is specified by,

$$\log\left[-\log(R(t|\boldsymbol{z}))\right] = s_0(x, \boldsymbol{\gamma}) + \boldsymbol{z}^T\beta, \tag{6}$$

where s_0 is a restricted cubic spline and $x = \log(t)$. We refer to this model as the NRS model. Andersson et al. altered the base functions of the restricted cubic splines in (6) to establish a cure model, which we will refer to as the ARS model [22]. In this model, the excess hazard is restricted to be zero beyond the last knot of the splines, resulting in a flat relative survival after this point. Lastly, we consider the flexible mixture cure (FMC) model introduced by Jakobsen et al. [18], where $S_u(t)$ is modelled by the splines of the Royston-Parmar model, i.e.,

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})]\exp\left(-\exp\left(s_0(x, \boldsymbol{\gamma}) + \boldsymbol{z}^T\beta\right)\right).$$

Due to the their flexibility, these models often provide similar estimates, but differ in the tail of their distributions, which controls the trajectory beyond the available follow-up. The NRS model is linear on the log-log scale beyond the last knot, while the ARS model is constant after the last knot. The splines of the FMC model are also linear beyond the last knot, but the relative survival is bounded downward by $\pi(\boldsymbol{z})$. The parameters of each model can be estimated by maximum likelihood.

Cure points

For 1 - the probability of cure (2), the probability of cancer-related death (3), and the loss of lifetime (4), survival improvement corresponds to a decrease in the comparison measure, G. Hence, for a clinical relevant margin, ϵ, the cure point is estimated by solving the equation

$$\hat{G}(t|\boldsymbol{z}) = \epsilon, \tag{7}$$

with respect to t. The variance of the cure point estimate can be calculated by the delta method (see Section S1) after appropriate smoothing of the general population survival.

We implemented the method in the `R`-package `cuRe`, see `https://github.com/LasseHjort/cuRe`. The package contains functions for fitting FMC models as well as computing the probability of cure (2), the probability of cancer-related death (3), and the loss of lifetime (4). The comparison measures can be computed using both the NRS, ARS, and the FMC model, except for the probability of cure, which can only be computed from cure models. The package also contains functions for computing cure points and corresponding variances based on either of the three comparison measures.

3 Simulation study

3.1 Simulation design

To investigate the performance of the proposed method, we conducted a simulation study. Data were simulated using a relative survival approach, asumming independence between the relative survival and the general population survival [23]. That is, the observed follow-up time was $\min(x, x^*, c)$, where x, x^*, and c were realizations generated from $R(t)$, $S^*(t)$, and a censoring distribution $C(t)$. The status indicator was $\mathbf{1}[\min(x, x^*) \leq c]$.

The general population survival function, S^*, was derived using a Danish life table from the Human mortality database [24]. For simplicity, all patients were assumed to be 60-year-old females diagnosed in 1980. The relative survival was simulated from the mixture cure model in (1) using a Weibull distribution for S_u. We considered three scenarios with varying levels of severity mainly controlled by the cure proportion, π (see Figure S1 for the relative survival trajectories). To mimic register data, the censoring distribution was chosen to be uniform(0, 10). The simulations were repeated 500 times using a sample size of 2,000.

3.2 Simulation results

For a given clinical relevant margin, ϵ, and comparison measure, $G(t)$, the true cure point, t_ϵ, was obtained by inserting the true hazard into $G(t)$ and solving (7). Cure point estimates, $\hat{t}_\epsilon$, were obtained by fitting a relative survival model and inserting the estimated hazard into $G(t)$ and solving (7). We considered 4 relative survival models for this purpose:

- The ARS model with knots placed at the 0, 25, 50, 75, and 95 percentiles of the uncensored follow-up times and the last knot at 10 years. If the

95 percentile was smaller than 4 years, an additional knot was placed at 6 years.

- The FMC model with knots placed at the 0, 33, 67, and 100 percentiles of the uncensored follow-up times. The cure proportion was modelled using a logit link function as suggested by Lambert et al [12].
- The NRS model with knots placed at the 0, 25, 50, 75, and 100 percentiles of the uncensored follow-up times.
- The Weibull mixture cure model from which the data were simulated. This model was included to assess the performance using the "true" model.

			ARS model			FMC model			NRS model			Weibull mixture		
	ϵ	t_ϵ	Bias	$\widehat{\text{Var}}(\hat{t}_\epsilon)$	ECP (%)	Bias	$\widehat{\text{Var}}(\hat{t}_\epsilon)$	ECP (%)	Bias	$\widehat{\text{Var}}(\hat{t}_\epsilon)$	ECP (%)	Bias	$\widehat{\text{Var}}(\hat{t}_\epsilon)$	ECP (%)
Loss of lifetime														
Scenario 1														
	1.00	3.34	0.05	0.06	92.8	0.04	0.04	95.0	0.32	0.64	96.2	0.01	0.02	94.4
	2.00	2.86	0.04	0.02	93.8	0.03	0.02	95.6	0.15	0.17	97.2	0.01	0.01	94.8
	3.00	2.55	0.03	0.01	93.6	0.02	0.01	95.6	0.07	0.07	96.8	0.01	0.01	95.2
Scenario 2														
	1.00	4.15	0.09	0.07	92.8	0.06	0.06	95.2	0.89	2.14	94.8	0.02	0.03	94.6
	2.00	3.55	0.05	0.03	94.0	0.04	0.03	94.6	0.33	0.36	97.0	0.01	0.02	94.4
	3.00	3.15	0.04	0.02	94.4	0.03	0.02	95.2	0.15	0.13	96.2	0.01	0.01	94.4
Scenario 3														
	1.00	3.05	0.08	0.04	88.4	0.22	0.28	96.6	1.12	1.49	90.4	0.07	0.06	93.8
	2.00	2.16	0.05	0.03	92.4	0.11	0.07	95.0	0.37	0.22	94.2	0.05	0.03	94.4
	3.00	1.61	0.03	0.02	93.4	0.07	0.03	93.6	0.20	0.06	91.8	0.04	0.02	92.6
Conditional probability of cancer-related death														
Scenario 1														
	0.05	3.39	0.05	0.07	92.8	0.05	0.05	95.0	0.45	3.30	92.8	0.02	0.02	94.4
	0.10	2.89	0.04	0.03	93.6	0.03	0.02	95.4	0.17	0.56	94.4	0.01	0.01	94.6
	0.15	2.57	0.03	0.02	93.6	0.03	0.02	95.6	0.06	0.20	95.0	0.01	0.01	95.0
Scenario 2														
	0.05	4.24	0.11	0.09	91.8	0.07	0.07	95.2	2.06	14.47	88.8	0.02	0.03	94.4
	0.10	3.62	0.06	0.04	94.2	0.05	0.04	94.8	0.68	2.08	95.8	0.02	0.02	94.8
	0.15	3.22	0.05	0.02	94.6	0.04	0.02	95.2	0.31	0.54	97.0	0.02	0.02	94.4
Scenario 3														
	0.05	3.15	0.09	0.05	87.6	0.31	0.71	96.2	2.78	9.56	85.2	0.09	0.07	94.0
	0.10	2.22	0.06	0.03	92.2	0.16	0.15	96.2	0.82	1.07	95.6	0.06	0.04	94.4
	0.15	1.64	0.03	0.02	93.0	0.10	0.05	94.8	0.41	0.24	94.4	0.05	0.02	93.0
Conditional probability of cure														
Scenario 1														
	0.05	3.39	0.05	0.07	92.4	0.05	0.05	95.0				0.02	0.02	94.4
	0.10	2.90	0.04	0.03	93.6	0.03	0.02	95.2				0.01	0.01	94.6
	0.15	2.58	0.03	0.02	93.6	0.03	0.02	95.6				0.01	0.01	95.0
Scenario 2														
	0.05	4.25	0.12	0.09	91.8	0.07	0.08	95.2				0.02	0.03	94.4
	0.10	3.63	0.06	0.04	94.4	0.05	0.04	95.0				0.02	0.02	94.8
	0.15	3.23	0.05	0.02	94.8	0.04	0.02	95.2				0.02	0.02	94.4
Scenario 3														
	0.05	3.17	0.09	0.05	87.6	0.32	1.50	96.0				0.09	0.08	94.2
	0.10	2.24	0.06	0.03	92.2	0.16	0.29	96.2				0.06	0.04	94.4
	0.15	1.66	0.04	0.02	93.0	0.10	0.09	95.6				0.05	0.02	93.6

Table 1: Bias, variance, and coverage of the cure point estimate in simulated data. The cure point estimates were based on the loss of lifetime function, the probability of cancer-related death, and the probabiltiy of cure. The NRS model was not evaluated for the latter measure since this is not a cure model. ARS: relative survival model by Andersson et al. [22], FMC: flexible mixture cure model by Jakobsen et al. [18], NRS: relative survival model by Nelson et al. [21], ECP: empirical coverage probability.

The cure point bias was computed as $\frac{1}{500}\sum_{j=1}^{500}\hat{t}_{\epsilon,j} - t_\epsilon$, where $\hat{t}_{\epsilon,j}$ is the cure point estimate in the j$^{\text{th}}$ simulation. The empirical coverage probability

(ECP) was calculated as the proportion of simulations where t_ϵ was within the 95% confidence interval of $\hat{t}_\epsilon$ and we denote by $\overline{\text{Var}}(\hat{t}_\epsilon)$ the empirical mean variance across all simulations. The bias, mean variance, and ECP of the estimated cure points are displayed in Table 1 for each scenario and for varying clinical relevant margins. The cure point was calculated using the loss of lifetime (4), the probability of cancer-related death (3), and the probability of cure (2).

Across all three measures, the results were fairly similar displaying smaller biases for larger clinical relevant margins. The biases obtained by using the ARS, FMC, and Weibull mixture cure models were low, while larger biases were seen for the NRS model. However, the tail of the NRS model, which is not a cure model, varies from that of the mixture cure model used to generate the survival times. Since both the loss of lifetime (4) and conditional probability of cancer-related death (3) require extrapolation, this likely explains the inferior performance of the NRS model.

The variances and biases of the ARS and FMC models were similar, but slightly elevated compared to the true Weibull model. In scenario 3, the FMC model showed subpar performance with respect to both bias and variance, but the ECP remained sufficiently high. In fact, the FMC model was the only model with a majority of the ECPs exceeding 95%. The ECP of the ARS model was slightly lower than expected in scenario 3 using the lowest margin. This may be explained by the combination of the biased estimates and a relatively small cure point variance. Since the ARS model is restricted to have zero excess hazard beyond its last knot, the cure point estimate will typically be affected by the location of this knot. For general cure point estimation, we therefore suggest that other models are used or that the last knot of the ARS model is placed well beyond the point at which the excess hazard can be assumed to be zero. In these simulation, the good performance of the ARS model may be explained by the relative survival models used for simulating, where a clear cure pattern was assumed within the follow-up period.

The NRS model was not assessed for the probability of cure (2) since a cure model formulation of the relative survival is required. In two out of 500 simulations, no solution to (7) was found within 80 years for the probability of cure using the FMC model. From inspection of these cases, the models seemed to fit the data well, although the cure proportion estimates were substantially downward biased which directly affects the probability of cure, and thus the cure point estimate. This identifiability issue is the main criticism of using (2) as comparison measure. In real world data, the assumption of statistical cure will typically be more difficult to verify. Therefore, it is expected that this problem occurs more frequently than in these simulations. This problem was not observed for the ARS model, due to its restrictions after the last knot, nor the Weibull mixture cure model from which the data were simulated.

4 Analysis of Danish cancer register data

4.1 Data description

To illustrate the proposed method, we analyzed patient data from three malignant diseases: CC, AML (not otherwise specified), and DLBCL retrieved from the Danish Colorectal Cancer Group Database [25], the Danish National Acute Leukemia Registry [26], and the Danish National Lymphoma Registry [27], respectively. Each register ensures accurate follow-up on deaths by merging with the Danish Civil Registration System [28]. The selection of the three diseases was based on previous studies displaying statistical cure in CC ($\geq$20 years of age), AML, and DLBCL (18-50 years of age) patients [8, 10, 29]. For CC and AML all adult patients ($\geq$18 years of age) diagnosed between 2000 and 2016 were included, while only DLBCL patients between 18 and 50 years of age were included. Follow-up was measured as the time from diagnosis until death or censoring (June 2017). The Danish general population mortality rates were obtained from a publicly available life table retrieved from the Human Mortality Database [24]. When mortality rates for calendar years beyond those available in the life table were needed, the age- and sex-specific rates from the last available calendar year were used. The study was approved by the Danish Data Protection Agency (2008-58-0028). The FMC model was fitted separately to each disease using four knots placed at the 0, 33, 67, and 100 percentiles of the uncensored follow-up times and a logit link function for the cure proportion.

4.2 Results

In total, 42,380 CC, 1,887 AML, and 762 DLBCL patients were included in the study. The 5-year Kaplan-Meier estimate was 49% (95% CI, 48-49%), 15% (95% CI, 13-17%), and 83% (95% CI, 80%-86%) for CC, AML, and DLBCL, respectively. The fitted FMC models are shown in Figure 1 together with a non-parametric relative survival estimate calculated by the Ederer II method [30]. Estimates of the 5-year relative survival, the cure proportion, the probability of dying due to cancer, and the baseline (at time zero) loss of lifetime are shown in Table 2. Each disease showed an immediate steep relative survival function which ultimately flattens out. The plateau level of the diseases differed greatly. Among the young DLBCL patients, the cure proportion was rather high while for the AML patients it was low, with few patients alive after 10 years.

The loss of lifetime function and the conditional probability of cancer-related death are shown in Figure 2. For the AML patients, the loss of lifetime was initially large followed by a steep decrease during the first five years. The same pattern was seen in the conditional probability of cancer-related death.

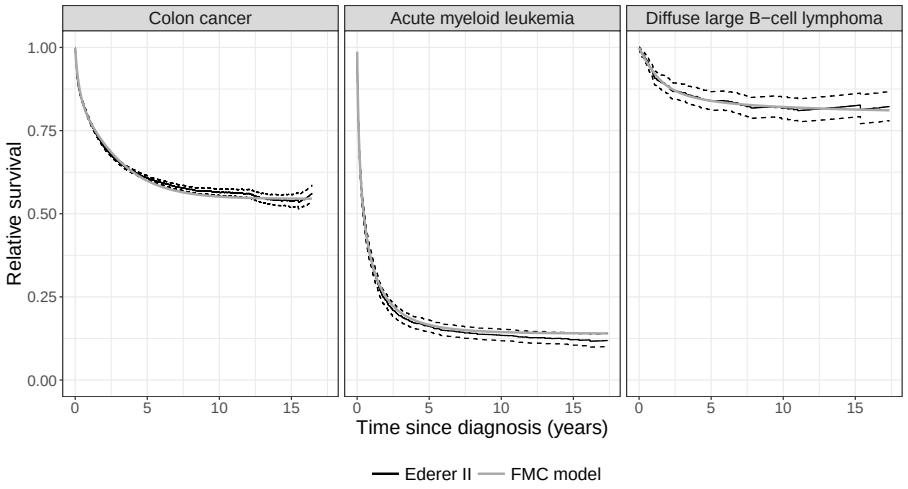

Figure 1: The relative survival of Danish CC, AML, and DLBCL patients calculated by the Ederer II method including 95% confidence intervals (dashed lines) and the flexible mixture cure (FMC) model.

	CC (n = 42,380)	AML (n = 1,887)	DLBCL (n = 762)
Mean age (range)	72(22-105)	67(18-99)	40(18-50)
5-year RS (Ederer II)	0.61(0.60-0.61)	0.16(0.15-0.18)	0.84(0.81-0.87)
5-year RS (parametric)	0.60(0.59-0.61)	0.17(0.15-0.18)	0.84(0.81-0.87)
Cure proportion	0.54(0.53-0.56)	0.14(0.12-0.16)	0.79(0.67-0.88)
Probability of dying due to cancer	0.42(0.41-0.42)	0.83(0.81-0.85)	0.20(0.15-0.25)
Baseline loss of lifetime (years)	5.87(5.77-5.97)	15.13(14.81-15.45)	7.43(5.93-8.92)

Table 2: Relative survival estimates, cure proportion, probability of dying due to cancer, and baseline loss of lifetime estimates. CC: colon cancer, AML: acute myeloid leukemia, DLBCL: diffuse large B-cell lymphoma, RS: relative survival.

The loss of lifetime trajectories were similar between the young DLBCL patients and the CC patients despite the superior relative survival of the DLBCL patients. This is likely explained by the lower age of the DLBCL patients. However, the conditional probability of cancer-related death was higher in CC compared to the DLBCL patients. Thus, while the relatively small excess mortality among the young DLBCL patients heavily influences the loss of lifetime function, it does not have as big an impact on the probability of cancer-related death.

Figure 3 displays the estimated cure point in each disease, obtained by solving (7), for a varying level of clinical relevance. Whenever the margin of clinical relevance was low, small changes to the margin implied substantial changes in the estimated cure point, but for larger values, the cure point became less sensitive towards the choice of margin. For example, increasing

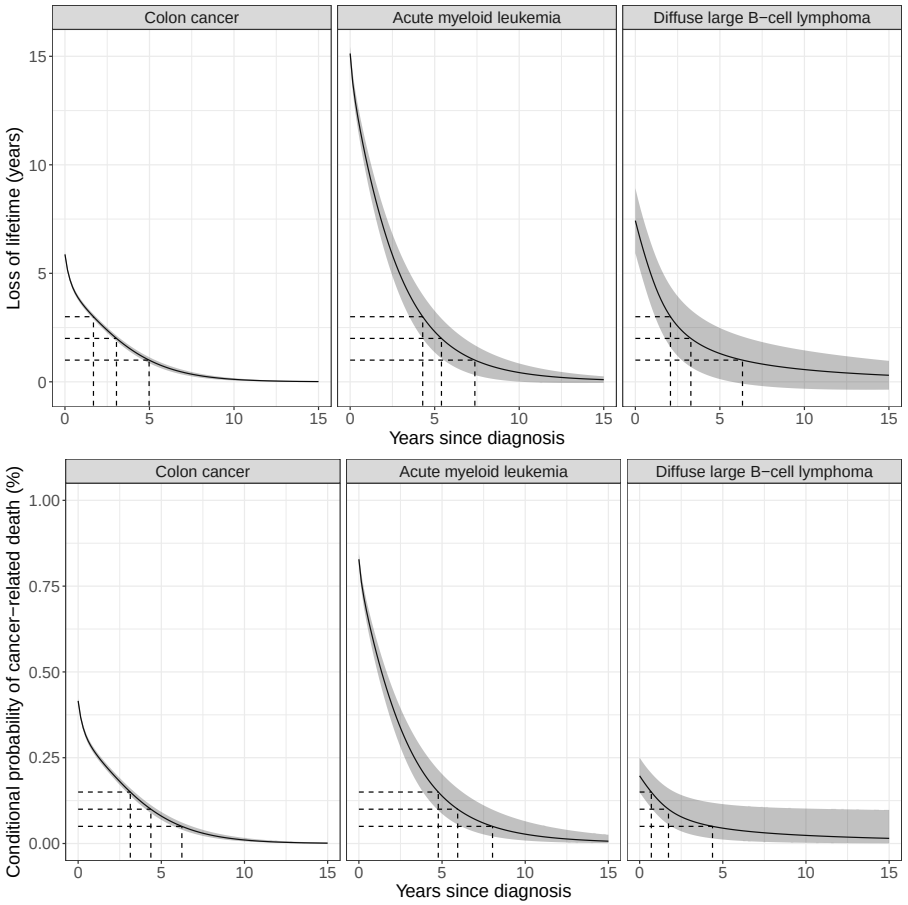

Figure 2: Upper: The loss of lifetime function within the first 15 years after diagnosis for Danish CC, AML, and DLBCL patients. The dashed lines indicate cure point estimates based on three different clinical relevant margins; 1, 2, and 3 years. Lower: The conditional probability of cancer-related death within the first 15 years after diagnosis for Danish CC, AML, and DLBCL patients. The clinical relevant margins are 0.05, 0.10, and 0.15. The shaded areas indicate pointwise 95% confidence intervals.

the margin for the loss of lifetime function from 3 years to 4 years in DLBCL resulted in a decrease in the cure point estimate from 2.08 to 1.39 years, while increasing the margin from 0.5 to 1 year resulted in a decrease from 10.90 to 6.34 years.

For colon cancer, the patients were stratified according to age group (-60, 60-70, 70-80, 80-), gender, and clinical stage (UICC stage I-II, III-IV) and the FMC model was fitted to each subgroup separately. The stratified cure point estimates computed using the conditional probability of cancer-related death (see Figure S2 and S3) as comparison measure and three clinical relevant

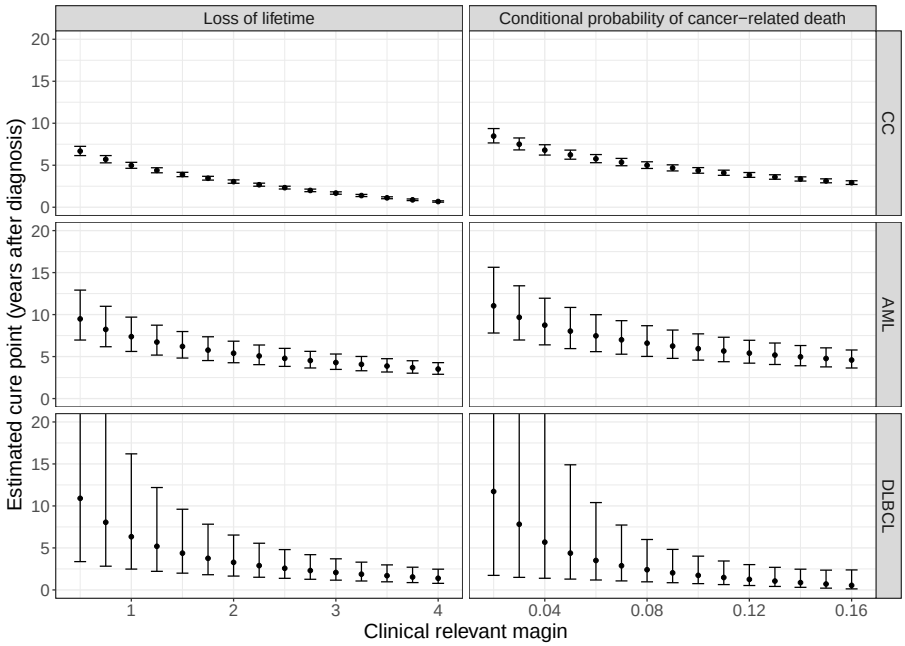

Figure 3: The estimated cure points with 95% confidence intervals against the clinical relevant margin in CC, AML, and DLBCL patients. For the loss of lifetime function, the margin is given in years. CC: colon cancer, AML: acute myeloid leukemia, DLBCL: diffuse large B-cell lymphoma.

margins (0.025, 0.05, and 0.075) are shown in Figure 4.

The cure point estimates of female and male patients displayed similar trends. In certain strata, the confidence interval of the cure point was wide, which makes the usability of these estimates difficult. For other subgroups, the variance was reasonably small and led to stable estimates. The cure point for low stage patients >80 years of age was very small, and due to the slope of the conditional probability of cancer-related death (Figure S2 and S3), the corresponding confidence interval was narrow.

5 Discussion

Cure points enable communication of prognostic information to cancer patients and are particularly useful for patients attending routine follow-up. Also health care planners may find cure points useful, e.g., for deciding the duration of the follow-up period. If an early cure point is detected, the duration may be adjusted accordingly, which potentially lowers the cost of the follow-up program and avoids unnecessary patient anxiety. However, the detection of long term toxicities is typically also a goal of routine follow-up

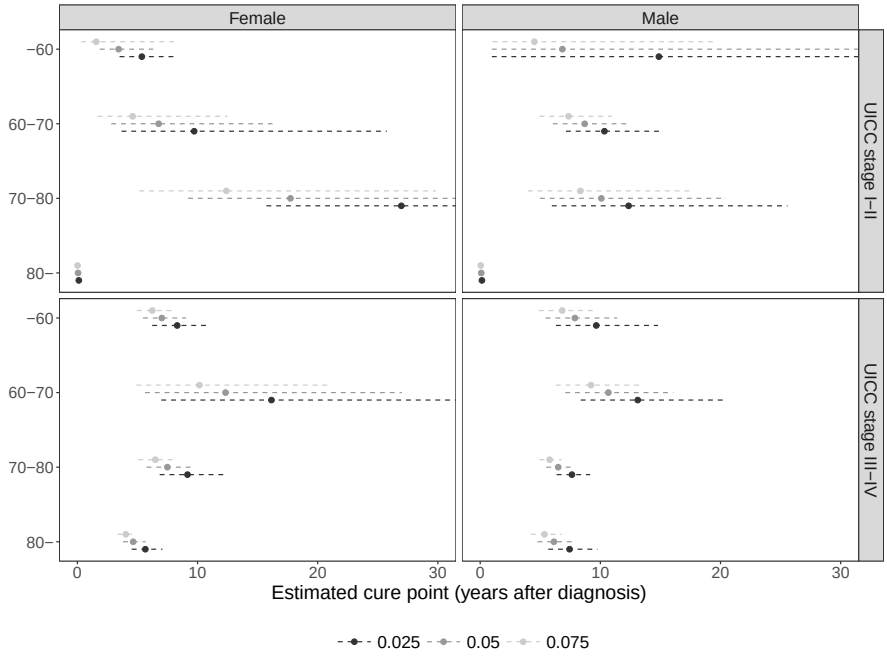

Figure 4: The estimated cure point for Danish colon cancer patients stratified on age group (-60, 60-70, 70-80, 80-), gender, and stage (UICC stage I-II vs. III-IV). The estimates are based on the conditional probability of cancer-related death (see Figure S2 and S3). UICC: Union for International Cancer Control.

programs and should be considered alongside the mortality risk.

We presented a general framework for the estimation of cure points from cancer survival data. Such a framework has not previously been established and researchers often consider ad-hoc methodologies to obtain cure point-like estimates. In recent studies of lymphoma, Maurer et al. and Hapgood et al. used sequential testing of standardized mortality ratios to evaluate the survival improvement of the patients [6, 9]. In our previous work, we combined the same standardized mortality ratio approach with restricted loss of lifetime estimates within DLBCL and concluded that the patients had a sufficiently low restricted loss of lifetime after 2 years of event-free survival [10]. Since hypothesis testing is generally sample size dependent, using a sequential testing approach implies that the cure point is likely to occur later for larger data sets. Therefore, approaches that directly involve hypothesis testing are not appropriate for cure point estimation. Dal Maso et al. [7] computed the cure point of various cancers using the conditional relative survival approach, while Chauvenet et al. [5] computed the cure point of colorectal patients as the time at which the relative survival of the uncured fell below

10%. Andersson et al. used the loss of lifetime approach to evaluate the survival progression of colon cancer patients and argued that the loss of lifetime was sufficiently low after 8-10 years of survival [8].

In studies of statistical cure or similar measures in cancer survival, the results should be interpreted with respect to the inclusion criteria of the study. For instance, if a study includes patients treated with a specific therapy, which is typically not given to frail patients, or merely includes patients with high socioeconomic status, the resulting patient population may be in slightly better shape than the general population. Thus, for patients achieving statistical cure, a significant excess mortality may still exist if compared to non-diseased individuals with similar physical capabilities or similar socioeconomic status.

The main issue with the loss of lifetime function and the conditional probability of cancer-related death is the need for extrapolation for which the accuracy cannot be assessed. Therefore, sensitivity analyses are recommended to accompany these estimates [18]. To avoid extrapolation, the model by Andersson et al. where cure is assumed at a specific time point, can be used to compute these measures [22]. However, estimating the cure point from a model where the cure point is explicitly defined seems counter-intuitive and we do generally not recommend this approach. Nevertheless, Boussari et al. used this model to estimate the cure point by using the probability of cure (2) [13]. As demonstrated in the simulation study, (2) could also be computed from regular cure models where cure occurs at an asymptote (the FMC model or the Weibull mixture cure model). However, computing (2) would require reliable cure proportion estimates which can be problematic in some scenarios due to the intrinsic identifiability issue of cure models which makes the cure proportion sensitivity towards the tail of the parametric distribution [12].

Deciding on margins of clinical relevance is an essential part of the present methodology. The approach is known from non-inferiority testing where hypothesis testing is conducted using a non-inferiority margin [31]. The choice of clinical relevant margin may be based on, e.g., the age and gender distribution of the considered patient population, or if single-patient cure points are of interest, the specific characteristics of the patient. Therefore, it is important that researchers with experience within the field of research aid in deciding on a clinical relevant margin which also emphasizes the need for considering comparison measures that are interpretable to non-statisticians.

The search for surrogate end-points to be used in clinical trials in order to increase the pace at which these are executed has been the focus of recent cancer studies [32]. Cure points may be used to derive new surrogate endpoints, since prolonging the study period beyond the cure point may not be necessary. However, as previously observed by Stephens et al. [33], long term risks cannot be observed in studies with short follow-up and additional

validation of new end-points in a series of clinical trials is required [34].

Using the conditional probability of cancer-related death, stratified cure points were computed for the colon cancer patients. However, to obtain cure points more applicable to the clinical setting and to potentially decrease the variance, modelling of the covariates could be utilized instead. For relative survival models, proportional excess hazards models are commonly used, and non-proportional hazards can readily be obtained by including time-varying covariate effects [35, 36]. Based on the fitted relative survival model, the individual cure point estimate can be computed by solving (7) using a patient-specific clinical relevant margin.

Because of the varying survival trajectory of cancers, cure points are not useful for all diseases, particularly not cancers where statistical cure is not achieved. To accompany analyses of statistical cure, a formal test for the assumption of statistical cure would be convenient. Formal tests based on simple parametric models have previously been introduced [37, 38], but these are not commonly used and suffer from a number of practical disadvantages [39].

The sensitivity of the cure point towards the choice of clinical relevance is a key criticism of the present approach. As demonstrated in Section S1, the variance of the cure point is inversely proportional to the derivative of the comparison measure evaluated at the estimated cure point. Therefore, the cure point is fairly robust whenever the progression measure is steep and sensitive whenever the progression measure is flat. Also, if only a slow improvement is seen in the comparison measure, the cure point may change dramatically if the margin is slightly changed. We suggest that the sensitivity should be tested and discussed whenever cure points are estimated.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

References

[1] P. Tralongo, M. S. McCabe, and A. Surbone. Challenge for cancer survivorship: improving care through categorization by risk. *Journal of Clinical Oncology*, 35(30):3516–3517, 2017.

[2] H.-C. Yeh, B. B. Duncan, M. I. Schmidt, N.-Y. Wang, and F. L. Brancati. Smoking, smoking cessation, and risk for type 2 diabetes mellitus. *Annals of Internal Medicine*, 152(1):10–17, 2010.

[3] M. P. Vessey, N. H. Wright, K. McPherson, and P. Wiggins. Fertility after stopping different methods of contraception. *British Medical Journal*, 1(6108):265–267, 1978.

[4] S. Linn, S. C. Schoenbaum, R. R. Monson, B. Rosner, and K. J. Ryan. Delay in conception for former 'pill' users. *JAMA: The Journal of the American Medical Association*, 247(5):629–632, 1982.

[5] M. Chauvenet, C. Lepage, V. Jooste, V. Cottet, J. Faivre, et al. Prevalence of patients with colorectal cancer requiring follow-up or active treatment. *European Journal of Cancer*, 45(8):1460–1465, 2009.

[6] M. J. Maurer, H. Ghesquières, J.-P. Jais, T. E. Witzig, C. Haioun, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *Journal of Clinical Oncology*, 32(10):1066–1073, 2014.

[7] L. Dal Maso, S. Guzzinati, C. Buzzoni, R. Capocaccia, D. Serraino, et al. Long-term survival, prevalence, and cure of cancer: a population-based estimation for 818 902 Italian patients and 26 cancer types. *Annals of Oncology*, 25(11):2251–2260, 2014.

[8] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, A. Sjövall, M. Lambe, et al. The loss in expectation of life after colon cancer: a population-based study. *BMC Cancer*, 15(1):412, 2015.

[9] G. Hapgood, Y. Zheng, L. H. Sehn, D. Villa, R. Klasa, et al. Evaluation of the Risk of Relapse in Classical Hodgkin Lymphoma at Event-Free Survival Time Points and Survival Comparison With the General Population in British Columbia. *Journal of Clinical Oncology*, 34(21):2493–2500, 2016.

[10] L. H. Jakobsen, M. Bøgsted, P. d. N. Brown, B. Arboe, J. Jørgensen, et al. Minimal loss of lifetime for patients with diffuse large B-cell lymphoma in remission and event free 24 months after treatment: a Danish population-based study. *Journal of Clinical Oncology*, 35(7):778–784, 2017.

[11] D. Rabinowitz and L. Ryan. Lower confidence bounds for time to cure. *Biometrika*, 80(3):681–687, 1993.

[12] P. C. Lambert, J. R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, 2007.

[13] O. Boussari, G. Romain, L. Remontet, N. Bossard, M. Mounier, et al. A new approach to estimate time-to-cure from cancer registries data. *Cancer Epidemiology*, 53:72–80, 2018.

[14] S. Eloranta, P. C. Lambert, T. M.-L. Andersson, M. Björkholm, and P. W. Dickman. The application of cure models in the presence of competing risks. *Epidemiology*, 25(5):742–748, 2014.

[15] F. Ederer, L. M. Axtell, and S. J. Cutler. The relative survival rate: a statistical methodology. *National Cancer Institute monograph*, 6:101–121, 1961.

[16] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, M. Lambe, and P. C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, 2013.

[17] C. Jackson, J. Stevens, S. Ren, N. Latimer, L. Bojke, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Medical Decision Making*, 37(4):377–390, 2017.

[18] L. H. Jakobsen, T. M.-L. Andersson, J. Biccler, T. C. El-Galaly, and M. Bøgsted. Estimation of the loss of lifetime function using flexible parametric relative survival models. *Submitted for publication*, 2018.

[19] P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine*, 29(7-8):885–895, 2010.

[20] P. Royston and M. K. B. Parmar. 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(15):2175–2197, 2002.

[21] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, 2007.

[22] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, and P. C. Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, 2011.

[23] M. J. Rutherford, P. W. Dickman, and P. C. Lambert. Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology*, 36(1):16–21, 2012.

[24] Human mortality database. URL: `www.mortality.org`. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Accessed 15 September 2017.

[25] P. Ingeholm, I. Gögenur, and L. H. Iversen. Danish Colorectal Cancer Group Database. *Clinical Epidemiology*, 8:465–468, 2016.

[26] L. S. G. Østgård, J. M. Nørgaard, K. K. Raaschou-Jensen, R. S. Pedersen, D. Rønnov-Jessen, et al. The Danish National Acute Leukemia Registry. *Clinical Epidemiology*, 8:553–560, 2016.

[27] B. Arboe, P. Josefsson, J. Jørgensen, J. Haaber, P. Jensen, et al. Danish National Lymphoma Registry. *Clinical Epidemiology*, 8:577–581, 2016.

[28] C. B. Pedersen. The Danish Civil Registration System. *Scandinavian Journal of Public Health*, 39(Suppl 7):22–25, 2011.

[29] H. Bower, T. M.-L. Andersson, M. Björkholm, P. W. Dickman, P. C. Lambert, et al. Continued improvement in survival of acute myeloid leukemia patients: an application of the loss in expectation of life. *Blood Cancer Journal*, 6(2):e390, 2016.

[30] F. Ederer and H. Heise. Instructions to IBM 650 programmers in processing survival computations. *Methodological note No. 10, End Results Evaluation Section. Bethesda:National Cancer Institute*, 1959.

[31] E. Walker and A. S. Nowacki. Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, 26(2):192–196, 2011.

[32] Q. Shi, C. R. Flowers, W. Hiddemann, R. Marcus, M. Herold, et al. Thirty-month complete response as a surrogate end point in first-line follicular lymphoma therapy: an individual patient-level analysis of multiple randomized trials. *Journal of Clinical Oncology*, 35(5):552–560, 2017.

[33] D. M. Stephens, H. Li, M. L. LeBlanc, S. D. Puvvada, D. Persky, et al. Continued risk of relapse independent of treatment modality in limited-stage diffuse large B-cell lymphoma: final and long-term analysis of southwest oncology group study S8736. *Journal of Clinical Oncology*, 34(25):2997–3004, 2016.

[34] T. R. Fleming and J. H. Powers. Biomarkers and surrogate endpoints in clinical trials. *Statistics in Medicine*, 31(25):2973–2984, 2012.

[35] M. Pohar and J. Stare. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81(3):272–278, 2006.

[36] P. C. Lambert and P. Royston. Further development of flexible parametric models for survival analysis. *Stata Journal*, 9(2):265–290, 2009.

[37] Y. Peng, K. B. G. Dear, and K. C. Carriere. Testing for the presence of cured patients: a simulation study. *Statistics in Medicine*, 20(12):1783–1796, 2001.

[38] R. A. Maller and S. Zhou. Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 51(4):1197–1205, 1995.

[39] X. Yu, R. De Angelis, T. Andersson, P. Lambert, D. O'Connell, et al. Estimating the proportion cured of cancer: some practical advice for users. *Cancer Epidemiology*, 37(6):836–842, 2013.

Supplementary

S1 Variance of the cure point estimator

In the following, we derive a variance estimator for the cure point estimator using the delta method under the assumption that the hazard function $h(t, \boldsymbol{\theta})$ has a parametric form with parameters $\boldsymbol{\theta}$. Denote the parameter estimator by $\hat{\boldsymbol{\theta}}$, and assume that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with mean $\mathbf{0}$ and variance Σ, where $\boldsymbol{\theta}_0$ is the true parameter value and Σ is the inverse information matrix, i.e., minus the inverse of the expected Hessian matrix of the likelihood function evaluated at $\boldsymbol{\theta}_0$.

Let $G(t, \boldsymbol{\theta}) = G(h, h^*)(t)$ be the strictly monotone comparison measure at time t obtained by inserting the parameters of the hazard function into the comparison measure and assume that G is continuously differentiable with respect to $\boldsymbol{\theta}$ and t. Furthermore, let $t_\epsilon = G^{-1}(\epsilon, \boldsymbol{\theta})$ and $\hat{t}_\epsilon = G^{-1}(\epsilon, \hat{\boldsymbol{\theta}})$ for a fixed clinical relevant margin, ϵ. The variance of $\hat{t}_\epsilon$ can then be approximated directly by using the the delta method, i.e.,

$$\operatorname{Var}\left[\hat{t}_\epsilon\right] \approx \frac{1}{n} \left(\nabla_\theta t_\epsilon|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right) \Sigma \left(\nabla_\theta t_\epsilon|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^T. \tag{8}$$

Due to the definition of t_ϵ,

$$\nabla_{\boldsymbol{\theta}} G(t_\epsilon, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0},$$

and by the chain rule of vector functions we have that

$$\nabla_{\boldsymbol{\theta}} G(t_\epsilon, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\partial G(t, \boldsymbol{\theta})}{\partial t}|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \nabla_\theta t_\epsilon|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \nabla_{\boldsymbol{\theta}} G(t, \boldsymbol{\theta})|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Thus,

$$\nabla_\theta t_\epsilon|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\left(\frac{\partial G(t, \boldsymbol{\theta})}{\partial t}|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^{-1} \nabla_{\boldsymbol{\theta}} G(t, \boldsymbol{\theta})|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Inserting into (8) yields

$$\begin{aligned}
\operatorname{Var}\left[\hat{t}_\epsilon\right] =& \frac{1}{n} \left(\frac{\partial G(t, \boldsymbol{\theta})}{\partial t}|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^{-2} \left(\nabla_{\boldsymbol{\theta}} G(t, \boldsymbol{\theta})|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right) \Sigma \left(\nabla_{\boldsymbol{\theta}} G(t, \boldsymbol{\theta})|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^T \\
\approx& \left(\frac{\partial G(t, \boldsymbol{\theta})}{\partial t}|_{t=\hat{t}_\epsilon, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^{-2} \operatorname{Var}\left[G(t, \hat{\boldsymbol{\theta}})\right]|_{t=\hat{t}_\epsilon},
\end{aligned}$$

where $\operatorname{Var}\left[G(t, \hat{\boldsymbol{\theta}})\right]|_{t=\hat{t}_\epsilon}$ is the variance of $G(\hat{t}_\epsilon, \hat{\boldsymbol{\theta}})$ without taking into account the uncertainty of $\hat{t}_\epsilon$, i.e., the point-wise variance of G evaluated at the point $\hat{t}_\epsilon$. For obtaining a non-negative confidence interval for the cure point, $\hat{t}_\epsilon$, the

variance of the log-transformed estimator is computed by the delta method:

$$\text{Var}\left[\log(\hat{t}_\epsilon)\right] \approx \frac{1}{\hat{t}_\epsilon^2}\text{Var}\left[\hat{t}_\epsilon\right] \approx \frac{1}{\hat{t}_\epsilon^2}\left(\frac{\partial G(t,\boldsymbol{\theta})}{\partial t}|_{t=\hat{t}_\epsilon,\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^{-2}\text{Var}\left[G(t,\hat{\boldsymbol{\theta}})\right]|_{t=\hat{t}_\epsilon}.$$

S2 Additional figures

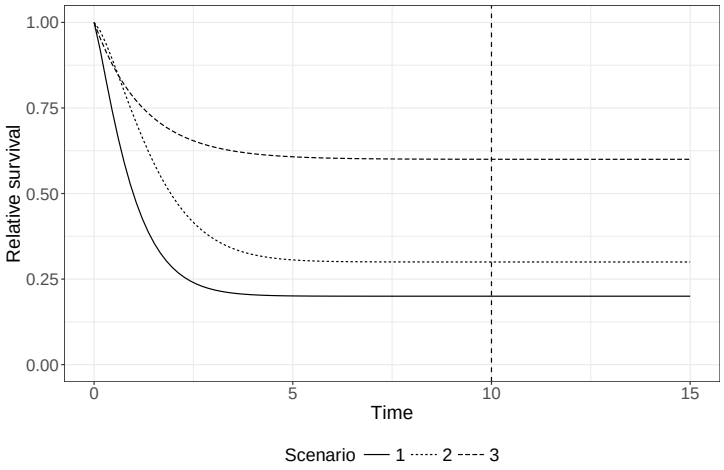

Figure S1: The relative survival trajectories of the models from which the simulated data were generated.

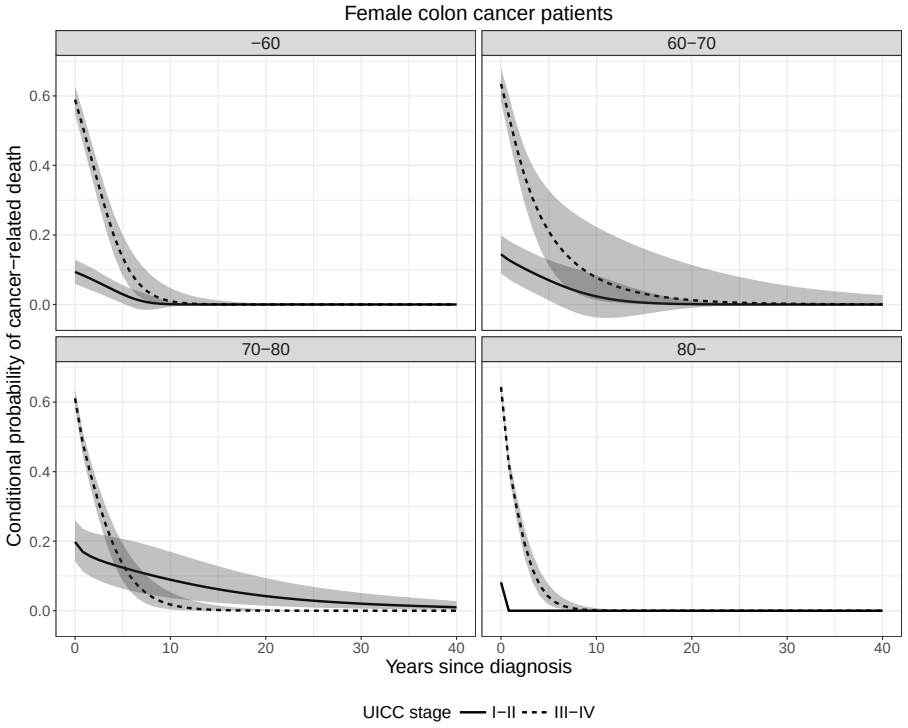

Figure S2: The conditional probability of cancer-related death in Danish female colon cancer patients stratified on age group (-60, 60-70, 70-80, 80-) and stage (UICC stage I-II vs III-IV). UICC: Union for International Cancer Control.

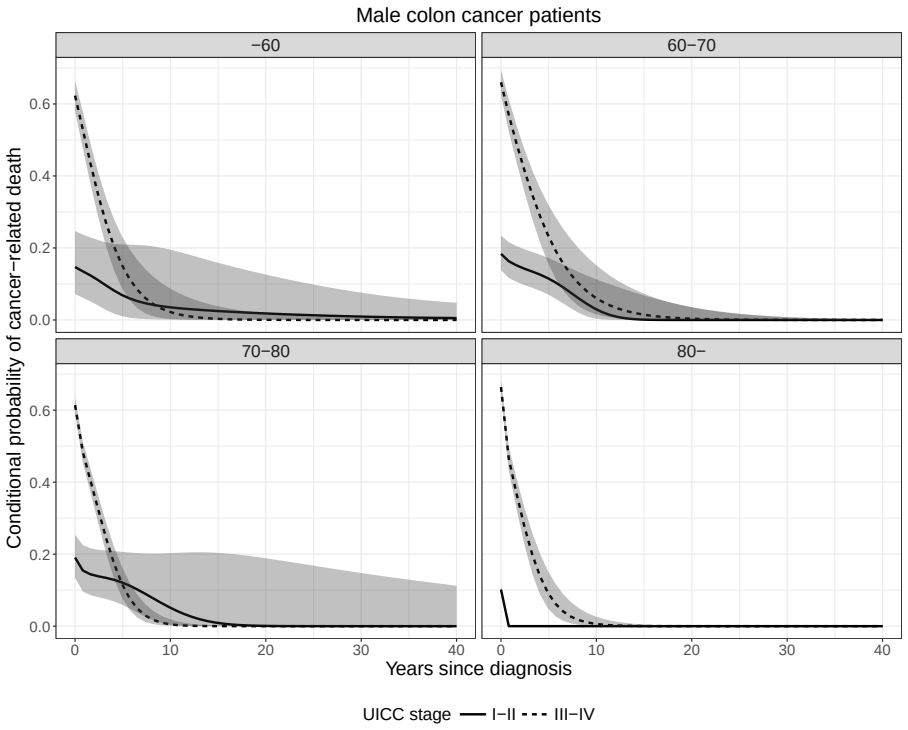

Figure S3: The conditional probability of cancer-related death in Danish male colon cancer patients stratified on age group (-60, 60-70, 70-80, 80-) and stage (UICC stage I-II vs III-IV). UICC: Union for International Cancer Control.

Paper IV

Generalized parametric cure models

Lasse H. Jakobsen[1,2], Martin Bøgsted[1,2], and Mark Clements[3]

1. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; 2. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; 3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

In preparation.

Description
In this paper, we introduced a general framework for estimating parametric cure models. Cure models are particular useful in situations, where some individuals will never experience the event of interest, or in situations where the observed survival becomes the same as the survival of the general population. These two scenarios are indicated by a plateau in the survival function and relative survival function, respectively. The primary entity of interest in cure models is the proportion of the individuals who are cured. In a relative survival context cured individuals have the same mortality as the general population. The aim of this study was to i) develop a framework for estimating cure models, which allows for a wide range of models specifications, and ii) assess the performance of cure models, which incorporate the assumption of a cured proportion in different ways.

Abstract

Cure models are used in time-to-event analyses in cases where not all individuals are expected to experience the event of interest. Cure models have a long history in the statistical literature and during the last two decades, cure models for the relative survival have also been discussed. The main parameter of interest in cure models is the proportion of individuals who are cured, termed the cure proportion, but also the survival function of the uncured individuals may be estimated. Within relative survival, the cure proportion denotes the fraction of individuals who experience the same survival function as the general population. These entities can be estimated from two general classes of models, namely explicit and latent cure models, which deviate in their inclusion of explicit parameters for the cure proportion.
In this article, we introduce a general parametric formulation of both explicit and latent cure models together with a general estimation framework and software, which enable fitting of a wide range of different models. Through simulations, we assess the estimation accuracy of both explicit and latent cure models with respect to the cure proportion and the survival of the uncured individuals. Finally, we illustrate the models on colon cancer data and data on diffuse large B-cell lymphoma patients, for whom cause of death information was available.
As demonstrated in the simulations, explicit cure models which are not guaranteed to be constant after a finite time point tend to produce accurate estimates of the cure proportion and the survival of the uncured. However, these models are very unstable in certain cases, whereas latent cure models generally provide stable results at the price of more biased estimates.

1 Introduction

In time-to-event analyses, it is typically assumed that all included individuals are susceptible to the event of interest. However, in some scenarios it is reasonable to assume that a fraction of the individuals will never experience the event of interest. These individuals could be considered *cured* of the event and the observed population would be a mixture of cured and uncured individuals. For instance, a proportion of first-time mothers will never give birth to a second child, and thus the survival function corresponding to the time until birth of the second child reaches a plateau after a number of years [1]. Cure models, which have a long history in the statistical literature, are particularly useful for analyzing time-to-event data, which exhibits such patterns [2]. The two main entities of interest in cure models are the proportion of cured individuals and the time-to-event survival function of the uncured individuals.

While cure models have been employed for analyzing the time to death for cancer patients with the aim of estimating the proportion of long-term survivors [3–5], this may not always be appropriate as especially elderly patients can also die from other causes than the cancer. Instead the proportion of the individuals who have the same survival as the general population may be estimated by using a cure model formulation for the relative survival function [6–8]. Individuals with the same mortality as the general population are said to be *statistically cured* and the presence of such individuals is indicated by a plateau in the relative survival function given that the follow-up is sufficiently long. Relative survival has previously been used to quantify the net survival, i.e., the survival of patients with a specific disease in the scenario where they can only die from the considered disease [9]. Perme et al. gave a comprehensive discussion on the relation between net and relative survival [10].

The Cox proportional hazards model remains the most popular model in medical research even though parametric survival models have proven useful for many applications. Since standard parametric distributions are typically too restrictive to provide a sufficiently good fit, developing more flexible alternatives is an active research area, best exemplified by the spline-based Royston-Parmar model [11]. The Royston-Parmar model was recently generalized by Liu et al. by using a link-based approach that allows for a wide range of smoothers and functional forms [12].

Non-parametric cure models for the all-cause survival have previously been established by using the Kaplan-Meier estimator and, more recently, by utilizing an expectation-maximization scheme [4, 13]. However, non-parametric cure models for the relative survival function have not previously been developed. Parametric cure models are often based on maximum likelihood estimation and are therefore easily extendable to relative survival [6, 7]. The Royston-Parmar model has previously also been used to model the relative survival function, which only requires minor changes to the fitting procedure due to the parametric formulation of the model [14]. Andersson et al. suggested a cure model version of this model by forcing the splines of the Royston-Parmar model to be constant after the last knot [15].

In this article we introduce a general link-based formulation of two classes of parametric cure models, namely *explicit* and *latent* cure models. While explicit cure models contain explicit parameters for the cure proportion, latent cure models introduce a cure proportion by incorporating a plateau in the relative survival function similarly to Andersson et al. [15]. We introduce a general modelling framework using maximum likelihood, which enables various time and covariate effects for the cure proportion and the survival of the uncured individuals. For establishing the latent cure models, we introduce a methodology for restricting the trajectory of polynomial splines to be constant beyond their boundary knots. In a simulation study, we will examine

the ability of latent and explicit cure models to estimate the cure proportion and corresponding covariate effects. Finally, we analyze Danish colon cancer data, which often display statistical cure, and Danish lymphoma data with available cause of death information. Without loss of generality, we focus on parametric cure models for the relative survival function.

2 Methods

Assume we observe i.i.d. right-censored event times, i.e., triplets $(T_i, \delta_i, \boldsymbol{z}_i)$ for individuals $i = 1, ..., n$, where $T_i = \min(X_i, C_i)$, X_i is the time to event, C_i is the censoring time, $\delta_i = \mathbf{1}[X_i \leq C_i]$ is the event indicator, and $\boldsymbol{z}_i$ is a covariate vector for the i^{th} individual. We denote by $S(t|\boldsymbol{z}) = 1 - F(t|\boldsymbol{z}) = \mathrm{P}(X > t|\boldsymbol{z})$ the survival function associated with X, while $H(t|\boldsymbol{z}) = -\log[S(t|\boldsymbol{z})]$ is the cumulative hazard and $h(t|\boldsymbol{z}) = dH(t|\boldsymbol{z})/dt$ is the corresponding hazard function.

We denote by $S^*(t|\boldsymbol{z})$ the general population survival function and define the relative survival function as

$$R(t|\boldsymbol{z}) = \frac{S(t|\boldsymbol{z})}{S^*(t|\boldsymbol{z})}.$$

By using the relation between the hazard and the survival function, the hazard function can be written as

$$h(t|\boldsymbol{z}) = h^*(t|\boldsymbol{z}) + \lambda(t|\boldsymbol{z}),$$

where $h^*(t|\boldsymbol{z})$ is the general population hazard, and $\lambda(t|\boldsymbol{z})$ is termed the *excess hazard* function or *excess mortality*. The functions $h^*(t|\boldsymbol{z})$ and $S^*(t|\boldsymbol{z})$ can be computed from publicly available life tables which contain mortality rates stratified on a number of demographic variables such as age, sex, and calendar year. The formulation of the hazard function directly leads to

$$H(t|\boldsymbol{z}) = H^*(t|\boldsymbol{z}) + \Lambda(t|\boldsymbol{z}),$$

where $\Lambda(t|\boldsymbol{z}) = \int_0^t \lambda(u|\boldsymbol{z})du$ is termed the cumulative excess hazard.

In cure models, it is assumed that some individuals are cured, i.e., have a constant relative survival of 1 (see Section S1 for the connection between cure and the relative survival function). Let Y_i be the unobserved random variable denoting whether the i^{th} individual is cured ($Y_i = 1$) or uncured ($Y_i = 0$) and assume that the Y_is are independent. The primary parameter of interest is $\pi(\boldsymbol{z}) = \mathrm{E}[Y_i|\boldsymbol{z}] = \mathrm{P}(Y_i = 1|\boldsymbol{z})$, i.e., the probability of being cured, also known as the cure proportion.

In some situations, the time point at which there is no mortality difference

between the considered individuals and the general population, is also of interest. This time point is known as the *cure point* and inspired by Rabinowitz and Ryan [16], we define it as

$$t_c = \inf \{t | \forall s \geq t \text{ and } \forall \boldsymbol{z}, \lambda(s|\boldsymbol{z}) = h(s|\boldsymbol{z}) - h^*(s|\boldsymbol{z}) = 0\}.$$

This definition implies that the relative survival is constant for all $t \geq t_c$, which is an important feature of latent cure models as described in Section 2.2.

2.1 Explicit cure models

We define explicit cure models as models with explicit parameters for the cure proportion, π. In cure model literature, this class of cure models is commonly divided into mixture and non-mixture cure models [7, 17].

Mixture cure models

In the mixture cure model [2], the relative survival is specified by

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})] S_u(t|\boldsymbol{z}), \tag{1}$$

where $S_u(t|\boldsymbol{z})$ is the relative survival of the uncured individuals. That is, a proportion of the considered individuals, $\pi(\boldsymbol{z})$, have the same survival as the general population, while the remaining individuals have a worse survival described by $S^*(t|\boldsymbol{z}) S_u(t|\boldsymbol{z})$.

We introduce a new formulation of the mixture cure model by using a bijective function $g : (0, 1) \to \mathbb{R}$ for S_u, similarly to Liu et al. and Younes and Lachin [12, 18]. By introducing an additional link function, $g_\pi : (0, 1) \to \mathbb{R}$, for π, the mixture cure model is formulated by

$$g(S_u(t|\boldsymbol{z})) = \eta(t, \boldsymbol{z}; \boldsymbol{\theta}) = X(t, \boldsymbol{z})\boldsymbol{\theta} \tag{2}$$

and

$$g_\pi(\pi(\boldsymbol{z})) = \eta_\pi(\boldsymbol{z}; \boldsymbol{\beta}) = X_\pi(\boldsymbol{z})\boldsymbol{\beta}, \tag{3}$$

where $X(t, \boldsymbol{z})$ and $X_\pi(\boldsymbol{z})$ are design matrices for the survival of the uncured and the cure proportion, respectively. Thus, S_u is modelled by a time-varying design matrix together with some link function, spanning a wide range of different models. Note, the design matrix associated with the cure proportion is not time-varying. Although the domain of the link functions is often assumed to be [0, 1], the identity function has previously been proposed for the cure proportion [7]. Denoting the inverse link functions by $G = g^{-1}$ and

$G_\pi = g_\pi^{-1}$, the excess hazard is given by

$$\lambda(t|\boldsymbol{z}) = \frac{-\left[1 - G_\pi(\eta_\pi(\boldsymbol{z};\boldsymbol{\beta}))\right] G'(\eta(t,\boldsymbol{z};\boldsymbol{\theta}))}{G_\pi(\eta_\pi(\boldsymbol{z};\boldsymbol{\beta})) + \left[1 - G_\pi(\eta_\pi(\boldsymbol{z};\boldsymbol{\beta}))\right] G(\eta(t,\boldsymbol{z};\boldsymbol{\theta}))} X_D(t,\boldsymbol{z})\boldsymbol{\theta},$$

where

$$X_D(t,\boldsymbol{z}) = \frac{\partial X(t,\boldsymbol{z})}{\partial t}. \tag{4}$$

Generally, (4) can be computed by numerical differentiation. The cumulative hazard is given by

$$\begin{aligned}\Lambda(t|\boldsymbol{z}) &= -\log\left(R(t|\boldsymbol{z})\right) \\ &= -\log\left[G_\pi(\eta_\pi(\boldsymbol{z};\boldsymbol{\beta})) + \left[1 - G_\pi(\eta_\pi(\boldsymbol{z};\boldsymbol{\beta}))\right] G(\eta(t,\boldsymbol{z};\boldsymbol{\theta}))\right].\end{aligned}$$

Simple parametric models for S_u, such as the exponential and Weibull distribution, can be obtained through the linear predictor in (2). A Weibull model without covariates is obtained by choosing $g(x) = \log(-\log(x))$, and $X(t_i, \boldsymbol{z}_i) = [1, \log(t_i)]$. Although these simple models are included in the general model class, our focus will be on more flexible time effects such as splines. Since the time effect is represented by the general linear predictor, $\eta(t,\boldsymbol{z};\boldsymbol{\theta})$, different smoothers can be employed. For the purpose of this article, we focus on polynomial B-splines. These can be transformed into natural cubic splines (NCSs), which are similar to the restricted cubic splines used in the Royston-Parmar model [11]. Also covariate effects may be modelled using smoothers, and time-varying coefficients can be defined by modelling interactions between covariates and a time effect determined by a given smoother. The general linear predictor, $\eta(t,\boldsymbol{z};\boldsymbol{\theta})$, enables the use of smoothers which are not limited to be functions of $\log(t)$ as in the Royston-Parmar model [11]. Due to the numeric differentiation in (4), smoothers that are functions of, e.g., t and $\sqrt{t}$ can readily be applied.

Non-mixture cure models

Non-mixture cure models [7, 17, 19, 20] are specified by

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})},$$

where $\widetilde{F}$ is a proper distribution function with a positive domain. The change in notation from $S_u(\cdot)$ to $\widetilde{F}(\cdot)$ is to emphasize the interpretational differences between these two functions. While $S_u(\cdot)$ is the survival of the uncured individuals, $\widetilde{F}(\cdot)$ cannot be interpreted as the distribution function of the uncured individuals and is only used to model the time effect of the relative

survival. However, a simple transformation of the non-mixture cure model,

$$R(t|\boldsymbol{z}) = \pi(\boldsymbol{z}) + [1 - \pi(\boldsymbol{z})]\frac{\pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})} - \pi(\boldsymbol{z})}{1 - \pi(\boldsymbol{z})},$$

allows for the same interpretation as in mixture cure models. That is, the survival of the uncured can be derived from a non-mixture cure model by computing

$$S_u(t|\boldsymbol{z}) = \frac{\pi(\boldsymbol{z})^{\widetilde{F}(t|\boldsymbol{z})} - \pi(\boldsymbol{z})}{1 - \pi(\boldsymbol{z})}. \tag{5}$$

Thus, the non-mixture cure model deviates from the mixture cure model only in the formulation of S_u in which π is also included. Using the link functions, g and g_π, we model π similarly to (3) and the distribution function by

$$g(1 - \widetilde{F}(t|\boldsymbol{z})) = \eta(t, \boldsymbol{z}; \boldsymbol{\theta}) = X(t, \boldsymbol{z})\boldsymbol{\theta},$$

which leads to

$$\lambda(t|\boldsymbol{z}) = \log\left[G_\pi(\eta_\pi(\boldsymbol{z}; \boldsymbol{\beta}))\right] G'(\eta(t, \boldsymbol{z}; \boldsymbol{\theta}) X_D(t, \boldsymbol{z})\boldsymbol{\theta}$$

and

$$\Lambda(t|\boldsymbol{z}) = -\log\left[G_\pi(\eta_\pi(\boldsymbol{z}; \boldsymbol{\beta}))\right]\left[G(\eta(t, \boldsymbol{z}; \theta)) - 1\right].$$

Again X_D can be computed by numerical differentiation.

2.2 Latent cure models

Instead of modelling the relative survival with a mixture or non-mixture cure model, a model for the relative survival may be formulated without specification of a cure proportion, i.e.,

$$g(R(t|\boldsymbol{z})) = \eta(t, \boldsymbol{z}; \boldsymbol{\theta}) = X(t, \boldsymbol{z})\boldsymbol{\theta}. \tag{6}$$

The cumulative excess hazard is then $\Lambda(t|\boldsymbol{z}) = -\log[G(\eta(t, \boldsymbol{z}; \boldsymbol{\theta}))]$ and the corresponding excess hazard function is

$$\lambda(t|\boldsymbol{z}) = -\frac{G'(\eta(t, \boldsymbol{z}; \boldsymbol{\theta}))}{G(\eta(t, \boldsymbol{z}; \boldsymbol{\theta})))} X_D(t, \boldsymbol{z})\boldsymbol{\theta}.$$

However, additional assumptions are needed in order to estimate the cure proportion from the model in (6). In particular, assuming that $t_c < \infty$, we obtain for a general relative survival function that

$$R(t|\boldsymbol{z}) = R(t_c|\boldsymbol{z}) \quad \text{for all } t \geq t_c. \tag{7}$$

If we further assume an underlying mixture cure model, following the argument in Section S1, the condition in (7) implies that $\pi(\boldsymbol{z}) = R(t_c|\boldsymbol{z})$ and $S_u(t) = 0$ for all $t \geq t_c$. Additionally, the survival of the uncured, S_u, is computed by

$$S_u(t|\boldsymbol{z}) = \frac{R(t|\boldsymbol{z}) - R(t_c|\boldsymbol{z})}{1 - R(t_c|\boldsymbol{z})},$$

which is clearly 0 beyond t_c. Therefore, the cure proportion and the survival of the uncured may be computed post hoc from the estimated model in (6), which does not contain explicit parameters for the cure proportion.

In mixture cure models, restricting S_u to be zero beyond t_c requires additional considerations about the link function, g. The relative survival model in (6) provides an alternative approach for estimating cure proportions that requires R to be constant beyond t_c, which is not challenged by the choice of link function. However, in order to fit the model in (6), the value of t_c has to be chosen. As simple parametric distributions, such as the Weibull model, cannot appropriately be restricted to be constant beyond t_c, these are not appropriate models for R in (6) if t_c is assumed to be finite. The model in (6) is rather intended for smoothers for which the trajectory can be restricted at t_c, such as polynomial splines.

The described cure model is simply a generalization of the model by Andersson et al. [15]. They employed a proportional excess hazards model, corresponding to $G(x) = \exp(-\exp(x))$, with the log-cumulative baseline hazard modelled by restricted cubic splines. The time effects were altered to be constant after the last knot of the splines. Andersson et al. showed that this model was equivalent to a non-mixture cure model and thus contained explicit parameters for the cure proportion [15]. However, this does not hold in general for any link function, g. Similar restrictions have previously been used for establishing non-mixture cure models in a Bayesian framework [21]. In the following, we propose a method for restricting the model in (6) to satisfy the condition in (7) by modelling the time-effects with B-splines.

Cure by QR decomposition of B-splines

Since g is bijective, $R(t)$ is constant beyond t_c if and only if $\eta(t, \boldsymbol{z}; \boldsymbol{\theta})$ is constant beyond t_c. Therefore, the constraint in (7) is equivalent to

$$\eta'(t, \boldsymbol{z}; \boldsymbol{\theta}) = X_D(t, \boldsymbol{z})\boldsymbol{\theta} = 0 \quad \text{for } t \geq t_c \text{ and } \forall \boldsymbol{z}, \tag{8}$$

For now, we consider the case where no time-varying covariate effects are included and use the partition $X(t, \boldsymbol{z}) \equiv [X_1(t) \quad X_2(\boldsymbol{z})]$, where $X_1(t)$ contains the time effects and $X_2(\boldsymbol{z})$ contains the covariate effects. Let the time effect, $X_1(t)$, be modelled by B-splines of order $m + 1$ with K knots $k_1 < \cdots < k_K$. A constant relative survival function beyond $t_c = k_K$ can be obtained by

incorporating a series of linear constraints,

$$B\boldsymbol{\theta}_1 = 0,$$

where $\boldsymbol{\theta}_1$ is the p_1 model coefficients corresponding to the time effect and

$$B_{ij} = \left[X_1^{(i)}(t_c)\right]_j \quad \text{for } i = 1, ..., m$$

with $X_1^{(i)}(t_c)$ being a vector containing the i^{th} derivative of $X_1(t_c)$ with respect to t. For cubic splines, it is therefore sufficient to have the first and second order derivative of the base functions being equal to zero at t_c. The linear constraints can be absorbed into the model fitting procedure by using the QR decomposition of B^T [22]. We may write

$$B^T = Q \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where Q is a $p_1 \times p_1$ orthogonal matrix and R is an $m \times m$ upper triangular matrix. Using the partitioning $Q \equiv [Q_1 \quad Q_2]$, where Q_2 is a $p_1 \times (p_1 - m)$ matrix, parameter estimates satisfying (8) can be obtained by a transformation of $X_1(t)$. In particular, the linear predictor $\widetilde{X}(t, z)\widetilde{\boldsymbol{\theta}} \equiv [X_1(t)Q_2 \quad X_2(z)]\,\widetilde{\boldsymbol{\theta}}$ satisfies the constraints in (8). Note that $X_1(t)Q_2$ is an $n \times (p_1 - m)$ matrix and thus $\widetilde{\boldsymbol{\theta}}$ contains m fewer entries than $\boldsymbol{\theta}$.

For NCSs, the trajectory before k_1 is restricted to be linear and so an additional constraint, $X_1^{(2)}(k_1) = 0$, needs to be added. For models with time-varying covariate effects, the QR-decomposition approach can be applied to each time effect separately.

2.3 Estimation

Let $\boldsymbol{\Theta}$ be all parameters of the cure model. Then the log-likelihood function is written as

$$\ell(\boldsymbol{\Theta}) = \sum_{i=1}^{n} \delta_i \log\left[h^*(t_i|\boldsymbol{z}_i) + \lambda(t_i|\boldsymbol{z}_i)\right] - \Lambda(t_i|\boldsymbol{z}_i). \tag{9}$$

In parametric relative survival models there are usually no restrictions on the excess hazard function [14], but a non-negativity constraint is applied to the all-cause hazard, i.e., $h(t|\boldsymbol{z}_i) = h^*(t|\boldsymbol{z}_i) + \lambda(t|\boldsymbol{z}_i)$. Similarly to Liu et al. [12], we subtract a penalty term from the log-likelihood function to ensure a

non-negative hazard,

$$\ell_P(\boldsymbol{\Theta}) = \ell(\boldsymbol{\Theta}) - \frac{\kappa}{2}\sum_{i=1}^{n} h(u_i|\boldsymbol{z}_i)^2 \mathbf{1}[h(u_i|\boldsymbol{z}_i) < 0]. \tag{10}$$

Initially, the likelihood is maximized with $\kappa = 1$. If one or more hazard values are negative, the likelihood is re-maximized after doubling κ until all hazards are non-negative. Negative hazard values in (9) are replaced by some small value, $\epsilon \approx 10^{-16}$. For the mixture cure models, the survival of the uncured individuals, $S_u(t)$, may be considered a proper survival function. Therefore, for mixture cure models, $h(\cdot)$ in the penalty term in (10) is replaced by $h_u(\cdot) = -d/dt \log[S_u(t|\boldsymbol{z})]$. Similar restrictions are employed for the non-mixture cure models, where $\widetilde{S}(t) = 1 - \widetilde{F}(t)$ is a proper survival function, and $h(\cdot)$ is replaced by $\widetilde{h}(\cdot) = -d/dt \log[1 - \widetilde{F}(t|\boldsymbol{z})]$. Note, this restriction automatically implies a non-negativity excess hazard.

Initial values

For latent cure models, initial values were found by fitting a Cox proportional hazards model including all covariates, but excluding time-varying covariate effects. The g-transformed predicted survival probabilities from the Cox model, $g\left[\hat{S}(t_i|\boldsymbol{z}_i)\right]$, were then used as response in a linear model with $X(t_i, \boldsymbol{z}_i)$ as design matrix. The obtained parameter values were used as initial values for optimizing (9). To adjust the initial values to relative survival, the predicted survival probabilities were normalized by $\hat{S}^*(t_i|\boldsymbol{z}_i) = \exp\left(-h^*(t_i|\boldsymbol{z}_i)t_i\right)$, assuming a constant general population hazard.

For the explicit cure models, two different approaches were used to obtain initial values. The likelihood was maximized using both sets of initial parameters and the best model was selected as the model resulting in the largest likelihood. Since mixture and non-mixture cure models are related through (5), we only describe the two approaches in terms of the mixture cure model.

The first approach finds initial values by fitting a parametric mixture cure model following the implementation by Lambert [23], with a logistic link function for π and a Weibull model for S_u excluding time-varying covariate effects. The predicted cure probabilities, $\hat{\boldsymbol{y}}$, were used as response in a generalized linear model with X_π as design matrix. For S_u, a generalized linear model is fitted with $\hat{S}_u(t_i|\boldsymbol{z}_i)$ as response and $X(t_i, \boldsymbol{z}_i)$ as design matrix.

The second approach fits the relative survival model in (6) with a log-log link function (a proportional excess hazards model) and the baseline time effect modelled by an NCS without restricting it to be constant beyond the last knot. The estimates $\hat{\boldsymbol{y}}$ are obtained by calculating $\hat{R}(\tau + c|\boldsymbol{z}_i)$ for $i = 1, ..., n$, where τ is the largest follow-up time and c is a small constant (selected to be 0.1). The procedure from the previous approach is repeated to obtain initial

values for π. For S_u, the initial values are found by fitting the generalized linear model,

$$g\left(\frac{\hat{R}(t_i|\boldsymbol{z}_i)-\hat{y}_i}{1-\hat{y}_i}\right)=X(t_i,\boldsymbol{z}_i)\boldsymbol{\theta}.$$

2.4 Identifiability

Identifibability is a key issue in cure models. A cure model is said to be identifiable if $R(t|\boldsymbol{z},\boldsymbol{\Theta}_1)=R(t|\boldsymbol{z},\boldsymbol{\Theta}_2)$ for $0\leq t<\tau$ implies that $\boldsymbol{\Theta}_1=\boldsymbol{\Theta}_2$.

Following the argument of Hanin and Huang, mixture cure models, non-mixture cure models, and latent cure models are identifiable if $S_u(\tau|\boldsymbol{z})=0$ for all $\boldsymbol{z}$ [24]. This can be ensured by having $\tau=\infty$ if S_u $(\widetilde{F})$ is proper in $\mathbb{R}^+$ or by assuming $t_c\leq\tau$. In these cases, the cure model is identifiable regardless of the model formulation for π [25]. However, these conditions are not necessarily satisfied in either explicit or latent cure models. We refer the reader to Hanin and Huang for further identifiability results [24].

2.5 Useful summary measures

The primary entities of interest in cure models are the cure proportion and the survival of the uncured. However, other useful summary measures can be computed from cure models. From both explicit and latent cure models the conditional probability of belonging to the cured group given survival beyond time t can be obtained [8, 17]. It is given as

$$\mathrm{P}(Y=1|T>t,\boldsymbol{z})=\frac{\mathrm{P}(T>t|Y=1,\boldsymbol{z})\mathrm{P}(Y=1|\boldsymbol{z})}{\mathrm{P}(T>t|\boldsymbol{z})}=\frac{\pi(\boldsymbol{z})}{R(t|\boldsymbol{z})}.$$

The denominator is changed according to the chosen cure model and the numerator changes according to the specified model for π. Generally, if $R(t|\boldsymbol{z})$ is non-increasing, $\mathrm{P}(Y=1|T>t,\boldsymbol{z})\to 1$ as $t\to\infty$, but if $t_c<\infty$, we have $\mathrm{P}(Y=1|T>t_c,\boldsymbol{z})=1$.

Disease-specific mortality measures

By using that the excess hazard can be interpreted as a cause-specific hazard associated with a given disease, crude cumulative incidences of death from the disease can be obtained from relative survival models as demonstrated by Lambert et al. [26]. For this, we let $E_i\in\{D,O\}$, denote the unknown eventual cause of death for the i^{th} individual, where D and O indicate disease-related death and death due to other causes, respectively. The crude probability of disease-related death is

$$\mathrm{P}(T\leq t,E=D|\boldsymbol{z})=\int_0^t S^*(u|\boldsymbol{z})R(u|\boldsymbol{z})\lambda(u|\boldsymbol{z})du, \tag{11}$$

and the crude probability of death from other causes than the disease can be obtained by replacing $\lambda(\cdot)$ with $h^*(\cdot)$, i.e.,

$$\mathrm{P}(T \leq t, E = O|\boldsymbol{z}) = \int_0^t S^*(u|\boldsymbol{z})R(u|\boldsymbol{z})h^*(u|\boldsymbol{z})du. \tag{12}$$

The probability of eventually dying from the disease, $\mathrm{P}(E = D|\boldsymbol{z}) = \mathrm{P}(T \leq \infty, E = D|\boldsymbol{z})$, provides a measure of the entire disease burden. By using this probability and (11), Eloranta et al. [27] proposed

$$\mathrm{P}(E = D|T > t, \boldsymbol{z}) = \frac{\mathrm{P}(E = D, T > t|\boldsymbol{z})}{\mathrm{P}(T > t|\boldsymbol{z})} = \int_t^\infty \frac{S^*(u|\boldsymbol{z})R(u|\boldsymbol{z})}{S^*(t|\boldsymbol{z})R(t|\boldsymbol{z})}\lambda(u|\boldsymbol{z})du$$

as an interpretable tool for risk communication. This yields the conditional probability of eventually dying from the disease given survival until time t. Without $t_c \leq \tau$, extrapolation is required in order to compute $\mathrm{P}(E = D|\boldsymbol{z})$. The parametric formulation of $R(\cdot)$ and $\lambda(\cdot)$ enables extrapolation while assumptions about the mortality rate in the future general population is required to extrapolate $S^*(\cdot)$ [28].

Cure models also enable estimation of the expected lifetime and loss of lifetime. The expected residual lifetime is given by

$$\mathrm{E}\left[T - t|T > t, \boldsymbol{z}\right] = \int_t^\infty \frac{S(u|\boldsymbol{z})}{S(t|\boldsymbol{z})du} = \int_t^\infty \frac{S^*(u|\boldsymbol{z})R(u|\boldsymbol{z})}{S^*(t|\boldsymbol{z})R(t|\boldsymbol{z})}du.$$

We define the loss of lifetime function as the difference between the expected residual lifetime of the considered individuals and the general population,

$$L(t|\boldsymbol{z}) = \int_t^\infty \frac{S^*(u|\boldsymbol{z})}{S^*(t|\boldsymbol{z})}du - \int_t^\infty \frac{S^*(u|\boldsymbol{z})R(u|\boldsymbol{z})}{S^*(t|\boldsymbol{z})R(t|\boldsymbol{z})}du. \tag{13}$$

This provides the number of years lost due to the disease given survival until time t. At $t = 0$, (13) provides the mean number of years lost due to the diagnosis, which was termed the loss in expectation of life by Andersson et al. [28]. Jakobsen et al. [29] investigated the accuracy of the loss of lifetime function using both explicit and latent cure models as well as the relative survival model of Nelson et al. [14]. They concluded that sensitivity analyses should be conducted when computing the loss of lifetime function since the considered models gave volatile results.

2.6 Implementation

The explicit cure models are implemented in the `R`-package `cuRe` (see `https://github.com/LasseHjort/cuRe`). The latent cure models are implemented in the `R` package `rstpm2` which is available from the Comprehensive `R` Archive

Network (CRAN). Both packages enable fitting of all-cause as well as relative survival cure models by specifying an argument for the general population hazard, $h^*(t_i|\boldsymbol{z}_i)$. All models are implemented for both right-censored and left-truncated time-to-event data and for a series of commonly used link functions. For the explicit cure models, an argument is available for disabling the penalization approach used to obtain proper estimates of S_u and $\widetilde{S}$.

All post-estimation procedures described in Section 2.5 are implemented in the `cuRe` package. All integrals are computed by Gauss-Legendre quadrature and the variances are computed by the delta method using numerical differentiation. For probability measures, appropriate link functions were applied to restrict the confidence interval to be within $[0, 1]$. The `rstpm2` package enables fitting of latent cure models to correlated time-to-event data using random effects which are estimated using maximum marginal likelihood (see Liu et al. for details [30]).

3 Cure proportion estimation

As explicit and latent cure models rely on different model assumptions, these may provide different estimates of the cure proportion, $\pi(\boldsymbol{z}) = \mathrm{E}[Y|\boldsymbol{z}]$, and the survival of the uncured, $S_u(t|\boldsymbol{z})$. In the following section, we evaluate the performance of several instances of both model types in a series of simulations with and without covariate effects.

3.1 Data simulation

Data were simulated from various relative survival models assuming that the time-to-event variable X associated with the relative survival, R, and the general population survival time, X^*, are independent. Along the lines of Rutherford et al. [31], we used the following scheme to generate the follow-up times:

1. Generate survival time, X, from the potentially improper $R(t|\boldsymbol{z})$.
2. Generate survival time, X^*, from $S^*(t|\boldsymbol{z})$.
3. Generate censoring time, C.
4. The follow-up time is $T = \min(X, X^*, C)$ and $\delta = \mathbf{1}[\min(X, X^*) \leq C]$.

For a specific $R(t|\boldsymbol{z})$, the event time was simulated as the root of the equation $R(t|\boldsymbol{z}) = U$, where U is uniformly distributed between 0 and 1 (a similar approach was used for $S^*(t)$) [32]. In cases where no root existed, the follow-up time was set to ∞. This may occur if a cure fraction is assumed for $R(t|\boldsymbol{z})$. A uniform distribution, Uniform(0, 15), was chosen as censoring distribution

in all simulations. The general population survival function, $S^*(t|z)$, was computed for Danish females aged 60 in 1980 using the Ederer I method [33] and a Danish life table retrieved from the Human Mortality Database [34]. In each simulation, the sample size was 1000 and the simulations were repeated 500 times.

Without covariates

For $R(t|z)$ we considered six scenarios described by a mixture cure model with varying cure proportions. The cure proportion in each scenario was 0.25, 0.25, 0.5, 0.75, 0.75, and 0, and the survival of the uncured was given by a Weibull model. The relative survival trajectories can be found in Figure S1. The trajectories were chosen to assess the cure proportion sensitivity in cases where cure occurs within the follow-up (scenario 1 and 4), just at the end of the follow-up (scenario 3), and beyond the follow-up (scenario 2, 5, and 6). To increase the complexity of the survival of the uncured, the simulations were repeated using polynomial splines instead of a Weibull model for S_u (see Supplementary S2 for details).

With covariates

To assess the ability of the cure models to capture covariate effects, we introduced a binary covariate, z, simulated from a Bernoulli distribution with $p = 0.5$. A covariate effect for both the cure proportion and the survival of the uncured was included by

$$\eta_\pi(z;\boldsymbol{\beta}) = \beta_0 + z\beta_1 \qquad \text{and} \qquad \eta(z,t;\boldsymbol{\theta}) = \theta_0 + z\theta_1 + \theta_2 \log(t).$$

The coefficients and link functions, which were used for the simulations, are shown in Table 1. Again, the survival function of the uncured was given by a Weibull model, which was obtained by using the link function $g(x) = \log(-\log(x))$.

Scenario	π	β_0	β_1	θ_0	θ_1	θ_2	g	g_π	LOR
1	0.25	-1.10	0.50	0.00	0.00	1.00	$\log[-\log(x)]$	$\log\left(\frac{x}{1-x}\right)$	0.50
2	0.50	-0.37	0.50	-2.30	0.50	1.40	$\log[-\log(x)]$	$\log[-\log(x)]$	-0.76
3	0.60	0.40	0.50	-0.69	0.50	1.00	$\log[-\log(x)]$	$\log\left(\frac{x}{1-x}\right)$	0.50

Table 1: Parameter values used for simulations including covariate effects. LOR: true log-odds ratio for the cure proportion.

3.2 Simulation results

Without covariates

For the simulations without covariates we considered five models; a (Weibull) mixture cure model, two explicit cure models and two latent cure models using NCSs with different knot placement (see Table 2 for details).

Model	Model	g_π	g	Smoother	Knot locations
A	Weibull mixture CM	$\log\left(\frac{x}{1-x}\right)$	$\log(-\log(x))$		
B	Explicit mixture CM	$\log\left(\frac{x}{1-x}\right)$	$\log(-\log(x))$	NCS	0%, 25%, 50%, 75%, and 100% quantiles of the uncensored event times.
C	Explicit mixture CM	$\log\left(\frac{x}{1-x}\right)$	$\log(-\log(x))$	NCS	Smallest uncensored event time, 0.5, 1, 3, and 5 years.
D	Latent CM		$\log(-\log(x))$	NCS	0%, 25%, 50%, 75%, 95%, and 100% quantiles of the uncensored event times.
E	Latent CM		$\log(-\log(x))$	NCS	0%, 25%, 50%, and 75% quantiles of the uncensored event times and additional knots at 8 and 30 years.

Table 2: Specification of the models used for estimating the cure proportion in simulations without covariates. CM: cure model, NCS: natural cubic spline.

In scenarios where statistical cure is clearly achieved within the follow-up (1, 3, and 4) all models performed fairly well (Figure 1). In scenarios where statistically cure was not reached within the follow-up (2, 5, and 6), the explicit cure models displayed very dispersed cure proportions, which is explained by identifiability issues. In these scenarios, the same relative survival trajectory may be obtained by several sets of parameter values. Notably, the dispersion was not heavily influenced by the flexibility of the explicit cure models as determined by the number of knots in the NCS (see Figure S2), but it was less pronounced for the Weibull model (A).

The latent cure models, on the other hand, gave stable cure proportion estimates in all scenarios, but due to the strict assumption of cure at a specific time point, these were overestimated in scenarios where statistical cure is not reached within the follow-up. Even though identifiability is not guaranteed for model E, because the last knot is placed beyond the available follow-up, the corresponding estimates were stable and generally slightly lower than those of model D.

With covariates

We fitted five models similar to those in Table 2, but now including an effect for the binary covariate, z. In particular, for the mixture cure models, the covariate was added both for the cure proportion and the survival of the uncured, and for the latent cure models it was added both in a proportional hazards and proportional odds model (see details in Table 3). To equalize the

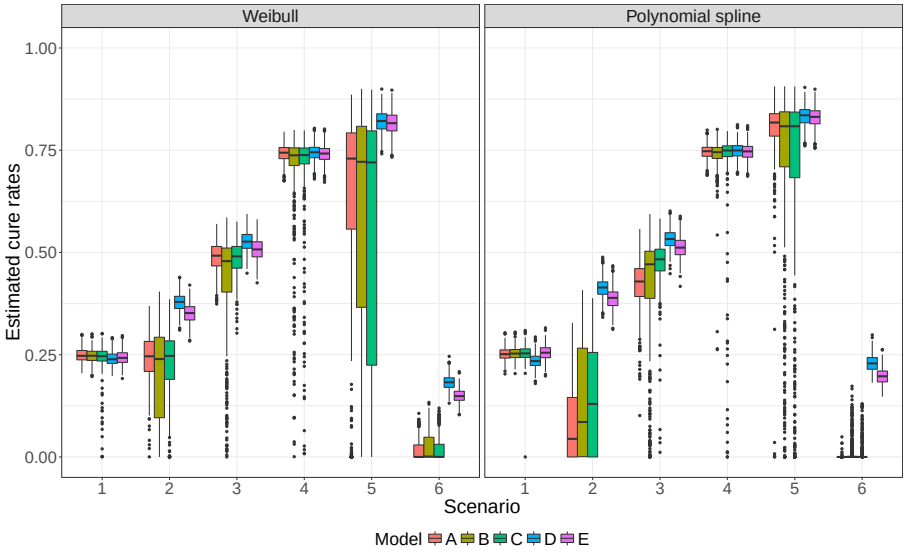

Figure 1: Estimated cure proportions from the models listed in Table 2 in simulations of the scenarios in Figure S1.

number of parameters, z was modelled as a time-varying effect in the latter two models using an NCS with two knots.

Model	Model	g_π	g	Smoother	Knot locations
A	Weibull mixture CM	$\log\left(\frac{x}{1-x}\right)$	$\log(-\log(x))$		
B	Explicit mixture CM	$\log\left(\frac{x}{1-x}\right)$	$\log(-\log(x))$	NCS	0%, 25%, 50%, 75%, and 100% quantiles of the uncensored event times.
C	Explicit mixture CM	$\log(-\log(x))$	$\log(-\log(x))$	NCS	0%, 25%, 50%, 75%, and 100% quantiles of the uncensored event times.
D	Latent CM		$\log(-\log(x))$	NCS	0%, 20%, 40%, 60%, 80%, 95% and 100% quantiles of the uncensored event times.
E	Latent CM		$-\log\left(\frac{x}{1-x}\right)$	NCS	0%, 20%, 40%, 60%, 80%, 95% and 100% quantiles of the uncensored event times.

Table 3: Specification of covariate dependent models assessed in the simulation study. CM: cure model, NCS: natural cubic spline.

For the cure proportion, the effect of the covariate was measured by the log-odds ratio (LOR) which was obtained by

$$\text{LOR} = \log\left(\frac{\hat{\pi}(z=1)/(1-\hat{\pi}(z=1)}{\hat{\pi}(z=0)/(1-\hat{\pi}(z=0)}\right).$$

The true LOR can be found in Table 1. The variance of the estimated LOR was computed by using the delta method and numerical differentiation. The mean absolute error (AE) was used to measure the accuracy of the covariate-

specific cure proportion. The AE was computed as

$$\text{AE}(\hat{\pi}, \pi) = \frac{1}{2} \sum_{z=0}^{1} |\hat{\pi}(z) - \pi(z)|,$$

where $\pi(z)$ is the true cure proportion. The integrated mean AE (IAE) was used to measure the accuracy of the survival of the uncured and the entire relative survival function. It was computed by

$$\text{IAE}(\hat{S}_u, S_u) = \int_0^{15} \frac{1}{2} \sum_{z=0}^{1} |\hat{S}_u(t|z) - S_u(t|z)| dt, \tag{14}$$

where S_u is the true survival of the uncured individuals. The IAE for the relative survival was obtained by replacing S_u with R in (14). The integral was computed by Gauss-Legendre quadrature.

The empirical mean LOR, median variance, AE, and IAEs of each model are shown in Table 4. In the first scenario, the cure proportion AE was low for all models, but model E showed the worst empirical mean LOR. In the second scenario, the models were more biased, with the latent cure models (D and E) underestimating the LOR and the spline-based explicit cure models (B and C) overestimating the LOR. The AE of the cure proportion was smaller for model D and E compared to model B and C. The same pattern was seen in scenario 3. The AEs of the cure proportion obtained from the reference model (A) and the latent cure models (model D and E) were low in all scenarios.

In general, differences between the two latent cure models were minimal, but larger differences were observed between the latent and explicit cure point models. In some simulations, the variance of the LOR could not be computed or was negative because the covariance matrix was not positive semidefinite, which stems from identifiability issues. However, the median variances of model A, B and C in scenario 2 and 3 were only slightly elevated in comparison with model D and E.

In scenario 1, the IAEs of S_u were fairly similar between the five models, whereas in scenario 2 and 3, the IAE of S_u was lower for model D and E compared to B and C. Notably, the relative survival IAE of each model was similar in all scenarios despite the differences in the estimation of the cure proportion and survival of the uncured. This highlights the identifiability issues related to the considered explicit cure models.

Model	LOR	Var(LOR)	AE($\hat{\pi}(z), \pi$)	IAE($\hat{S}_u, S_u$)	IAE($\hat{R}, R$)
Scenario 1 (LOR = 0.50)					
A	0.51	0.02	0.018	0.060	0.254
B	0.51	0.03	0.019	0.092	0.260
C	0.51	0.03	0.019	0.090	0.260
D	0.47	0.02	0.019	0.107	0.264
E	0.39	0.02	0.020	0.114	0.286
Scenario 2 (LOR = -0.76)					
A	-0.76	0.05	0.030	0.307	0.298
B	-0.61	0.05	0.051	0.631	0.304
C	-0.52	0.05	0.051	0.622	0.304
D	-1.01	0.03	0.037	0.528	0.333
E	-0.95	0.02	0.032	0.468	0.301
Scenario 3 (LOR = 0.50)					
A	0.51	0.03	0.023	0.185	0.290
B	0.62	0.03	0.039	0.626	0.298
C	0.69	0.03	0.041	0.642	0.297
D	0.37	0.02	0.025	0.293	0.312
E	0.33	0.02	0.026	0.317	0.325

Table 4: Emperical mean LOR, median variance, AE, and IAE of the five models in Table eftab:modelscov. The IAE was calculated both for the survival of the uncured and the entire relative survival function. LOR: log-odds ratio, AE: absolute error, IAE: integrated absolute error.

4 Real life data examples

4.1 Colon cancer data

Data on colon cancer survival were retrieved from the Danish Colorectal Cancer Group Database [35]. The patients were diagnosed in the period 2000-2016 and follow-up was measured from the date of diagnosis until death or censoring (July 2017). We restricted the analysis to adult patients, i.e., patients $\geq$18 years at diagnosis. The age-, gender-, and calendar year-matched general population hazard rates were obtained from the Human Mortality Database [34]. The use of the data was approved by the Danish Data Protection Agency (2008-58-0028).

Model coefficients with and without the assumption of cure

In total 38,082 patients were analyzed. Stratified on diagnostic age (-55, 55-65, 65-75, and 75-), we fitted two parametric relative survival models of the form $\log[-\log(R(t))] = s_0(x)$, where $x = \log(t)$ and $s_0(x)$ is an NCS with 8 degrees of freedom. These models were fitted using the `R`-function `stpm2`. One model was fitted using the QR-decomposition approach described in

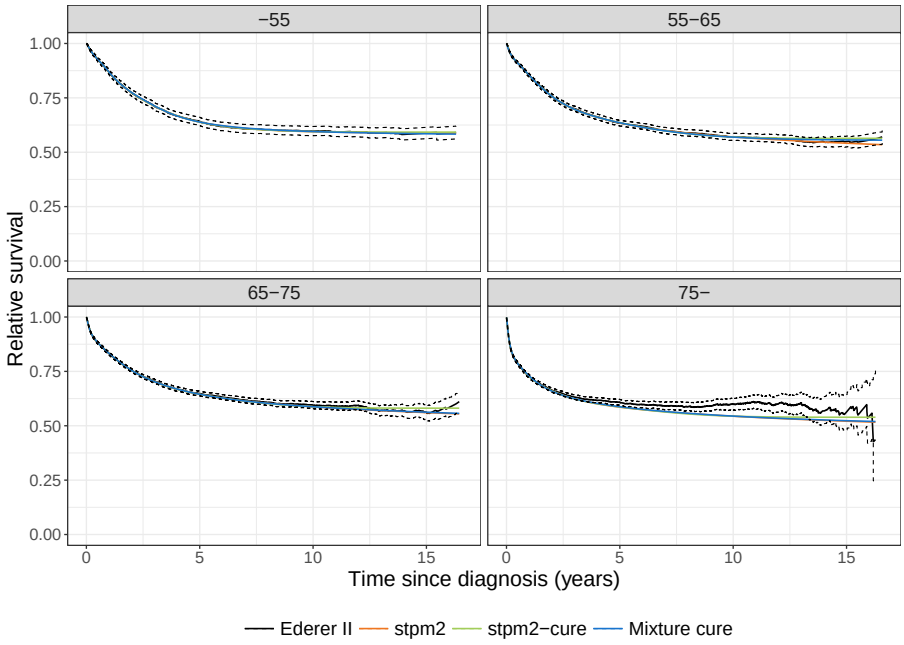

Figure 2: The relative survival of each age group in the colon cancer data estimated by the Ederer II estimator, an NCS-based relative survival model (stpm2), an NCS-based latent cure model (stpm2-cure), and an NCS-based mixture cure model. NCS: natural cubic spline.

Section 2.2, while no restrictions were forced upon the other. We term these, the stpm2-cure and stpm2 model, respectively. Additionally, a mixture cure model was fitted with $\log[-\log(S_u)] = s_0(x)$, where $s_0(x)$ is an NCS with 7 degrees of freedom. Together with the non-parametric Ederer II estimator of the relative survival [33], these are shown for each age group in Figure 2.

In all age groups, a plateau was seen in the relative survival function. The three parametric models displayed fairly similar trajectories with only small differences seen at the end of the follow-up. For patients >75 years of age, the non-parametric relative survival variance was high by the end of the follow-up, which is explained by only few patients being at risk after 15 years. For patients <55 years of age, the models produced virtually identical trajectories.

We extended the stpm2 and stpm2-cure models with five covariates, namely age at diagnosis, gender (female as reference), Charlson score (<2 vs. $\geq$ 2), a metastases indicator, and clinical stage (UICC, I-II vs. III-IV). The covariates

were included in a proportional excess hazards model, i.e.,

$$\log(-\log(R(t|\boldsymbol{z}))) = s_0(x) + \sum_{i=1}^{5} z_i\beta_i. \tag{15}$$

To compare the coefficients with those of a relative survival model, which is not based on smoothers, we also fitted the piecewise constant excess hazard model introduced by Estève et al. [36]. The model is formulated similarly to (15), but a different formulation is used for $s_0(x)$ which ensures that the baseline excess hazard is a piecewise constant function. The model is fitted by maximum likelihood and is implemented in the `R`-package `relsurv` [37]. The change points for the baseline was selected as zero and the 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, 87.5%, and 100% quantiles of the uncensored follow-up times.

The coefficients, confidence intervals, and p-values of each model are shown in Table 5. Age, Charlson score, metastases indicator, and stage were associated with an increasing excess hazard in all three models. In particular, assuming cure at the last uncensored follow-up time seemed to change the coefficients only slightly.

	Estève ASM		stpm2		stpm2-cure	
	β	P	β	P	β	P
Age	0.024(0.022;0.025)	0.000	0.024(0.022;0.026)	0.000	0.024(0.022;0.026)	0.000
Male gender	0.004(-0.033;0.041)	0.834	0.007(-0.03;0.043)	0.724	0.006(-0.031;0.042)	0.756
Charlson $\geq$2	0.438(0.398;0.477)	0.000	0.438(0.399;0.477)	0.000	0.438(0.398;0.477)	0.000
Metastases	1.668(1.622;1.714)	0.000	1.647(1.601;1.692)	0.000	1.646(1.601;1.692)	0.000
Stage III-IV	1.133(1.062;1.205)	0.000	1.096(1.027;1.164)	0.000	1.098(1.029;1.166)	0.000

Table 5: Regression coefficients from proportional excess hazards models. ASM: additive survival model.

The model in (15) was further extended to incorporate time-varying covariate effects. The time effect of each covariate was modelled by an NCS on the log-time scale with 3 degrees of freedom, i.e.,

$$\log(-\log(R(t|\boldsymbol{z}))) = s_0(x) + \sum_{i=1}^{5} z_i s_i(x), \tag{16}$$

For the stpm2-cure model, the QR-decomposition was applied to s_i for $i = 0,..,6$ separately, with a common last knot placed at the last uncensored follow-up time. The time-varying coefficients of the two models are shown in Figure 3.

The two models display similar time-varying effects only with small differences in the effect of the Charlson score, gender, and stage. Deviations from the proportional excess hazard assumption were observed in all covariates expect the metastases indicator. In addition, the covariate effects in the

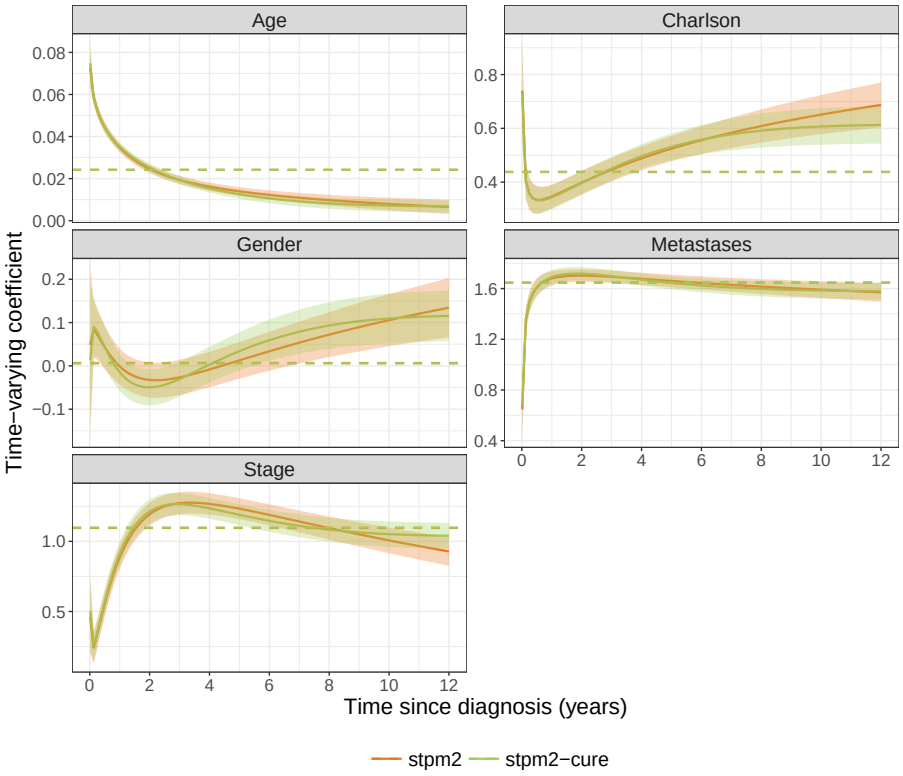

Figure 3: The time-varying effect of age, Charlson score ≥2, male gender, the presence of metastases, and UICC stage ≥2 in Danish colon cancer patients. The horizontal lines indicate the estimated coefficients from the proportional excess hazards models, i.e., the models without time-varying coefficients.

cure model were almost constant after 7 years, while those of the stpm2 model were less restricted.

For estimating the cure proportion, we also fitted a mixture cure model formulated by

$$\log(-\log(S_u(t|z))) = s_0(x) + \sum_{i=1}^{5} z_i s_i(x),$$
$$\log\left(\frac{\pi(z)}{1-\pi(z)}\right) = \beta_0 + \sum_{i=1}^{5} z_i \beta_i, \tag{17}$$

where 7 degrees of freedom were used for s_0 and 2 degrees of freedom was used for s_i where $i \neq 0$. The cure proportion estimated by (17) and (16) is shown in Figure 4 for female colon cancer patients without metastases.

Although similar trends were obtained from these models, relatively large differences were also observed in the estimated cure proportions.

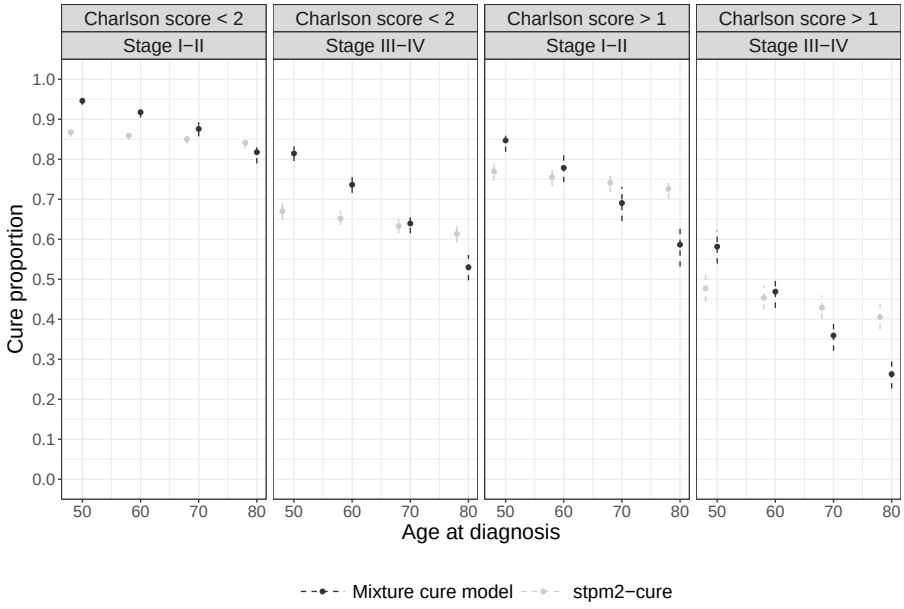

Figure 4: Estimated cure proportion for a female colon cancer patient without metastases based on a latent (stpm2-cure) and an explicit (mixture) cure model.

4.2 Lymphoma data

We analyzed data on 1,621 diffuse large B-cell lymphoma (DLBCL) patients obtained from the Danish National Lymphoma Registry [38]. These data were originally reported in the study of Jakobsen et al. who also provided details on the patient population [39]. Patients were included if they achieved complete remission after immunochemotherapy, i.e., no residual disease was observed after first line treatment. The follow-up was measured from end of treatment until death or censoring. The data set contains cause of death information grouped into categories: lymphoma-related, related to cardiovascular disease, related to other cancers, related to other causes, and no information/no contact to hospital. The cause-specific cumulative incidences computed by the Nelson-Aalen estimator [40] and reported by Jakobsen et al. [39] are shown in Figure S3. These revealed that lymphoma is the main cause of death among the DLBCL patients even after achieving complete remission, which is mainly explained by the occurrence of relapses. In the following, we compare the cumulative incidence and cause-specific hazard

obtained by relying on the available cause of death information and relative survival.

Excess mortality from cause of death information and relative survival

The death causes were divided into 2 groups, i.e., lymphoma-related deaths and deaths due to other causes. Utilizing the cause of death information, we computed the age, gender, and calendar time-dependent cumulative incidences by the approach of Hinchliffe and Lambert [41]. Let S_L and h_L be the survival and hazard function, respectively, corresponding to lymphoma-related deaths when censoring competing risks. Conversely, let S_O and h_O be the analogue survival and hazard function, respectively, for the other causes. The cumulative incidence of lymphoma-related death can then be computed by

$$\int_0^t S_L(u)S_O(u)h_L(u)du$$

using numerical integration (Gauss-Legendre quadrature). The cumulative incidence of death due to other causes is obtained by replacing h_L with h_O. We considered two models for S_L, both formulated as

$$\log(-\log(S_L(t|a))) = s_0(x) + s_a(a) \times s_1(x) + g \times s_2(x) + s_c(c) \times s_3(x), \quad (18)$$

where a is the diagnostic age, g is the gender, and c is calendar time. The functions s_i for $i = 0, 1, 2, 3$ were modelled with NCSs using four degrees of freedom for $i = 0$ and two degrees of freedom for $i \neq 0$. The NCSs s_a and s_c were formulated with two degrees of freedom. One model was fitted using the QR decomposition approach for each s_i ($i = 0, 1, 2, 3$) to incorporate the assumption of cure, while no further restrictions were made for the other. The last knot of the cure model was chosen to be the time of the last uncensored follow-up time. The model in (18) was also used for S_O, but without the assumption of cure.

In addition to these models, we also fitted two relative survival models formulated by (18) with S_L replaced by R. Again, one of the models was fitted using the QR decomposition approach. Using the two relative survival models, the cause-specific cumulative incidences were computed by (11) and (12). The cumulative incidences for a 50-year-old, 60-year-old, and a 70-year-old female patient completing treatment in 2008 are displayed in Figure 5.

A slight difference between using the relative survival approach and the cause of death approach was observed, with a larger lymphoma-specific mortality obtained by the relative survival approach. The same pattern can be seen from the lymphoma-specific hazard functions (Figure 6). Notably, the hazards obtained by the relative suvival approach was larger than those of the cause of death approach late in the follow-up. This suggests that the

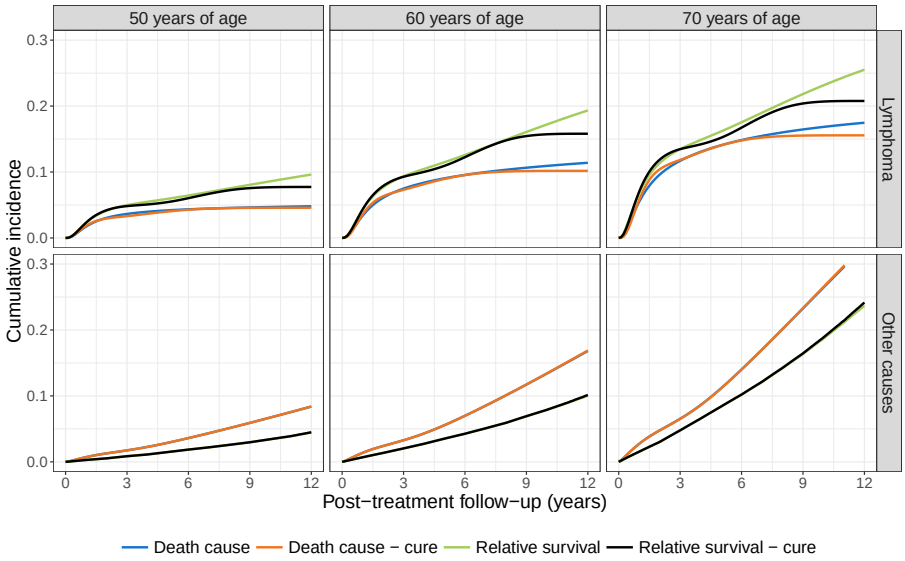

Figure 5: Cause-specific cumulative incidences for a 50-year-old, 60-year-old and 70-year-old female DLBCL patient after achieving first complete remission in 2008. The cumulative incidences were estimated by using four models; two models utilizing cause of death information and two relative survival models. Additionally, two models were fitted using the QR-decomposition approach, which incorporates a cure proportion.

cause of death approach may not capture the entire disease burden, likely because it can be difficult in practice to determine the exact cause of death or weather the death is related to the course of the disease [42]. The differences in the cumulative incidences between the cure models and the non-cure models were only minor and occurred late in the follow-up.

5 Discussion

We presented a general framework for fitting parametric cure models from right-censored and left-truncated time-to-event data using a link function approach. We introduced two general classes of cure models, namely explicit and latent cure models, a general estimating procedure for these, and associated post-estimation summary measures computable from both model classes. These estimating procedures was implemented in the R-packages `rstpm2` and `cuRe`.

The performance of explicit and latent cure models was tested in a simulation study. The explicit cure models may in some scenarios suffer from identifiability issues and provide biased cure proportion estimates, e.g., if the follow-up is not sufficiently long. The latent cure models provide fairly sta-

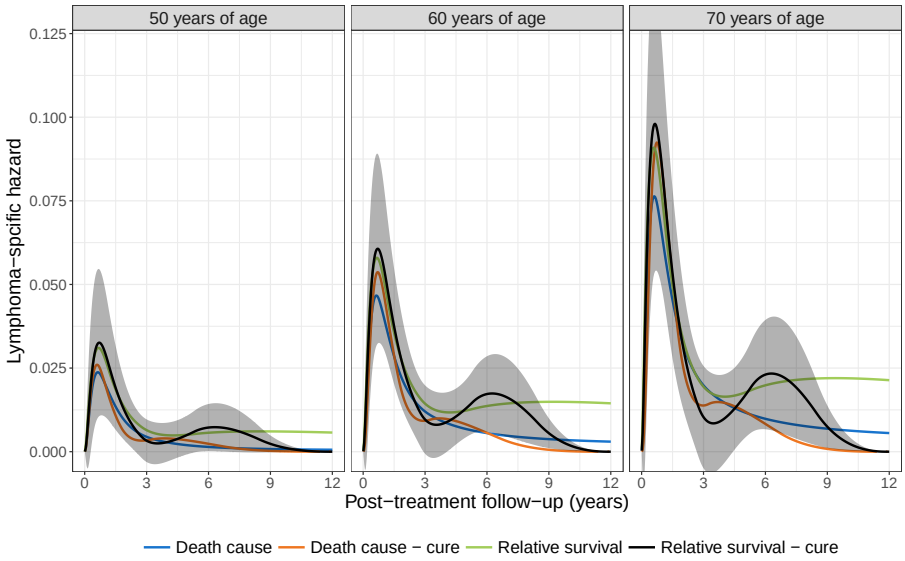

Figure 6: Lymphoma-specific hazards for a 50-year-old, 60-year-old and 70-year-old female DLBCL patient after achieving first complete remission in 2008. The cumulative incidences were estimated by using four models; two models utilizing cause of death information and two relative survival models. Additionally, two models were fitted using the QR-decomposition approach, which incorporates a cure proportion.

ble, but often upward biased estimates since the relative survival is restricted to be constant beyond a finite time point. In particular, choosing a model for estimating the cure proportion is essentially a trade-off between identifiability and bias. While our results guarantee identifiability for models with a cure point within the follow-up, more research is needed to depict the identifiability of the explicit cure models. The explicit cure models considered in the simulation were not restricted to be constant after a specific time point. Thus, the results of the simulations reflects the behaviour of the models under different assumptions about the cure point rather than the formulation of the explicit cure models. In particular, such a restriction could be incorporated in the formulation of $S_u(t)$ and $\widetilde{S}$ by letting $S_u(t_c) = \widetilde{S}(t_c)) = 0$. In practice this could be carried out by using splines restricted to satisfy this property, but this requires additional considerations about the corresponding link function.

Following the methodology of Lambert et al. [26], we applied relative survival models to compute the cause-specific cumulative incidences of lymphoma patients. The relative survival approach offers an alternative to using cause of death information, which may more accurately reflect the excess mortality caused by a particular disease. We used a cause-specific hazard ap-

proach to compute the cumulative incidences based on cause of death information. Alternatively, direct modelling of the cumulative incidence function using splines could have been conducted [43]. In this context, cure models have been proposed for the cumulative incidence, but it is not recommended to assume cure for all competing risks simultaneously [43]. However, interpretation of the parameters from subdistribution hazards models is slightly challenging and thus this approach is often not preferred [44]. Furthermore, neither in our approach nor in the direct modelling approach, the sum of the cumulative incidences are restricted to be one as time goes to infinity.

In the present article, we only considered NCS as smoother. Rutherford et al. [45] found that restricted cubic splines which are similar to NCS, were able to capture the shape of even complex hazard functions given a sufficient number of knots. Other types of smoothers, such as thin plate splines and fractional polynomials, may be used for the explicit cure models, but for the latent cure models, this would require a scheme for restricting the splines to be constant beyond some finite t_c. Also, penalized splines are appealing since the location and number of knots become less important. However, a penalization term that incorporates the constant trajectory beyond the cure point was not readily available.

Model checks in survival analysis are important to assess the assumption of the used model. This can be conducted by computing various types of residuals, e.g., Cox-Snell residuals or deviance residuals, and assess their behavior. Versions of these residuals were recently proposed for the mixture cure model, but these utilizes specific properties of cure models formulated for the all-cause survival function [46]. Therefore, these are not immediately available for cure models for the relative survival. Additionally, an essential problem within cure models is model selection. Especially in explicit cure models this is a problem, since different variables can be used to model the cure proportion and the survival of the uncured. Approaches to incorporate L1-penalization on both terms has been proposed, but these do not fit into our modelling framework or are not readily implemented [47–49]. More research and implementation is needed to incorporate penalization into the presented cure models.

The present methodology was developed for relative survival models, but if total survival is of interest, we simply consider $S^*(t) = 1$ for all time points t. The general population hazard contribution to the generalized likelihood is then zero and all subsequent calculations follow directly after this.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

Acknowledgement

A special thanks to Søren Byg Vilsen for discussions about initial values for the likelihood optimization, and to Jorne Lionel Biccler for his input on the identifiability of cure models. Real-world data were obtained from the Danish Colorectal Cancer Group Database and the Danish National Lymphoma Registry. Additionally, cause of death information in the DLBCL data was collected by reviewing patient reports. We thank the data collectors their significant contribution.

References

[1] V. Bremhorst, M. Kreyenfeld, and P. Lambert. Fertility progression in Germany: an analysis using flexible nonparametric cure survival models. *Demographic Research*, 35:505–534, 2016.

[2] J. W. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11:15–53, 1949.

[3] Y. Peng and K. Dear. A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243, 2000.

[4] J. Zhang and Y. Peng. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*, 26(16):3157–3171, 2007.

[5] M. Othus, B. Barlogie, M. L. Leblanc, and J. J. Crowley. Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, 18(14):3731–3736, 2012.

[6] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–454, 1999.

[7] P. C. Lambert, J. R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, 2007.

[8] P. C. Lambert, P. W. Dickman, C. L. Weston, and J. R. Thompson. Estimating the cure fraction in population-based cancer studies by using finite mixture models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1):35–55, 2010.

[9] K. Seppä, T. Hakulinen, and A. Pokhrel. Choosing the net survival method for cancer survival estimation. *European Journal of Cancer*, 51(9):1123–1129, 2015.

[10] M. P. Perme, J. Stare, and J. Estève. On estimation in relative survival. *Biometrics*, 68(1):113–120, 2012.

[11] P. Royston and M. K. B. Parmar. 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(15):2175–2197, 2002.

[12] X.-R. Liu, Y. Pawitan, and M. Clements. Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5):1531–1546, 2018.

[13] R. Sposto, H. N. Sather, and S. A. Baker. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, 48(1):87–99, 1992.

[14] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, 2007.

[15] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, and P. C. Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, 2011.

[16] D. Rabinowitz and L. Ryan. Lower confidence bounds for time to cure. *Biometrika*, 80(3):681–687, 1993.

[17] R. Sposto. Cure model analysis in cancer: an application to data from the Children's Cancer Group. *Statistics in Medicine*, 21(2):293–312, 2002.

[18] N. Younes and J. Lachin. Link-based models for survival data with interval and continuous time censoring. *Biometrics*, 53(4):1199–1211, 1997.

[19] A. Cantor and J. Shuster. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, 11(7):931–937, 1992.

[20] A. Tsodikov. A proportional hazards model taking account of long-term survivors. *Biometrics*, 54(4):1508–1516, 1998.

[21] V. Bremhorst and P. Lambert. Flexible estimation in cure survival models using Bayesian P-splines. *Computational Statistics & Data Analysis*, 93:270–284, 2016.

[22] S. N. Wood. *Generalized additive models : an introduction with R*. CRC Press, second edition, 2017.

[23] P. C. Lambert. Modeling of the cure fraction in survival studies. *Stata Journal*, 7(3):351–375, 2007.

[24] L. Hanin and L.-S. Huang. Identifiability of cure models revisited. *Journal of Multivariate Analysis*, 130:261–274, 2014.

[25] C.-S. Li, J. M. Taylor, and J. P. Sy. Identifiability of cure models. *Statistics & Probability Letters*, 54(4):389–395, 2001.

[26] P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine*, 29(7-8):885–895, 2010.

[27] S. Eloranta, P. C. Lambert, T. M.-L. Andersson, M. Björkholm, and P. W. Dickman. The application of cure models in the presence of competing risks. *Epidemiology*, 25(5):742–748, 2014.

[28] T. M.-L. Andersson, P. W. Dickman, S. Eloranta, M. Lambe, and P. C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, 2013.

[29] L. H. Jakobsen, T. M.-L. Andersson, J. Biccler, T. C. El-Galaly, and M. Bøgsted. Estimation of the loss of lifetime function using flexible parametric relative survival models. *Submitted for publication*, 2018.

[30] X.-R. Liu, Y. Pawitan, and M. S. Clements. Generalized survival models for correlated time-to-event data. *Statistics in Medicine*, 36(29):4743–4762, 2017.

[31] M. J. Rutherford, P. W. Dickman, and P. C. Lambert. Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology*, 36(1):16–21, 2012.

[32] M. Crowther and P. Lambert. Simulating biologically plausible complex survival data. *Statistics in Medicine*, 32(23):4118–4134, 2013.

[33] F. Ederer and H. Heise. Instructions to IBM 650 programmers in processing survival computations. *Methodological note No. 10, End Results Evaluation Section. Bethesda:National Cancer Institute*, 1959.

[34] Human mortality database. URL: `www.mortality.org`. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Accessed 15 September 2017.

[35] P. Ingeholm, I. Gögenur, and L. H. Iversen. Danish Colorectal Cancer Group Database. *Clinical Epidemiology*, 8:465–468, 2016.

[36] J. Estève, E. Benhamou, M. Croasdale, and L. Raymond. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9(5):529–538, 1990.

[37] M. Pohar and J. Stare. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81(3):272–278, 2006.

[38] B. Arboe, P. Josefsson, J. Jørgensen, J. Haaber, P. Jensen, et al. Danish National Lymphoma Registry. *Clinical Epidemiology*, 8:577–581, 2016.

[39] L. H. Jakobsen, M. Bøgsted, P. d. N. Brown, B. Arboe, J. Jørgensen, et al. Minimal loss of lifetime for patients with diffuse large B-cell lymphoma in remission and event free 24 months after treatment: a Danish population-based study. *Journal of Clinical Oncology*, 35(7):778–784, 2017.

[40] T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data*. Springer New York, 2006.

[41] S. Hinchliffe and P. Lambert. Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC Medical Research Methodology*, 13:13, 2013.

[42] C. Percy, E. Stanek, L. Gloeckler, and L. Gloeckler. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American Journal of Public Health*, 71(3):242–250, 1981.

[43] S. Mozumder, M. Rutherford, and P. Lambert. Direct likelihood inference on the cause-specific cumulative incidence function: A flexible parametric regression modelling approach. *Statistics in Medicine*, 37(1):82–97, 2018.

[44] P. K. Andersen and N. Keiding. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, 31(11-12):1074–1088, 2012.

[45] M. J. Rutherford, M. J. Crowther, and P. C. Lambert. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, 85(4):777–793, 2015.

[46] S. Scolas, C. Legrand, A. Oulhaj, and A. El Ghouch. Diagnostic checks in mixture cure models with interval-censoring. *Statistical Methods in Medical Research*, 27(7):2114–2131, 2018.

[47] X. Liu, Y. Peng, D. Tu, and H. Liang. Variable selection in semiparametric cure models based on penalized likelihood, with application to breast cancer clinical trials. *Statistics in Medicine*, 31(24):2882–2891, 2012.

[48] S. Scolas, A. El Ghouch, C. Legrand, and A. Oulhaj. Variable selection in a flexible parametric mixture cure model with interval-censored data. *Statistics in Medicine*, 35(7):1210–1225, 2016.

[49] A. Masud, W. Tu, and Z. Yu. Variable selection for mixture and promotion time cure rate models. *Statistical Methods in Medical Research*, 27(7):2185–2199, 2018.

Supplementary

S1 Interpretation of cure models

All-cause survival

Let X be a random variable denoting the time to the event of interest. On the all-cause survival scale, individuals are cured if $X = \infty$, i.e., the event never occurs. The probability of cure is therefore $\pi = \mathrm{P}(X = \infty)$ and $1 - \pi = \mathrm{P}(X < \infty)$. The random variable X has the improper distribution function, $F(t) = \mathrm{P}(X \leq t)$ for which we have

$$1 - \lim_{t \to \infty} S(t) = \lim_{t \to \infty} F(t) = 1 - \pi.$$

Thus, the cure probability is equal to the asymptote of the survival function, S. For cure models, where $S(t)$ is constant for $t \geq t_c$, we have that $\pi = S(t_c)$. Therefore, the cure probability can in these models be estimated by evaluating the survival function at t_c.

Relative survival

Let T_D be the time to death due to a particular disease of interest and let T_P be the time to death from other causes. Patients may be considered cured if $T_D = \infty$, i.e., if they never die from the disease. Given a vector of covariates, $\boldsymbol{z}$, the probability of cure is given as $\pi(\boldsymbol{z}) = \mathrm{P}(T_D = \infty | \boldsymbol{z})$ and $1 - \pi(\boldsymbol{z}) = \mathrm{P}(T_D < \infty | \boldsymbol{z})$. The follow-up times, T_D and T_P, are competing events, and we only observe $X = \min(T_D, T_P)$. Now, assume that deaths from the disease of interest make up a negligibly small part of the general population mortality. Using the survival distribution associated with each random variable (S_{T_D}, S_{T_P}, and S_X), we define the relative survival as

$$R(t|\boldsymbol{z}) = \frac{S_X(t|\boldsymbol{z})}{S_{T_P}(t|\boldsymbol{z})}.$$

Assume now that T_D and T_P are conditional independent given $\boldsymbol{z}$. Then we have,

$$\begin{aligned}
1 - \lim_{t\to\infty} R(t|\boldsymbol{z}) &= 1 - \lim_{t\to\infty} \frac{S_X(t|\boldsymbol{z})}{S_{T_P}(t|\boldsymbol{z})} = 1 - \lim_{t\to\infty} \frac{\mathrm{P}(T_D > t, T_P > t|\boldsymbol{z})}{\mathrm{P}(T_P > t|\boldsymbol{z})} \\
&= 1 - \lim_{t\to\infty} \frac{\mathrm{P}(T_D > t|T_P > t, \boldsymbol{z})\mathrm{P}(T_P > t|\boldsymbol{z})}{\mathrm{P}(T_P > t|\boldsymbol{z})} \\
&= 1 - \lim_{t\to\infty} \mathrm{P}(T_D > t|\boldsymbol{z}) = \lim_{t\to\infty} \mathrm{P}(T_D \leq t|\boldsymbol{z}) \\
&= 1 - \pi(\boldsymbol{z}).
\end{aligned}$$

Again, if $R(t|\boldsymbol{z})$ is constant for $t \geq t_c$, we have that $\pi(\boldsymbol{z}) = R(t_c|\boldsymbol{z})$. The assumption of conditional independence is vital in order to obtain the above result and was also a key assumption in the net survival discussion by Perme et al. [10]. However, this assumption is not identifiable from the data [44].

On the other hand, a different interpretation may be given to the cure probability. Instead of considering two competing causes of death, we consider the population a mixture of cured (Y = 1) and uncured (Y = 0) individuals. In addition, we assume that the relative survival for $Y = 1$ is a constant equal to one, and the relative survival for $Y = 0$ goes towards zero as time approaches infinity. The cure probability may then be interpreted similarly to mixing weights from regular mixture models, i.e., the probability of belonging to each group.

S2 Simulation parameters

In Section 3, the simulations were conducted using a Weibull model and polynomial splines for $S_u(t)$. The parametrization of the Weibull model was

$$S_u(t) = \exp\left(-\gamma_0 t_1^{\gamma}\right).$$

and the polynomial spline model was formulated as

$$S_u(t) = \exp\left(-\exp\left(\beta_0 + \sum_{i=1}^{5} v_i(t)\beta_i\right)\right).$$

The basis functions, $v_i(t)$, were obtained by the `R`-function `bs`, which computes the basis function using B-splines. The polynomials were of degree 3, resulting in cubic splines. Four knots placed at time 0, 1, 7, and 15 were selected for these simulations. Table S1 displays the exact parameter values and Figure S1 displays the relative survival trajectory of each simulation scenario.

Scenario	Weibull		Polynomial splines					
	γ_0	γ_1	β_0	β_1	β_2	β_3	β_4	β_5
1	1.0	1.0	-6.0	4.5	7.8	7.8	8.0	8.5
2	1.2	0.1	-6.9	4.8	5.9	7.0	7.5	7.7
3	1.4	0.1	-7.5	5.3	6.7	8.1	8.7	9.0
4	1.0	0.5	-6.0	4.5	6.8	7.8	7.5	8.0
5	1.0	0.1	-6.5	4.3	5.8	6.4	6.6	6.7
6	1.2	0.1	-6.9	4.8	5.9	7.0	7.5	7.7

Table S1: Parameter values used for simulating survival data.

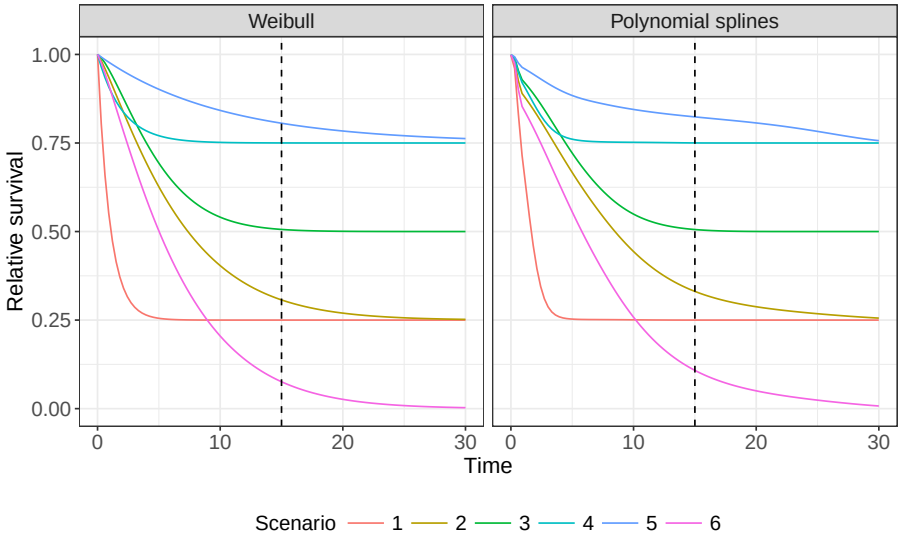

Figure S1: Relative survival functions from which the simulated data in Section 3 were generated. The vertical line indicates the maximum possible follow-up time (15 years).

S3 Additional simulation results

In this section, we re-estimated the cure proportion using model B of Section 3.2 with a varying number of knots. Figure S2 displays the estimated cure proportions in each scenario of the Weibull and polynomial spline simulations. In scenario 1, 3, 4, and 6, no substantial differences were seen between the models. In scenario 2 and 5 larger differences were seen between the models. In scenario 2 the simple model with two parameters had the best performance and increasing flexibility led to increased estimates of the cure proportion. In scenario 5, the dispersion of all models was rather large. However, except for the simple model, the dispersion was decreasing with more flexibility.

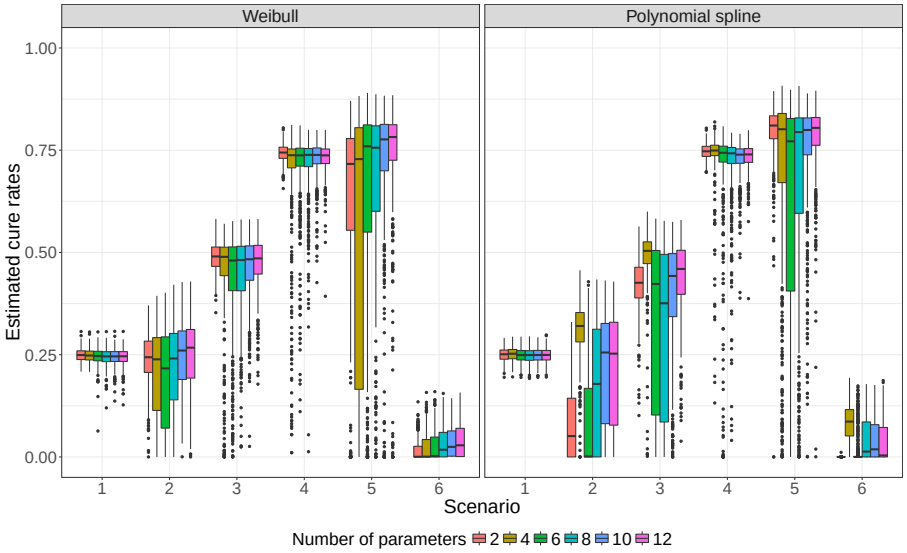

Figure S2: Estimated cure proportions in simulations of the scenarios in Figure S1 using model B of Section 3.2 with a varying number of knots, resulting in a varying number of parameters. The number of parameters used for the spline is displayed.

S4 Supplementary Figures

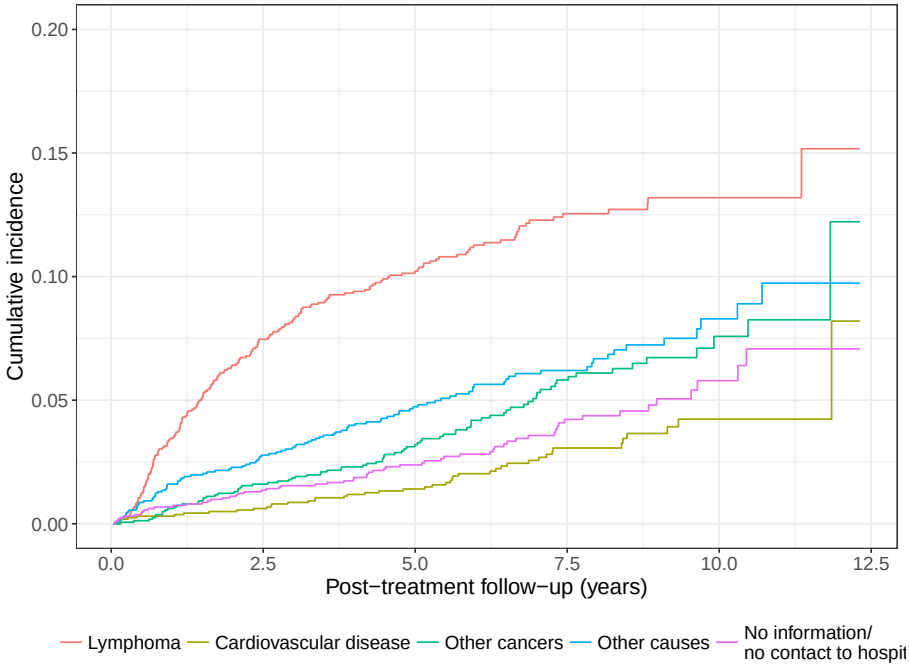

Figure S3: Cause-specific cumulative incidences of death in Danish DLBCL patients after achieving complete remission.

Paper V

Clinical prognostic scores are poor predictors of overall survival in various types of malignant lymphomas

Jorne L. Biccler[1,2], Tarec C. El-Galaly[1,2], Martin Bøgsted[1,2], Judit Jørgensen[3], Pater dN. Brown[4], Christian B. Poulsen[5], Jørg Starklint[6], Maja B. Juul[7], Jacob H. Christensen[8], Pär Josefsson[9], Andriette Dessau[10], and Lasse H. Jakobsen[1,2]

1. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; 2. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; 3. Department of Hematology, Aarhus University Hospital, Aarhus, Denmark; 4. Department of Hematology, Rigshospitalet, Copenhagen, Denmark; 5. Department of Hematology, Roskilde Sygehus, Roskilde, Denmark; 6. Department of Medicine, Hospitalsenheden Vest, Holstebro, Denmark; 7. Department of Hematology, Vejle Sygehus, Vejle, Denmark; 8. Department of Hematology, Odense University Hospital, Odense, Denmark; 9. Department of Hematology, Herlev Hospital, Herlev, Denmark; 10. Department of Hematology, Sydvestjysk Sygehus, Esbjerg, Denmark.

Letter submitted to Leukemia & Lymphoma

Description

As most clinical prognostic scores within lymphoma are based on dichotomized or grouped clinical variables, these do not utilize all available information for prognostication. By avoiding this, more accurate risk assessments may be obtained, in the form of survival probabilities or risk groupings. The aim of this paper was to evaluate the performance of commonly used prognostic scores within lymphoma. By utilizing data from the Danish Lymphoma Registry prognostic scores used within 11 common lymphoma types were evaluated.

Letter

Malignant lymphomas comprise a heterogeneous group of diseases with important differences in pathogenesis, management strategies, and outcome [1]. To obtain information about prognosis in daily practice, clinical prognostic models are often used. The most prominent examples of these are the international prognostic index (IPI) for diffuse large B-cell lymphoma (DLBCL) and the follicular lymphoma (FL)-specific international prognostic index (FLIPI) [2, 3]. For some lymphoma subtypes, treatment decisions are influenced by these scores, as exemplified by the ESMO guidelines for DLBCL where the age-adjusted IPI is used to differentiate treatment in patients younger than 60 years [4]. To simplify usability, prognostic scores are often constructed using dichotomized and/or grouped prognostic factors and predicted risks. However, the cost of dichotomization is a large loss of information resulting in inaccurate predictions of patient survival. Despite awareness of the disadvantages of dichotomization and grouping, recently developed prognostic models have also applied this approach [5, 6].

In the present study, we evaluated the predictive performance of established prognostic scores and two alternative models for 11 common subtypes of malignant lymphoma in Denmark. To do so, a nationwide cohort of adult patients diagnosed between 2006 and 2016 was extracted from the Danish National Lymphoma Registry [7]. Following the international prognostic score (IPS), only advanced stage lymphoma cases were included for Hodgkin lymphoma (HL) [8]. For the non-HL subtypes all patients were included. The follow-up was measured as the time from diagnosis to death or censoring (November 2016), whichever came first. The study was approved by the Danish Data Protection Agency (2008-58-0028).

For each lymphoma subtype, a conventional prognostic score, such as those recommended by the ESMO or NCCN guidelines, was selected for performance assessment. A list of the included prognostic indices is found in Table 1.

In order to assess the potential gain of more refined modelling, the prognostic scores were compared to a machine learning (ML) approach in which the best model was selected among a number of prespecified models. The models considered were two instances of random survival forest (RSF) including up to 20 clinical variables, a Cox proportional hazard (CPH) model, a penalized CPH model, and an accelerated failure time model (see supplementary). Furthermore, for each lymphoma subtype, a CPH model including age as continuous variable and performance status as categorical variable, termed the simple model, was tested. Age and performance status were selected since they are readily available and are also important survival predictors in the general population.

The performance of the prognostic models was measured by the time-

Lymphoma type	Index	Variables included in the prognostic index	End point	Outcome	N	Mean age (SD)	Gender ratio (M/F)
Advanced classical Hodgkin lymphoma	IPS [8]	Albumin, hemoglobin, sex, stage, age, leukocyte count, lymphocyte count	CSS	0, 1, 2, 3, 4, 5-	785	48(19.7)	1.4
Anaplastic large cell lymphoma	IPI [2]	Age, stage, LDH, performance status, extra nodal	OS	0-1, 2, 3, 4-5	176	57.3(17.6)	1.6
Angioimmunoblastic T-cell lymphoma	IPI [2]	Age, stage, LDH, performance status, extra nodal	OS	0-1, 2, 3, 4-5	127	68.8(12.6)	1.2
Burkitt lymphoma	[9]	Age, stage	RS	0-1, 2, 3, 4-	129	53(18.6)	3.6
Diffuse large B-cell lymphoma	IPI [2]	Age, stage, LDH, performance status, extra nodal	OS	0-1, 2, 3, 4-5	4420	67.4(13.5)	1.3
Extranodal marginal zone B-cell lymphoma (MALT-lymphoma)	MALT-IPI [6]	Age, LDH, stage	CSS	0, 1, 2-	629	67.3(14)	0.9
Follicular lymphoma	FLIPI [3]	Age, extra nodal, LDH, hemoglobin, stage	OS	0-1, 2, 3-5	2129	63.9(12.2)	1
Lymphoplasmacytic lymphoma	ISSWM [10]	Age, hemoglobin, thrombocyte count, β2 microglobulin count, IgM count	OS	0-1, 2, 3-	1215	71.6(9.9)	1.6
Mantle cell lymphoma	MIPI [11]	Age, performance status, LDH, Leukocyte count	OS	0, 1, 2	686	69.8(11)	2.2
Peripheral T-cell lymphoma, unspecified	IPI [2]	Age, stage, LDH, performance status, extra nodal	OS	0-1, 2, 3, 4-5	230	62.5(15.4)	1.6
Splenic marginal zone lymphoma	[12]	Hemoglobin, LDH, Albumin	CSS	0, 1, 2-	338	70.7(10.3)	0.9
Total					10,864	65.6(14.6)	1.3

Table 1: The chosen clinical prognostic indices for 11 common lymphoma types in the Danish National Lymphoma Registry. For each index the involved clinical variables, original end-point, and final risk groups are shown. For each lymphoma type, frequency, age, and gender ratio are also displayed. CSS: cause specific survival, OS: overall survival, RS: relative survival.

varying AUC (tAUC) and the Brier score (BS). The tAUC provides a measure of the correspondence between the predicted risk ranks and the observed survival statuses at specific time points, while the BS measures how close the predicted survival probabilities are to the survival as observed in the data. The tAUCs and BSs were estimated by ten-fold cross-validation.

In total, 10,864 patients distributed across 11 lymphoma subtypes (Table 1) were included (Figure S1). The median follow-up of the entire patient population was 4.98 years (reverse Kaplan-Meier method). The tAUCs of the prognostic scores, the simple model, and the ML approach are shown in Figure 1. Consistently across all lymphoma types, except peripheral T-cell lymphoma (PTCL), the tAUC was higher for the ML approach than for the prognostic indices. In addition, the same observation was made when the simple model was compared to the prognostic indices. Compared to the ML approach, the simple model had slightly improved discriminative capabilities in some subtypes (Burkitt lymphoma and splenic marginal zone lymphoma), while showing similar performance in six lymphoma types (HL, anaplastic large cell lymphoma, angioimmunoblastic T-cell lymphoma, lymphoplasmacytic lymphoma, and PTCL). Only in DLBCL, MALT-lymphoma, FL, and mantle cell lymphoma (MCL) it seemed beneficial to use the more complex modelling approach. The model performance as measured by the BS (Figure S2) was largely consistent with the tAUC showing inferior performance of the prognostic indices and only small changes in the ranking between the ML approach and the simple model. Finally, restricting the analyses to patients treated with curative intent led to a similar ranking of the three model approaches (see supplementary).

Dichotomization of the clinical variables is likely the main reason for the inferior performance of the established prognostic scores. For DLBCL it has previously been shown that especially the dichotomization of age impairs the predictive accuracy significantly [13]. Among the prognostic indices considered here, only the MCL-IPI utilized continuous clinical variables, but patients are still subsequently assigned to risks groups. We find it more natural that the primary goal of prognostic models is to yield survival probabilities instead of risk group assignments, which ignores the, often substantial, survival variability within each risk group. In practice, the use of cut-offs is sometimes necessary, e.g., to base treatment decisions on. In such scenarios, post-hoc risk grouping based on the predicted survival probabilities can be performed, e.g., by defining patients as high risk when the 2-year survival probability falls below 50%. Due to the differences in survival trajectories and patient characteristics, such cut-offs would have to be subtype-specific.

The introduction of novel drugs is increasing the therapeutic armamentarium against lymphoma. New treatment regimens often influence overall survival and challenge the validity of existing models predicting this endpoint. Attempts to update some of the considered prognostic scores to the era of

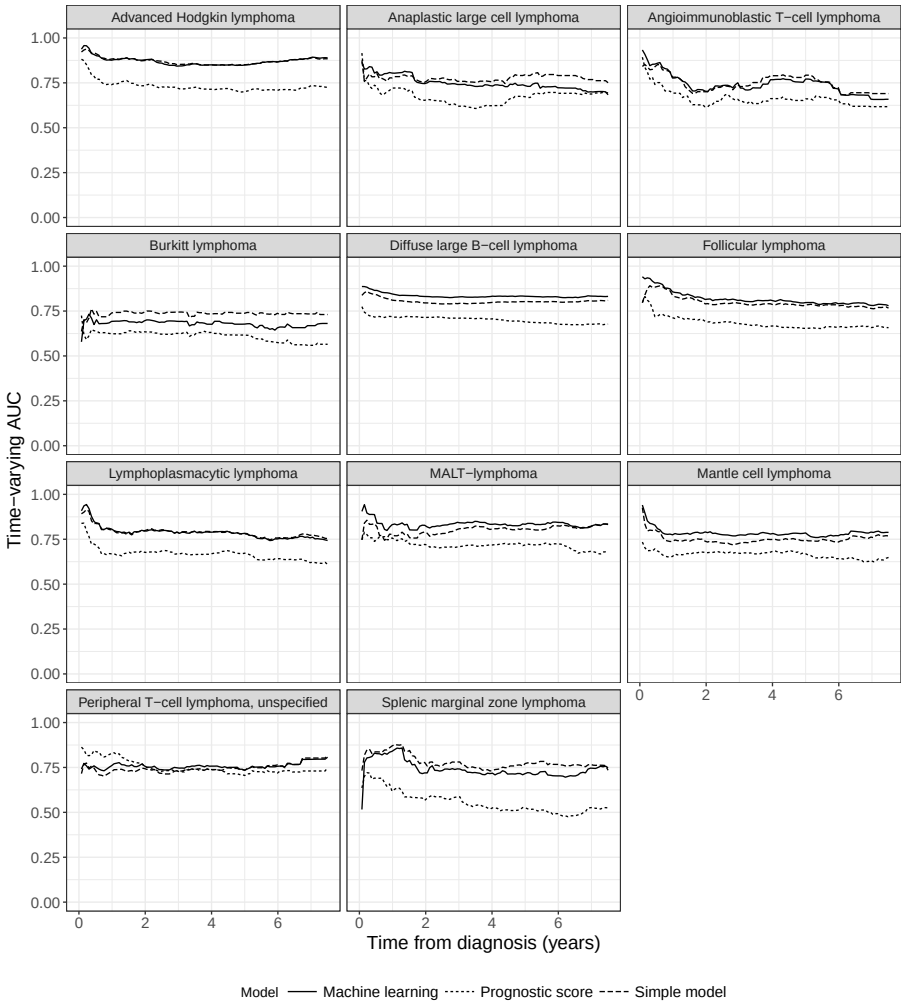

Figure 1: Time-varying AUC (tAUC). For each lymphoma type, the tAUC was computed for the corresponding prognostic score, the machine learning approach, and the simple model. The tAUCs were estimated by using 10-fold cross-validation.

immunochemotherapy have been made, e.g., by introducing the NCCN-IPI [5] for DLBCL, and FLIPI2 [14] for FL. However, these indices suffer from the same disadvantages (dichotomization and risk grouping) as many other prognostic indices and do not generally show good performance [13].

Another potential reason for the inferior performance of the established prognostic indices is that they are sometimes derived from clinical trial data [2, 3], which are well known to be subject to selection bias introduced by

strict inclusion criteria. Therefore, the indices may not be generalizable to the general patient population. By re-establishing the prognostic scores using nationwide register data, these biases could be minimized.

Due to incompleteness of cause of death and relapse information, OS was used as endpoint. However, not all prognostic scores were developed with OS as endpoint and hence their evaluation using the intended endpoint might have been more appropriate. Nonetheless, given the poor performance of the indices developed with OS as endpoint, we expect the same issues.

The rather discouraging increase in predictive performance of the ML approach, seen even for aggressive lymphomas, can be related to the fact that cancers are fundamentally genetic diseases. Attempts have been made to use genetic measurements such as gene expression data for prognostic modelling [15], however, such tools are still not a part of daily clinical practice.

In conclusion, the performance of commonly used prognostic indices for 11 common malignant lymphoma subtypes is subpar. This result was seen in a large nationwide cohort which was not hampered by selection bias. Interestingly, in the majority of the lymphoma subtypes, the ML approach did not outperform the simple model and only a moderate improvement was seen in the remaining lymphoma types. This suggests that age and performance status still are strong clinical predictors of survival in lymphoma, just as in the general population. Establishing new clinical prognostic models for each major lymphoma type is a natural next step in establishing risk assessment tools to be used in clinical practice.

Disclosures of interest

The authors report no conflicts of interest.

Supplementary

The Danish Nation Lymphoma Registry

The Danish National Lymphoma Registry (LYFO) has been nationwide since 2000 and contains detailed information on clinical baseline characteristics, treatment, and outcomes. In a recent quality assessment, the database was shown to be of high quality and, when compared to the Danish Cancer Register, it was shown that LYFO covers approximately 95% of all lymphoma cases in Denmark [7]. Complete follow-up information is ensured by regularly merging with the Danish Civil Registration System.

Patients

In order to properly assess the performance of the prognostic models in each lymphoma type, we restricted the analyses to subtypes with more than 100 patients diagnosed in the time interval of interest. Patients for whom a subtype could not be identified or patients diagnosed with small lymphocytic lymphoma were also excluded, resulting in 10,864 patients distributed across 11 lymphoma types (Table 1). The 11 considered lymphoma subtypes were based on the histology code grouping shown in Table S1. For HL, only patients who presented with advanced stage disease were included in the present study. Advanced stage HL was defined as Ann Arbor stage III-IV, stage II with B-symptoms, or bulky disease (patients with missing information on B-symptoms and stage were also excluded).

Prognostic indices

For each of the included malignant lymphoma subtypes, one established prognostic index was tested. The choice of prognostic model was based on the guidelines of the European Society for Medical Oncology (ESMO) and National Comprehensive Cancer Network (NCCN). For subtypes where prognostic models were not mentioned in the guidelines (Burkitt lymphoma, extranodal marginal zone (MALT) lymphoma, and splenic marginal zone lymphoma), we searched for prognostic indices using the search terms 'prognostic index' and 'prognostic factors' together with the lymphoma type in PubMed. In common for all prognostic models tested was the assignment of patients into predefined risk groups rather than resulting in predicted survival probabilities (see Table 1 for the number of risk groups).

Machine learning model

To establish a more advanced prognostic model for each lymphoma type, we selected the best model among: two random survival forest (RSF)[16]-based predictive models, the disease-specific prognostic index, a disease-specific CPH model which included the undichtomized/ungrouped variables of the corresponding prognostic index, a CPH model with a ridge penalty [17] which included a wider range of variables, and a disease-specific log-logistic accelerated failure time which included the undichtomized/ungrouped variables of the corresponding prognostic index. The best model was defined as the model with the lowest integrated BS as computed by cross-validation. We refer to this modelling approach as the machine learning (ML) model. The variables included in the penalized CPH model and the RSFs are specified in Table S2. The selection of these variables was based on their use in the existing prognostic scores and availability in LYFO. Establishing prognostic

models for rare lymphoma types is challenging for a number of reasons. Because of the small sample sizes, typically the models are unstable [18]. In an attempt to overcome these limitations, one of the RSF models was established in a pooled manner to share information across lymphoma types. This was done by adding lymphoma type as a prognostic factor and fitting the RSF using all patient records. Whenever a clinical variable is selected for splitting before all patients have been stratified according to the lymphoma type by the RSF procedure, information is essentially shared between two or more subtypes. For the second RSF-based model a separate RSF was fitted for each included lymphoma type.

Missing data handling

Missing data in the clinical variables were handled by single imputation using a conditional mean imputation technique based on random forest as implemented in the `R`-package `mlr` [19]. Conditional mean imputation was employed to avoid the excessive computational running time of multiple imputation, while preserving a similar predictive accuracy [20].

Cross validation and predictive performance

The performance of each model (prognostic index, simple model, and ML approach) was assessed by 10-fold cross validation. The 10-fold splitting of the data was conducted subtype-wise, i.e., in each fold one tenth of each lymphoma type was included. For the best model approach, internal 10-fold cross validation was performed to find the best model. The best model was defined as the model that minimized the integrated BS in the period from 0 to 7.5 years in the inner cross validation loop. Imputation of missing values was conducted inside both the outer cross validation and inner cross validation loop. In addition to the tAUC [21], the predictive performance of each model was inspected by estimating the BS [22]. The BS measures the calibration of the models and hence provides a measure of how close the predicted survival probabilities are to the observed survival statuses at different time points. To calculate the Brier score both in the inner and outer cross validation loop, weighting with the censoring distribution in the test data was needed. Since some subtypes were infrequent, some subtype-specific test sets were rather small. To cope with this, the censoring distribution was estimated lymphoma type-wise using all samples of that lymphoma type. The BSs of the prognostic indices, simple model, and ML approach can be found in Figure S2.

Curative treatment

In aggressive lymphoma types, the administration of palliative versus curative treatment is often strongly correlated with performance status and

age. To inspect whether this had an effect on the performance ranking of the different models, a restricted analysis was performed. In particular, for the aggressive lymphoma subtypes, the analyses were redone using only patients treated with chemotherapy (see Table S1 for a detailed description of the included treatments). The tAUCs and BS can be found in Figures S3 and S4, respectively. Although minor changes were observed, the main conclusion remains valid, i.e., the majority of the established prognostic indices can easily be improved upon.

Tables

Lymphoma type	**Histology code (n)**	**Treatment regimens considered with cureative intent**
Anaplastic large cell lymphoma	9714 (176)	CHOP/CHOEP/CEOP
Angioimmunoblastic T-cell lymphoma	9705 (127)	CHOP/CHOEP/CEOP
Burkitt lymphoma	9687 (129)	CHOP/CHOEP/CODOX-M/IVAC/BFM
Classical Hodgkin lymphoma	9650 (cHL, 126), 9663 (nodular sclerosis cHL, 457), 9651 (lymphcyte-rich cHL, 24), 9652 (mixed cellularity cHL, 159), and 9653 (lymphocyte-depleted cHL, 19)	BEACOPP/ABVD/MOPP
Diffuse large B-cell lymphoma	9680 (4420)	CHOP/CHOEP/CEOP
Extranodal marginal zone B-cell lymphoma (MALT-lymphoma)	9699 (629)	/
Follicular lymphoma	9690 (381), 9691 (grade I, 571), 9695 (grade II, 827), and 9698 (grade III, 350)	/
Lymphoplasmacytic lymphoma	9671 (1215)	/
Mantle cell lymphoma	9673 (686)	/
Peripheral T-cell lymphoma, unspecified	9702 (230)	CHOP/CHOEP/CEOP
Splenic marginal zone lymphoma	9689 (338)	/

Table S1: Lymphoma types included in the present study together with the defining histology codes for each type. "/" indicates indolent lymphoma types or lymphoma types for which there are no proven standard treatments. Treatment abbreviations: CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisolone); CHOEP (cyclophosphamide, doxorubicin, vincristine, etoposide, and prednisone); CEOP (cyclophosphamide, epirubicin, oncovin (vincristine), and prednisone); CODOX-M/IVAC (cyclophosphamide, doxorubicin, vincristine, methotrexate/ifosfamide, etoposide, high dose cytarabine); BFM (vincristine, prednisone, doxorubicin, cyclophosphamide, ifosfamide, mesna, etoposide, methotrexate, dexamethason); BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone); ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine); MOPP (vincristine, nitrogen mustard, procarbazine hydrochloride, and prednisone).

Variable	Description	Formatting
Age	Age at diagnosis	Continuous
Sex	Gender	Categoral - male or female
Ann Arbor stage	Ann Arbor stage	Categorial - 1, 2, 3, or 4
Performance status	ECOG performance status	Categorical - 0, 1, 2, 3, or 4
B-symptoms	B-symptoms	Categorical - yes or no
LDH value	Measured in μ/L. Standardized by the national upper reference (255 and 205 μ/L for patients $\geq$ 70 years and < 70 years, respectively)	Continuous
Extra nodal	Number of extranodal sites	Continuous
Hemoglobin value	Measured in g/L	Continuous
Albumin value	Measured in g/L	Continuous
Lymphocyte count	Billions per liter	Continuous
Leukocyte count	Billions per liter	Continuous
Tumor diameter	Maximal tumor diameter (cm)	Continuous
β2 microglobulin value	Measured in g/L	Continuous
IgG value	Measured in g/L	Continuous
IgM value	Measured in g/L	Continuous
IgA value	Measured in g/L	Continuous
Creatine value	Measured in mmol/L	Continuous
Thrombocyte count	Billions per liter	Continuous
CNS	CNS, eye, or leptomeninges involvement	Categorical - yes or no
Lymphoma type	Classifed lymphoma type	Categorial - values of first column in Table S1

Table S2: Variables used for predicting survival in Danish lymphoma patients.

Figures

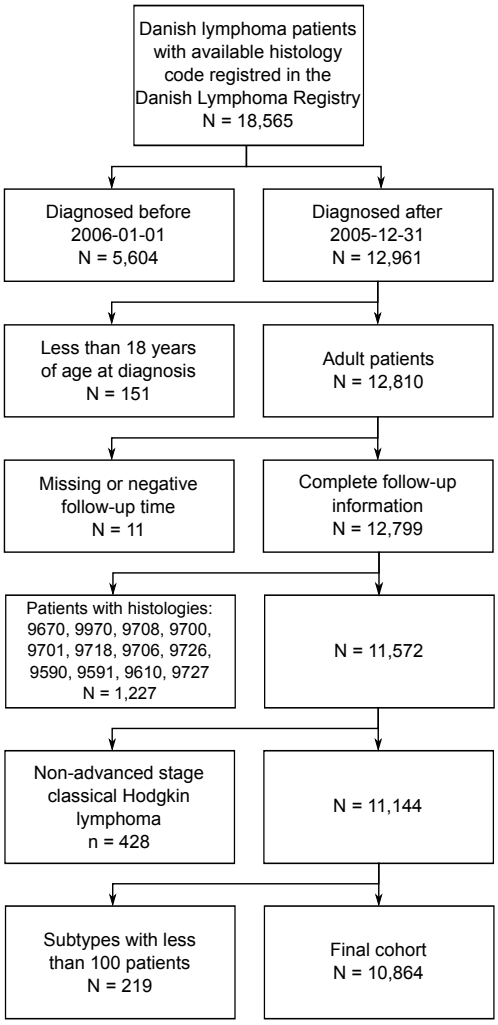

Figure S1: Flow chart of the inclusion criteria.

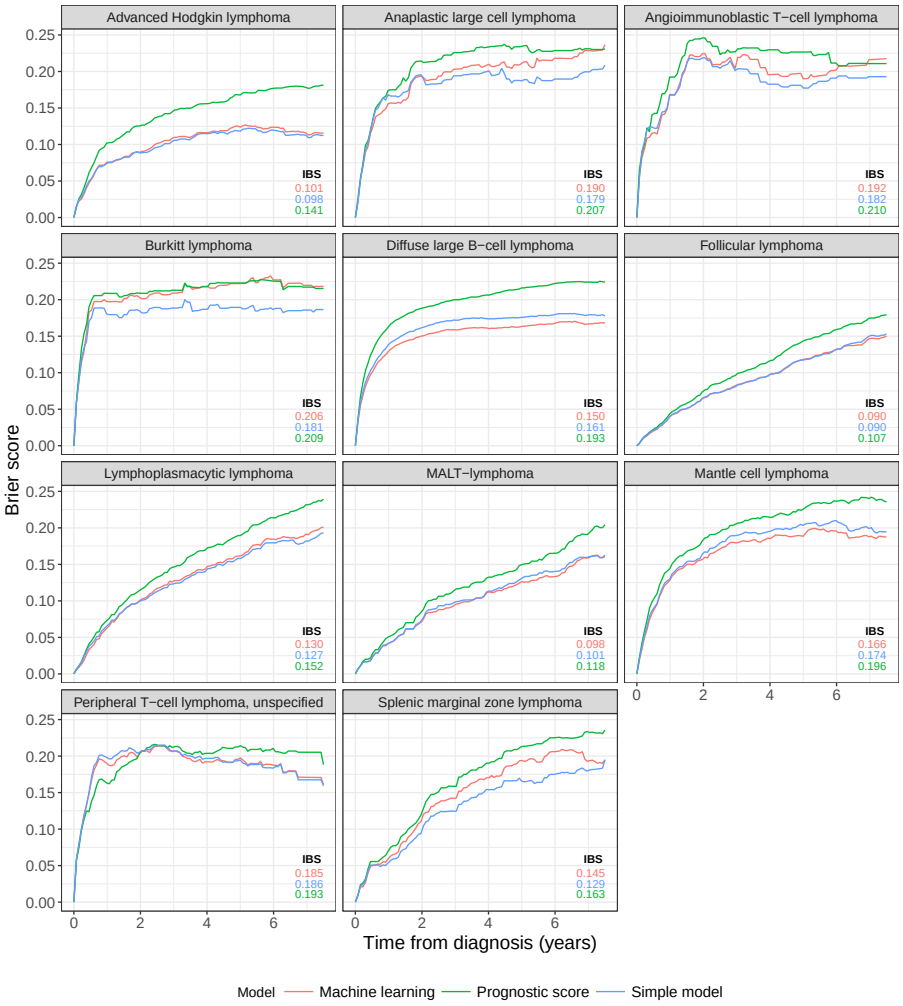

Figure S2: Time-varying Brier score, computed for each lymphoma type, for the prognostic index, the machine learning approach, and the simple model. The integrated Brier score (IBS) was computed and is shown in each plot window.

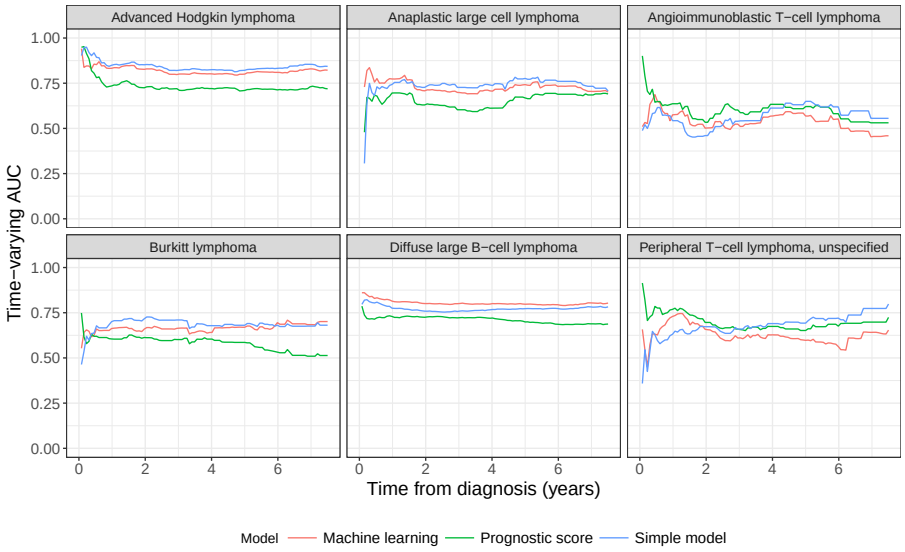

Figure S3: Time-varying area under the curve (AUC), computed for each aggressive lymphoma type, for the prognostic index, the machine learning approach, and the simple model. Only patients treated with curative intent were included.

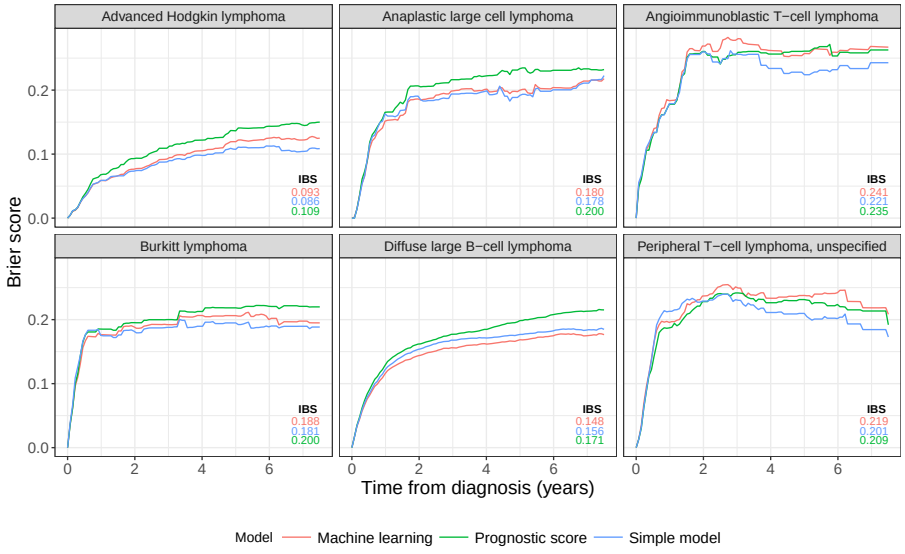

Figure S4: Time-varying Brier score, computed for each aggressive lymphoma type, for the prognostic index, the machine learning approach, and the simple model. Only patients treated with curative intent were included. The integrated Brier score (IBS) was computed and is shown in each plot window.

References

[1] S. H. Swerdlow, E. Campo, S. A. Pileri, N. L. Harris, H. Stein, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, 127(20):2375–2390, 2016.

[2] The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *New England Journal of Medicine*, 329(14):987–994, 1993.

[3] P. Solal-Céligny, P. Roy, P. Colombat, J. White, J. O. Armitage, et al. Follicular Lymphoma International Prognostic Index. *Blood*, 104(5):1258–1265, 2004.

[4] H. Tilly, M. Gomes da Silva, U. Vitolo, A. Jack, M. Meignan, et al. Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 26(Suppl 5):v116–v125, 2015.

[5] Z. Zhou, L. H. Sehn, A. W. Rademaker, L. I. Gordon, A. S. Lacasce, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood*, 123(6):837–842, 2014.

[6] C. Thieblemont, L. Cascione, A. Conconi, B. Kiesewetter, M. Raderer, et al. A MALT lymphoma prognostic index generated from the dataset of the IELSG-19 prospective clinical trial. *Blood*, 130(12):1409–1417, 2017.

[7] B. Arboe, T. C. El-Galaly, M. R. Clausen, P. S. Munksgaard, D. Stoltenberg, et al. The Danish National Lymphoma Registry: coverage and data quality. *PloS One*, 11(6):e0157999, 2016.

[8] D. Hasenclever, V. Diehl, J. O. Armitage, D. Assouline, M. Björkholm, et al. A Prognostic Score for Advanced Hodgkin's Disease. *New England Journal of Medicine*, 339(21):1506–1514, 1998.

[9] J. J. Castillo, E. S. Winer, and A. J. Olszewski. Population-based prognostic factors for survival in patients with Burkitt lymphoma: an analysis from the Surveillance, Epidemiology, and End Results database. *Cancer*, 119(20):3672–3679, 2013.

[10] P. Morel, A. Duhamel, P. Gobbi, M. A. Dimopoulos, M. V. Dhodapkar, et al. International prognostic scoring system for Waldenström macroglobulinemia. *Blood*, 113(18):4163–4170, 2009.

[11] E. Hoster, M. Dreyling, W. Klapper, C. Gisselbrecht, A. van Hoof, et al. A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood*, 111(2):558–565, 2008.

[12] L. Arcaini, M. Lazzarino, N. Colombo, S. Burcheri, E. Boveri, et al. Splenic marginal zone lymphoma: a prognostic model for clinical use. *Blood*, 107(12):4643–4649, 2006.

[13] J. Biccler, S. Eloranta, P. de Nully Brown, H. Frederiksen, M. Jerkeman, et al. Simplicity at the cost of predictive accuracy in diffuse large B-cell lymphoma: a critical assessment of the R-IPI, IPI, and NCCN-IPI. *Cancer Medicine*, 7(1):114–122, 2018.

[14] M. Federico, M. Bellei, L. Marcheselli, S. Luminari, A. Lopez-Guillermo, et al. Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project. *Journal of Clinical Oncology*, 27(27):4555–4562, 2009.

[15] A. Reddy, J. Zhang, N. S. Davis, A. B. Moffitt, C. L. Love, et al. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell*, 171(2):481–494.e15, 2017.

[16] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860, 2008.

[17] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5):1–13, 2011.

[18] E. O. Ogundimu, D. G. Altman, and G. S. Collins. Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology*, 76:175–182, 2016.

[19] B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, et al. mlr: machine learning in R. *Journal of Machine Learning Research*, 17(170):1–5, 2016.

[20] G. Ambler, R. Z. Omar, and P. Royston. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Statistical Methods in Medical Research*, 16(3):277–298, 2007.

[21] H. Hung and C.-T. Chiang. Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics*, 38(1):8–26, 2009.

[22] T. A. Gerds and M. Schumacher. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040, 2006.

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-221-4

AALBORG UNIVERSITY PRESS