

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/86285> holds various files of this Leiden University dissertation.

Author: Rüten-Budde, A.J.

Title: Personalised medicine for multiple outcomes : methods and application

Issue Date: 2020-03-10

Personalised medicine for multiple outcomes: methods and application

Anja Juana Rüten-Budde

The research in this thesis was supported by the Dutch Cancer Society (DCS) - KWF Kankerbestrijding [UL2015-8028].

Cover art by Niels D.S. Langeveld.

Personalised medicine for multiple outcomes: methods and application

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 10 maart 2020
klokke 16.15 uur

door

Anja Juana Rüten-Budde
geboren te Bremen
in 1990

Promotores:	Prof. dr. E. Eliel	Universiteit Leiden
	Prof. dr. H. Putter	Leids Universitair Medisch Centrum
Copromotor:	Dr. M. Fiocco	Universiteit Leiden, Leids Uni- versitair Medisch Centrum, Prinses Máxima Center

Samenstelling van de promotiecommissie:

Voorzitter:	Prof. dr. P. Stevenhagen	Universiteit Leiden
Secretaris:	Prof. dr. A. W. van der Vaart	Universiteit Leiden
	Prof. dr. H. Gelderblom	Leids Universitair Medisch Centrum
	Dr. F. Ieva	Politecnico di Milano
	Dr. R. Geskus	Oxford University

Contents

1	Introduction	1
§1.1	Introduction to survival analysis	2
§1.2	Frailty models	4
§1.3	Competing risks models	5
§1.4	Multi-state models	7
§1.5	Dynamic prediction	8
§1.6	AUC and C-index	11
§1.7	Multiple imputation	12
§1.8	Soft tissue sarcoma data	14
§1.9	Personalised sarcoma care app	15
§1.10	Outline of thesis	17
2	Investigating hospital heterogeneity with a competing risks frailty model	19
§2.1	Introduction	20
§2.2	Competing risks model	21
§2.3	Frailty model	22
§2.4	Competing risks frailty model	22
§2.4.1	Frailty decomposition	22
§2.4.2	Model estimation	23
§2.4.3	Implementation	24
§2.4.4	Estimation of the standard error	25
§2.4.5	Empirical Bayes estimates	26
§2.5	Data application	27
§2.5.1	Data description	27
§2.5.2	Competing risks model with independent frailties	29
§2.5.3	Competing risks model with correlated frailties	30
§2.5.4	Empirical Bayes estimates	31
§2.6	Simulation	36
§2.7	Discussion and Conclusion	42
	Appendix	
§2.A	Probabilities for E-step	44
§2.B	Observed information of regression parameters	46

3	Assessment of predictive accuracy of an intermittently observed binary time-dependent marker	51
§3.1	Introduction	52
§3.2	Time-specific AUC for binary marker	54
§3.2.1	Incident cases and dynamic controls	55
§3.2.2	Cumulative cases and dynamic controls	56
§3.2.3	AUC for Weibull illness-death model	57
§3.2.4	Estimation	58
§3.2.5	Estimation of incident/dynamic AUC	58
§3.2.6	Estimation of cumulative/dynamic AUC	59
§3.3	Illness-death models	59
§3.3.1	Cox model with time-dependent covariate	59
§3.3.2	Piecewise-constant model accounting for interval-censoring	60
§3.3.3	Weibull model accounting for interval-censoring	60
§3.3.4	M-spline model accounting for interval-censoring	61
§3.4	Simulation	61
§3.5	Application	70
§3.6	Discussion	72
	Appendix	
§3.A	Derivation of AUC	74
§3.A.1	Incident/dynamic AUC	74
§3.A.2	Cumulative/dynamic AUC	74
§3.B	Results for all scenarios	76
§3.C	Incident/dynamic AUC for scenarios D–L	79
§3.D	Cumulative/dynamic AUC for scenarios D–L	83
4	Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model	89
§4.1	Introduction	91
§4.2	Patients and methods	92
§4.2.1	Statistical analysis	93
§4.3	Results	94
§4.4	Discussion	100
5	A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)	103
§5.1	Introduction	105
§5.2	Methods	105
§5.2.1	Study population	106
§5.2.2	Study design	106
§5.2.3	Statistical analysis	107
§5.3	Results	108
§5.4	Discussion	118
	Appendix	

§5.A Competing risks model	120
6 Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma	123
§6.1 Introduction	124
§6.2 Methods	125
§6.2.1 Study design	125
§6.2.2 Patients and variables	125
§6.2.3 Statistical analysis	126
§6.3 Results	127
§6.4 Discussion	135
7 External Validation and Adaptation of a Dynamic Prediction Model for Patients with High-Grade Extremity Soft Tissue Sarcoma	137
§7.1 Introduction	139
§7.2 Methods	139
§7.2.1 Study design	139
§7.2.2 Patients and Variables	140
§7.2.3 Statistical analysis	141
§7.3 Results	142
§7.4 Discussion	149
8 Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model	151
§8.1 Introduction	153
§8.2 Methods	154
§8.2.1 Measures	154
§8.2.2 Statistical analysis	155
§8.2.3 Missing data	155
§8.3 Results	156
§8.4 Discussion	162
§8.5 Conclusion	164
Appendix	
§8.A Patient characteristics	165
§8.B Complete case analysis	165
9 Discussion and future perspectives	169
§9.1 Custom-made models	169
§9.2 Prediction models in clinical practice	172
§9.3 Future perspectives	173
Bibliography	177
Summary	193

Samenvatting	197
List of Publications	203
Curriculum Vitae	205
Acknowledgements	207

CHAPTER 1

Introduction

Statistical analysis aims to find patterns in data and to increase understanding of such data. The beauty of statistics is that it can be applied to a great variety of fields to answer relevant research questions. The particular field of statistics called survival analysis is where the topics of this thesis find their place. Survival analysis deals with life-time data. In this type of data the time from a specific starting point until an event of interest occurs are recorded. In medical research for example, time from diagnosis of disease until death could be studied. What characterizes life-time data, also called survival data, is that it is generally incomplete. Some individuals in the data might not have experienced the event of interest at the end of the study period or have dropped out of the study before the event has occurred. These data are called *right-censored*. The event time is unknown, it is known however, that the event had not occurred before the last observation time. To handle this particular type of missing data, and other similar types, special methodology is necessary summarized under the term survival analysis. Even though survival analysis is relevant to a large number of applications, the works in this thesis are all motivated by medical research. For this reason, examples are given in the context of clinical research.

Survival analysis is used by clinicians to identify risk factors associated with the occurrence of a clinical event of interest. For example in cancer research, clinicians use survival models to investigate if a patient's age, sex, tumor size, and other variables are associated to the risk of death. To describe the evolution of disease complex mathematical models are required. Patients may experience several disease related events in different orders. *Multi-state models* can be applied in such context. Another extension of survival models is to add a random effect, also called *frailty*. Frailty terms are used to model unobserved covariates which might have an effect on the event of interest. In all studies not all relevant patient or disease characteristics can be collected and therefore the survival model is incomplete. Random effects quantify the so called unobserved heterogeneity resulting from an incomplete model.

Survival models may be used to investigate the effect of risk factors on clinical events of interest and to predict survival probabilities. Such predictions inform both patients and clinicians of a patient's prognosis and may help in the shared decision making process. Prediction models are available for a variety of diseases and there is a demand for more and more sophisticated models. Ordinary prediction models are often limited to a single prediction time point. This means that predictions can only be made at a particular time, such as at time of diagnosis of disease. When a

patient comes back for a follow-up visit, such models are not able to provide accurate predictions. A patient may experience disease related events over time which are not taken into account by a model that considers only risk factors known at diagnosis or at start of treatment. *Dynamic prediction* models provide updated predictions from different time points during follow-up. They are able to include updated information as it becomes available. A simple idea to create dynamic prediction models is through the *landmarking* approach. Predictions are made from a chosen landmark time point by using a subset of the data consisting of patients still alive at that time. Multiple landmark times can be chosen to make predictions from different time points during follow-up.

The remainder of this chapter introduces basic concepts of survival analysis as well as more complex models that are used in this thesis. The following Section provides an introduction to survival analysis and explains simple survival models. Section 1.2, 1.3, 1.4, 1.5 introduce frailty models, competing risks models, multi-state models, and dynamic prediction models, respectively. In Section 1.6 and 1.7 the C-index and multiple imputation are explained, respectively. Section 1.8 and 1.9 introduce the motivating soft tissue sarcoma data set and the developed prediction tool, respectively. The last Section gives an overview of the remaining chapters of this thesis.

§1.1 Introduction to survival analysis

The concepts and definitions of this Section are introduced as in Klein and Moeschberger [92], which is referred to for further reading.

Survival analysis developed from the need to analyse life-time data. The structure of such data can be of different kind and often motivates the development of new methods. A first step in understanding survival concepts is in understanding the data it has to deal with.

The subject of study is the event time T . In medical research, T could represent the time from diagnosis until death. The event time for an individual may not be observed, if he dropped out of the study early, or the study ended before the event of interest occurred, or another event occurred. Denote by C the right censoring time for an individual. This is the last time a subject was observed in the study. The information observed for an individual is $\tilde{T} = \min(T, C)$, the minimum between right-censoring and event time, and $\delta = \mathbb{1}(T \leq C)$, the event time indicator. $\delta = 1$, if the event time was observed and $\delta = 0$, if it was not. Survival models for right-censored data assume that the event time T and the right-censoring time C are independent, sometimes conditional on covariates.

For other types of events, the exact event time cannot be observed directly. In cancer care for example, after removal of the tumor a patient attends scheduled follow-up visits where he is screened for recurrence of disease. If recurrence is found it is only known that it had occurred between the last negative screening and the first positive screening. The time until recurrence is *interval-censored*.

To study the distribution of the survival time T different parameters are studied. The most prominent function of interest is the survival function $S(t) = P(T > t)$,

which at time t is equal to the probability of being event-free at time t . The survival function is usually modelled by the hazard function

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

The hazard function at time t is equal to the conditional probability of experiencing the event in the next instant conditional on being event-free just before time t . The survival function can be defined in terms of the hazard function,

$$S(t) = \exp\left(-\int_0^t \lambda(u) du\right),$$

and it can be estimated nonparametrically with the *Kaplan-Meier estimator* [88]. Let $t_1 < t_2 < \dots < t_D$ be the ordered event times, d_i the number of events at time t_i , and Y_i denote the number of individuals at risk at time t_i . The Kaplan-Meier estimator of the survival function is given as follows

$$\hat{S}(t) = \begin{cases} 1, & \text{if } t < t_1 \\ \prod_{t_i \leq t} \left(1 - \frac{d_i}{Y_i}\right), & \text{if } t_1 \leq t. \end{cases} \quad (1.1.1)$$

An example of survival data set and the corresponding Kaplan-Meier curve are shown in Figure 1.1. Subject 2 and 10 are right-censored. The Kaplan-Meier curve changes at event times and remains unchanged at censoring times. The censoring times however affect the size of the jump the curve makes.

The effect of a covariate vector \mathbf{Z} is most commonly modelled with the Cox proportional hazards model [44] in which the hazard is defined as

$$\lambda(t \mid \mathbf{Z}) = \lambda_0(t) \exp(\beta^T \mathbf{Z}),$$

where $\lambda_0(t)$ is the baseline hazard and β is the vector of regression coefficients. In the Cox model, the effect of covariates is assumed to be multiplicative on the nonparametric baseline hazard. Let $t_1 < t_2 < \dots < t_D$ be the ordered event times, $\mathbf{Z}_{(i)}$ the covariates of the individual who experiences the event at time t_i , \mathbf{Z}_j the covariates of individual j , and $R(t_i)$ denote the set of individuals still at risk at time t_i . The vector of regression coefficients β is estimated, assuming all event times are distinct, by maximising the partial likelihood

$$L(\beta) = \prod_{i=1}^D \frac{\exp(\beta^T \mathbf{Z}_{(i)})}{\sum_{j \in R(t_i)} \exp(\beta^T \mathbf{Z}_j)},$$

and the baseline hazard $\lambda_0(t)$ can then be computed using the Breslow estimator [34].

The covariates \mathbf{Z} discussed so far are time-fixed and known at the time origin. However covariates can also change over time, like blood values which are repeatedly measured. Let $\mathbf{Z}(t)$ be a vector of time-dependent covariates, whose values change over time. The Cox model with time-dependent covariates is defined as

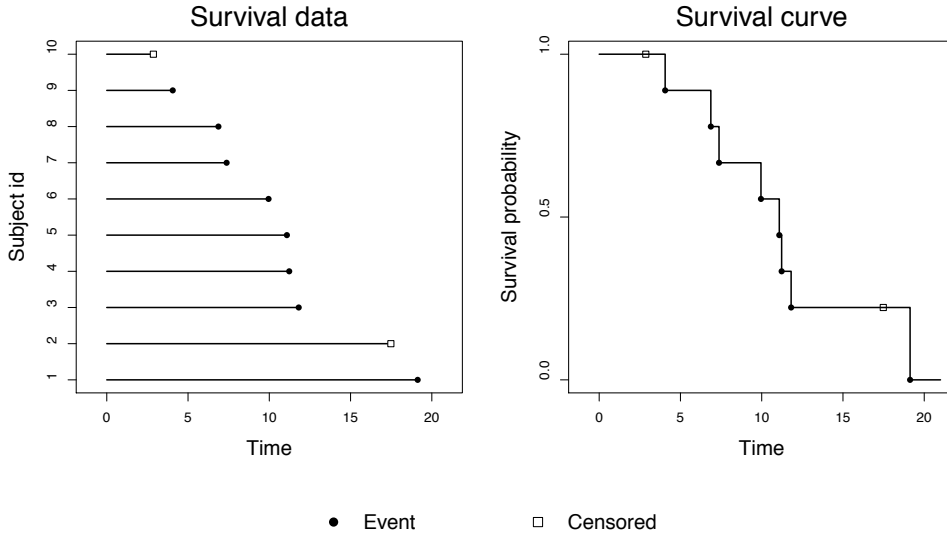


Figure 1.1: Left panel: survival data for 10 subjects. Right panel: the corresponding conditional Kaplan-Meier survival estimate.

$$\lambda(t | \mathbf{Z}(t)) = \lambda_0(t) \exp(\beta^T \mathbf{Z}(t)).$$

The partial likelihood for this Cox model is defined analogously to the model with only time-fixed covariates. Let $\mathbf{Z}_{(i)}(t_i)$ be the covariate vector at time t_i of the individual who experiences the event at time t_i , $\mathbf{Z}_j(t)$ the covariate vector of individual j at time t , and $R(t_i)$ the set of individuals still at risk at time t_i . Again assuming all event times are distinct, the vector of regression coefficients β is estimated by maximising the partial likelihood

$$L(\beta) = \prod_{i=1}^D \frac{\exp(\beta^T \mathbf{Z}_{(i)}(t_i))}{\sum_{j \in R(t_i)} \exp(\beta^T \mathbf{Z}_j(t_i))}.$$

§1.2 Frailty models

Survival regression models aim to explain the differences of survival times between individuals using covariate information. If the model is perfectly specified, the remaining variation reflects the randomness of the event time, conditional on the covariate values. However, often not all relevant covariates can be included in the model. The variation of survival time accounted for the missing covariates in the model is called unobserved heterogeneity. The effect of unobserved heterogeneity on the event time is called frailty [157].

In survival analysis, frailty can be modelled by a random effect included in a survival model. The variance of the random effect is a measure of the amount of unobserved heterogeneity. The frailty variable can be chosen subject specific or it can be shared for clusters of individuals. In univariate frailty models, a subject specific frailty models unobserved heterogeneity on the individual level. In multivariate frailty models, a shared frailty variable is used for a cluster of individuals which models unobserved heterogeneity on the cluster level.

The random frailty variable can be incorporated in a survival model with a multiplicative effect on the hazard. The cluster i specific frailty W_i has a multiplicative effect on the hazard,

$$\lambda(t | W_i) = W_i \lambda_0(t),$$

where $\lambda_0(t)$ is the baseline hazard. Often $E(W_i) = 1$, then $Var(W_i)$ describes the extent of unobserved heterogeneity. A univariate frailty model has cluster size equal to 1. In this case, the estimated frailty variance represents the unobserved heterogeneity between individuals. For cluster size bigger than 1 the estimated frailty variance represents the unobserved heterogeneity between clusters. The effect of a covariate vector \mathbf{Z} can be modelled by using a Cox model with frailty term

$$\lambda(t | W_i, \mathbf{Z}) = W_i \lambda_0(t) \exp(\beta^T \mathbf{Z}),$$

where $\lambda_0(t)$ is the baseline hazard and β are the regression coefficients. The frailty terms W_i are iid random variables with a specific distribution. The gamma distribution is a popular choice as frailty distribution due to its mathematical properties. An additional assumption of the frailty model is that censoring does not depend on the frailty [109].

§1.3 Competing risks models

In some applications, more than one type of terminal event is possible, such as in the study of different causes of death. A competing risks model is described by a starting state in which individuals are event-free and several end states, also referred to as causes of failure, see Figure 1.2.

The survival data of an individual has a different structure. Let T_1, T_2, \dots, T_J be the event times of J competing events and C the independent right-censoring time. For an individual, only the minimum of the first event or right-censoring time $T = \min(C, T_1, T_2, \dots, T_J)$ is observed together with an indicator $\delta = 0, 1, \dots, J$ indicating the cause of failure or censoring ($\delta = 0$).

A fundamental concept used in competing risks analysis is the *cause-specific hazard*,

$$\lambda_j(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \delta = j | T \geq t)}{\Delta t},$$

where $j, (j = 1, \dots, J)$ is one of the competing events. The cause-specific hazard represents the conditional probability of experiencing event j in the next instant

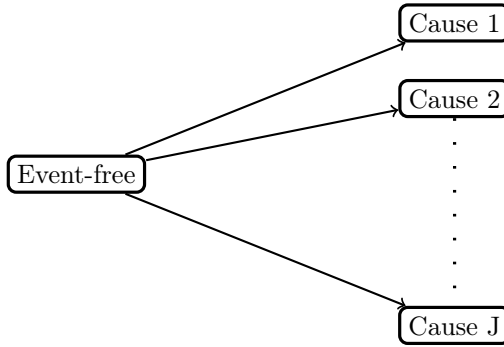


Figure 1.2: Competing risks model with J causes of failure.

given still event-free just before time t . Define the cumulative cause-specific hazard by

$$\Lambda_j(t) = \int_0^t \lambda_j(u) du.$$

A Cox model can be used to model the cause-specific hazard together with the effect of covariates \mathbf{Z} ,

$$\lambda_j(t | \mathbf{Z}) = \lambda_{j0}(t) \exp(\beta_j^T \mathbf{Z}),$$

where $\lambda_{j0}(t)$ and β_j are the cause-specific baseline hazard and the regression coefficients, respectively.

A quantity of interest, in particular in applications with competing risks, is the cumulative incidence function. This function corresponds to the probability of experiencing an event j before time t ,

$$I_j(t) = P(T \leq t, \delta = j) = \int_0^t \lambda_j(u) S(u) du,$$

where $S(u) = P(T > u)$ is the probability of being event-free at time u . In this context the survival function depends on all cause-specific hazards,

$$S(t) = \exp\left(-\sum_{j=1}^J \Lambda_j(t)\right),$$

where $\Lambda_j(t)$ is the cause-specific cumulative hazard at time t . The cumulative incidence function therefore not only depends on cause j but also on the cause-specific hazards of all the other causes.

In the competing risks setting, the cause-specific cumulative incidences are often the quantities of interest to answer questions such as, what is the probability of a recurrence of disease within a certain time frame. The Cox model can be used

to model covariate effects on the cause-specific hazards, however the effects on the cause-specific cumulative incidences are not straightforward, since they depend on all other cause-specific hazards simultaneously. Another approach to model the effect of risk factors was introduced by Fine and Gray [57]. The covariate effects are modelled directly on the cause-specific cumulative incidence through the *subdistribution hazard*,

$$\bar{\lambda}_j(t) = -\frac{d \log(1 - I_j(t))}{dt}.$$

To model covariate effects a proportional hazards model analogous to the Cox model was proposed,

$$\bar{\lambda}_j(t | \mathbf{Z}) = \bar{\lambda}_{j0}(t) \exp(\beta_j^T \mathbf{Z}).$$

This model can be estimated with a partial likelihood approach like the Cox model. The regression coefficients from Fine and Gray's model have an intuitive interpretation because they are regressed on the cause-specific cumulative incidence directly and therefore can be easily interpreted clinically.

§1.4 Multi-state models

Competing risks models extend standard survival models by adding more end states. Multi-state models allow multiple end points as well as transition states [119]. Figure 1.3, represents a particular multi-state model referred to as illness-death model. An individual starts in state 0, he can then move to state 1, which can represent a disease he may experience and subsequently move to state 2, death.

The transitions from state i to state j are modelled by the *transition hazard*,

$$\lambda_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

where T denotes the time of reaching state j from state i . The types and times of the occurrence of events of an individual define his path through the multi-state model. A common assumption, is that the multi-state model is a Markov model. Given the present state and the event history (the trajectory through the multi-state model so

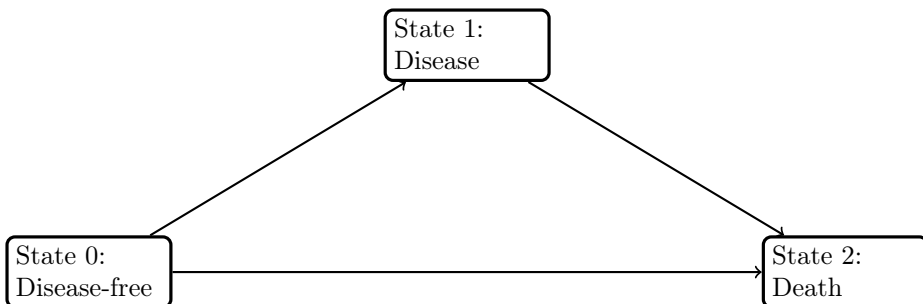


Figure 1.3: Illness-death model.

far) the next state to move to as well as the time of transition only depends on the present state.

To model the effect of covariates on the transition hazards in a Markov model the Cox model can be used,

$$\lambda_{ij}(t | \mathbf{Z}) = \lambda_{ij0}(t) \exp(\beta_{ij}^T \mathbf{Z}),$$

where $\lambda_{ij0}(t)$ and β_{ij} are the transition-specific baseline hazard and regression coefficients, respectively. Here the baseline hazards are transition-specific. One could however choose a subset of baseline transition hazards to be proportional to each other. For example in Figure 1.3, the two transitions towards state 2 (the death state) could be assumed proportional. This model would be similar to a single end point Cox model with the disease state as a binary (0, 1) time-dependent covariate. The multi-state model however would have the additional benefit of simultaneously modelling the rate of occurrence of the time-dependent covariate.

Similar to the competing risks setting the covariate effects are difficult to interpret. The effects act on the transition hazard which is the conditional probability of moving to state j at time t , given in state i just before t . From those transition hazards however, clinically more relevant and accessible quantities can be computed, for example, probabilities of future events. Particularly, the conditional probabilities of future events given a patients event history and covariate information \mathbf{Z} . In a Markov multi-state model instead of the entire event history only the current state is of relevance. For example in cancer care, given that a patient with particular characteristics is recurrence-free one year past surgery, the probability of getting a recurrence within the next 5 years may be of interest. Estimating these probabilities, also referred to as making predictions, is very relevant to patients and clinicians and can be used in the shared decision making process. The probabilities are expressed as $P_{ij}(u, t | \mathbf{Z})$, which represents the conditional probability of being in state j at time t given that the subject is in state i at time u . These probabilities can be computed from the transition hazards [119].

§1.5 Dynamic prediction

The previously discussed survival models may be used to model disease progression and to find risk factors for events of interest. They are also used to predict probabilities of future events. Prediction models are gaining popularity in the medical field, where they are used to inform clinician and patient about a patient's future prognosis. Survival estimates can help in the shared decision making process between patient and clinician. Many prediction models use baseline covariates; covariates measured before the time origin, such as time of diagnosis or time of treatment.

Some disease markers however, are measured and updated during follow-up. For example, blood values could be measured regularly, or recurrence of disease could be diagnosed during follow-up. Updating predictions over time with new information is defined as dynamic prediction. Dynamic prediction models can provide survival predictions at different time points, using all available information at that time point.

Among the models previously discussed only multi-state models are able to use updated information, if it can be expressed by an intermediate state. For example, the information of disease recurrence (yes vs. no) could be modelled by a state in a multi-state model.

Another approach to dynamic prediction is to use a landmark model [13, 149, 151]. Landmarking was introduced to avoid a common mistake in the analysis of tumor response. Patients were treated with chemotherapy and then followed to see whether the tumor responded to treatment. A common malpractice was to make two groups of patients, responders and nonresponders and to compare their survival times from start of treatment. This is known in literature as *immortal time bias*. The groups are compared from time of treatment, the status of response however, is not yet observed at that time. In this situation information from the future is used. An individual needs to survive long enough for a response to develop and to be observed. Individuals who die early and did not have time to develop a response yet are automatically grouped into the nonresponse group, giving this group an unfair disadvantage. On the other hand, responders must first survive long enough to become a responder and have therefore an unfair survival advantage. In this grouping scheme, responders are immortal until their time of response, hence the name immortal time bias.

A solution proposed by [13] is to evaluate the effect of tumor response with a landmark model. The idea is to choose a specific time point t_{LM} called *landmark* as new time origin for the analysis. At the landmark time the response of patients is evaluated and can be treated as a baseline covariate. The outcome is survival from landmark time and the interpretation of the response variable is response before landmark time. For this analysis only patients who are still in the risk set, meaning alive and in follow-up, are considered. Patients who developed a response before the landmark will be grouped in the response group. Patients who did not have a response before the landmark are grouped in the nonresponse group. Note that patients in the nonresponse group may develop a response later on. Figure 1.4 shows 3 landmark data sets and their corresponding Kaplan-Meier curves. The figures for landmark at time 0 show the original data and corresponding Kaplan-Meier survival curve. At time $t = 0$ no patient has developed a tumor response therefor all patients are grouped in the nonresponse group. The data for landmark at time 5 show that only individuals still in follow-up at that time are selected and that only individuals who developed a response before time 5 are grouped into the response group. The other individuals are grouped in the nonresponse group, even though they may develop a response later on.

In [151] it was suggested to use landmark models to make dynamic predictions. The Cox model is applied to the landmark data set to make predictions from the landmark time t_{LM} up until a prediction horizon $t_{LM} + w$, defined as

$$\lambda(t | \mathbf{Z}, t_{LM}, w) = \lambda_0(t | t_{LM}, w) \exp(\beta_{LM}^T \mathbf{Z}), \quad t_{LM} \leq t \leq t_{LM} + w,$$

where $\lambda_0(t | t_{LM}, w)$ is the baseline hazard for landmark t_{LM} and prediction window w , β_{LM} are the landmark-specific regression coefficients, and \mathbf{Z} are regression coefficients. The Cox landmark model can be used to predict survival probabilities from the landmark time, using the updated information of response status of a patient.

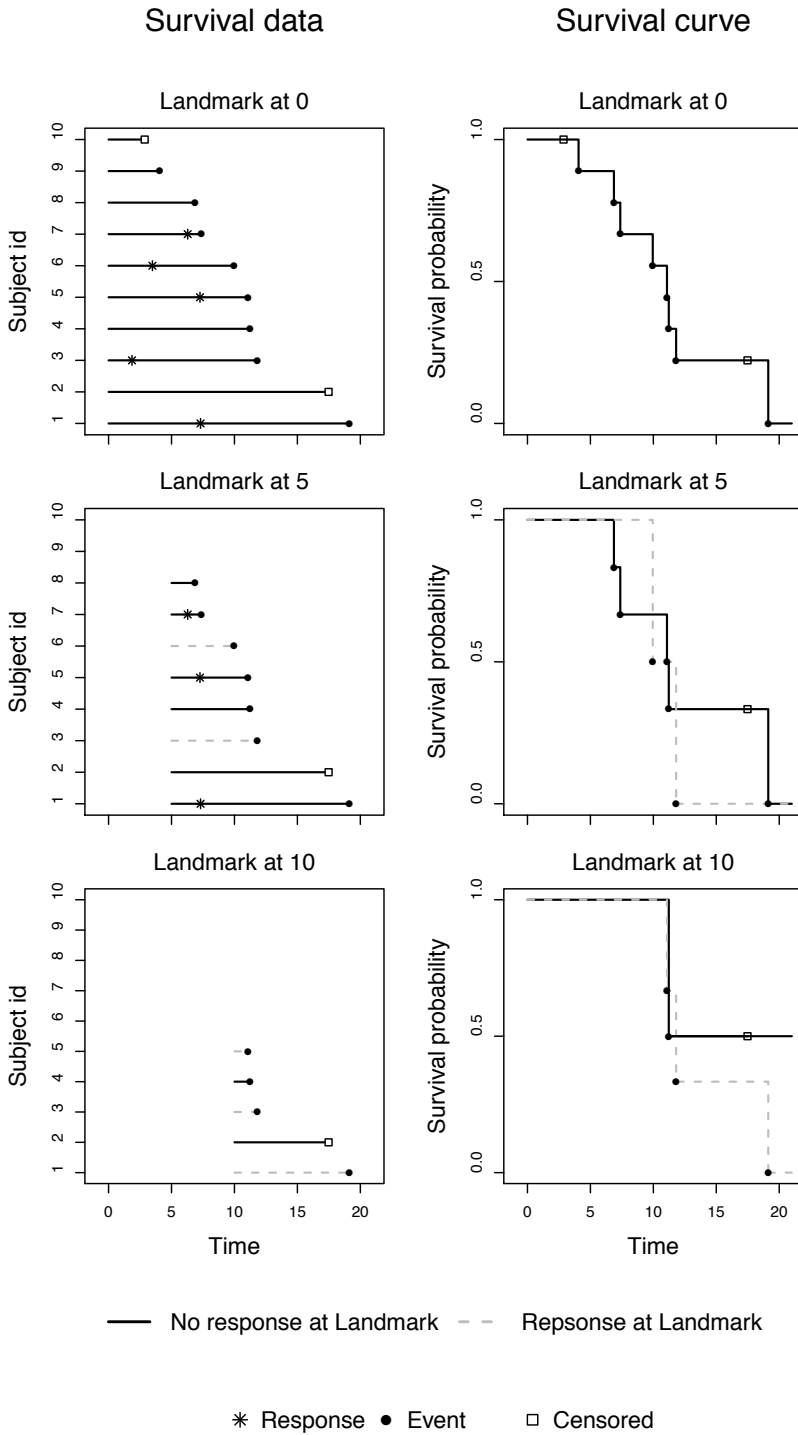


Figure 1.4: Left panel: landmark data sets. Right panel: the corresponding Kaplan-Meier survival estimate.

to make predictions from multiple landmarks one can choose to create a different landmark model for each landmark time, or combine them to a landmark *supermodel* [151].

§1.6 AUC and C-index

Survival predictions are important to patients and clinicians. More and more prediction models are becoming available for a variety of diseases. Before such a prediction model is published its clinical importance must be investigated. Does the model provide accurate predictions? To validate a prediction model different prediction aspects are considered [150].

Existing methods, such as sensitivity and specificity, were extended to survival analysis by [79], [80], and [171]. Sensitivity and specificity were originally defined for a binary outcome B , where $B = 1$ is considered a ‘case’ and $B = 0$ is considered a ‘control’. For a covariate X and a classification criterion c a simple prediction rule is to predict individuals to be cases if $X > c$ and otherwise controls. The correct classification rates, sensitivity(c) = $P(X > c \mid B = 1)$ and specificity(c) = $P(X \leq c \mid B = 0)$, summarize the accuracy of this classification rule.

A graphical summary which illustrates the whole range of sensitivity and specificity for different values of c is the Receiver Operation Characteristic (ROC) curve which plots sensitivity against 1-specificity, illustrating the difference of the marker distribution between cases and controls. In case the marker distributions are the same, the ROC curve lies on the 45 degree line, which indicates that the marker does not contribute in distinguishing cases from controls. A summarizing measure of concordance between marker and outcome, which can be used to measure predictive accuracy is the Area Under the ROC Curve (AUC).

Heagerty and Zheng (2005) [80] extended the concepts of sensitivity and specificity to survival analysis by defining time-specific ‘cases’ and ‘controls’. They give several different definitions, where particularly their *incident cases* and *dynamic controls* is of interest. At time t incident cases are those individuals that experience the event at time t and dynamic controls are those individuals who survive beyond time t . The time-specific AUC based on this definition of cases and controls is a time-specific measure of discrimination.

Discrimination refers to how well a model can distinguish between high and low risk individuals. A prediction model discriminates well if it predicts high risk for individuals who experience the event earlier and lower risk for individuals who experience the event late or not at all during the follow-up.

A weighted average of the time-specific incident/dynamic AUC coincides with Harrel’s *concordance index* (C-index), which is a popular measure of model discrimination [149, 78]. It was originally defined as the proportion of evaluable ordered pairs for which prediction and outcome are concordant. Ordered pairs are individuals (i, j) where the observation time of individual i is shorter or equal to the observation time of individual j . An ordered pair is evaluable if it was observed that i experienced the event of interest before j . Pairs in which i was censored before j experienced the event or was censored are not evaluable, since it is unknown who experienced the

event first. Concordant pairs are pairs in which the prediction model predicted higher risk for the individual experiencing the event of interest earlier. In the case that both patients have the same predicted risk 0.5 is added to the count of concordant pairs instead of 1. The C-index is computed by dividing the number of concordant pairs by the number of evaluable pairs. The values of the C-index are between 0 and 1, where a value of 0.5 indicates no discriminative ability.

The incident/dynamic AUC measures the ability of the model to discriminate at a particular time t . The C-index measures the overall ability of the model to discriminate between individuals. Considering dynamic predictions, one might be interested in the ability of the model to discriminate within a time window $[t, t + w]$. A dynamic C-index can be obtained by computing the proportion of ordered pairs for which prediction and outcome are concordant, only for individuals who are at risk at time t and considering event times censored at time $t + w$ [149]. For predictions made at time t , this index measures the discriminative ability of the model within a time window $[t, t + w]$.

In the competing risk setting, [165] proposed a different concordance index by defining evaluable and concordant patient pairs differently. For the event of interest, they define an ordered pair as evaluable if the first patient experiences the event at a time at which the second patient is still at risk. The risk set at time t in this case is made of patients who did not yet experience any event and are still in follow-up and those individuals who experienced a competing event. Furthermore, an ordered evaluable pair is defined as concordant if the first patient to experience the event of interest has a higher predicted risk than the other patient. In case that both patients have the same predicted risk 0.5 is added instead of 1 to the count of concordant pairs.

§1.7 Multiple imputation

Most statistical methods cannot be applied when missing information are present in the data. By default many statistical programs will remove observations with missing values and analyse only complete observations. This approach reduces the amount of subjects and therefor the power of the statistical tests and can lead to biased results in some cases [100].

Multiple imputation is a general approach to handling missing data which uses all available information, even for subjects with missing values [125, 126]. The method increases statistical power and reduces bias compared to a complete case analysis. The idea behind imputation of missing values is to generate likely values for the missing values to create a complete data set. Multiple imputation uses an imputation model to generate multiple complete data sets. For these data sets the observed values are the same, however the missing values are different. Statistical methods can be applied to each complete data set and results can be pooled using Rubin's rule [125].

The concept of multiple imputation was introduced by Rubin [125]. The idea is to draw m values for each missing value from the posterior predictive distribution of the missing values under a Bayesian model for the data and the missing-data mechanism. From the resulting m complete data sets m complete-data statistics $\hat{Q}_1, \dots, \hat{Q}_m$ and

the corresponding variance-covariance U_1, \dots, U_m can be computed and combined. The estimate from an analysis using multiple imputation can be computed by averaging the complete-data estimates

$$\bar{Q} = \frac{1}{m} \sum_{l=1}^m \hat{Q}_l.$$

The variance-covariance of \bar{Q} is equal to

$$T = \bar{U} + \frac{m+1}{m} B,$$

where

$$\bar{U} = \frac{1}{m} \sum_{l=1}^m U_l,$$

and

$$B = \frac{1}{m-1} \sum_{l=1}^m (\hat{Q}_l - \bar{Q})(\hat{Q}_l - \bar{Q})^T.$$

The terms \bar{U} and B correspond to the within-imputation variability and the between-imputation variability.

Throughout this thesis the package *Amelia II* was used to generate multiple imputations [82]. The assumptions for the imputation model in *Amelia II* are that the data are missing at random and that the complete data are multivariate normal. Missing at random means that the distribution of missingness only depends on the observed data. The multivariate normal distribution may only crudely approximate the true data distribution, however there is evidence that it works well even for categorical or mixed data [82, 129, 130]. The imputation models employed in this thesis included all variables used in the analysis together with the event status. Categorical variables were modelled as such, using the `noms` option of the `amelia` function and age and size were modelled using a square root transformation, so that no negative values could be imputed. For categorical covariates `amelia` determined the number of categories p and substituted them with $p-1$ binary variables to specify each category. These variables were treated as if they were continuous variables and missing information received continuous imputations. Those were scaled into probabilities for each category and from the resulting multinomial distribution one category was drawn so that the original multinomial variable is reconstructed.

The number of imputations was chosen to be equal 5 in Chapter 6 and 8 and 10 in Chapter 7.

§1.8 Soft tissue sarcoma data

Soft tissue sarcomas (STS) are a rare type of cancer that make up approximately 1% of all adult cancers [137]. It begins in the bodies connecting tissues, such as muscle, fat, blood vessels, nerves, tendons and the lining of joints [3]. It can appear anywhere in the body but most commonly in the extremities (about 60% of STS cases) [41]. The standard treatment for primary STS patients is surgical removal of the tumor and potentially (neo)adjuvant radiotherapy or chemotherapy [56].

Figure 1.5 shows the different disease states a patient may follow. After surgery a patient can remain disease free, or develop local recurrence (LR) or distant metastasis (DM), or die. A LR is diagnosed if evidence of tumor at the previously treated tumor bed is found, while DM is diagnosed if spread of tumor is found at another location.

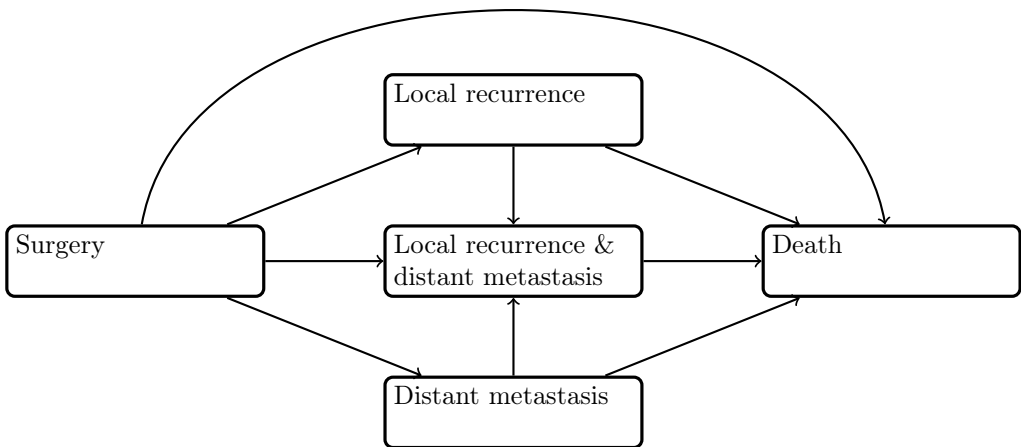


Figure 1.5: Soft tissue sarcoma data.

Because of the rarity of STS, many studies conducted have been subject to small sample size and large heterogeneity in the study population [53, 102]. Some prognostic factors for survival, such as histology, grade, depth and size were generally recognized [83, 117, 169, 53, 73, 146, 140, 98, 139, 29]. The effect of surgical margin and LR however, was long unclear [140, 106, 164, 74, 110, 102, 111]. Surgical margin is the amount of healthy tissue that is removed surrounding the tumor during primary surgery. Generally, it is desired to remove the tumor surrounded by healthy tissue, but this can be challenging depending on the tumor location. The effect of surgical margin has been of great interest, because of the effect on the quality of life after surgery.

The lack of an established prognostic profile for STS patients motivated a group of researchers to start an interdisciplinary project between the Leiden University Medical Center and the Mathematical Institute of Leiden University in 2016. The aim was to collect STS data on an international scale and to develop statistical models

to investigate the effect of risk factors on clinical outcomes, with particular interest in surgical margin, as well as to provide reliable survival predictions for patients. The funding granted by the Dutch Cancer Society (DCS) - KWF Kankerbestrijding allowed this project to become reality. The work in this thesis is the result of this collaboration.

Clinical data was collected retrospectively over the period of this project by contacting tertiary centers treating high-grade STS patients of the extremities. Patients were selected based on histological subtype, if they were treated surgically with curative intent from 2000 on, and if they had high-grade disease (as defined by FNCLCC larger than grade 2 [145]). All the collaborating centers adhered to the guidelines of the European Society for Medical Oncology for follow-up [56]. Chapter 4 is the first published article resulting from this collaboration. For each Chapters 5, 6, 7 more data has been added.

§1.9 Personalised sarcoma care app

For clinical decision making a patients prognosis always has played an important role. Reliable survival predictions are an important information to clinicians to consider in the patient care. Nowadays particularly in cancer care, the clinical community embraces the concept of shared decision making. In the shared decision making approach, the patient is involved in the choice of treatment. An important information, for this process are a patient's survival predictions. Prediction models have become popular in the clinical world [1, 6, 7]. Their increase and availability for various disease reflects the demand.

To support the shared decision making process, a prognostic prediction model, the PERsonalised SARcoma Care (PERSARC) model, for patients with high-grade STS of the extremities was developed in Chapter 5 [20]. It predicts from time of surgery a patient's probability of developing LR and survival. The model was internally validated by means of discrimination and calibration. To make predictions accessible to clinicians a mobile application was developed, which is available in the Apple and Google Play store [4, 5]. In the app patient specific characteristics can be entered and predictions are returned, see Figure 1.6 for illustration.

A group of researchers from Leiden University Medical Center was granted funding from the Dutch Cancer Society (KWF) to implement shared decision making for high-grade soft tissue sarcoma patients in the Netherlands. The goal is to ensure that soft tissue sarcoma patients receive personalised care, in which risks and benefits of treatment options and patient preferences are balanced. Part of the implementation strategy is the introduction of the PERSARC app to clinical practice.

An updated version of the PERSARC app is expected to be released in 2020. PERSARC version 2.0 will be able to provide dynamic predictions of survival as described in Chapter 7. Predictions will be made taking LR and DM events, that occur during follow-up, into account.

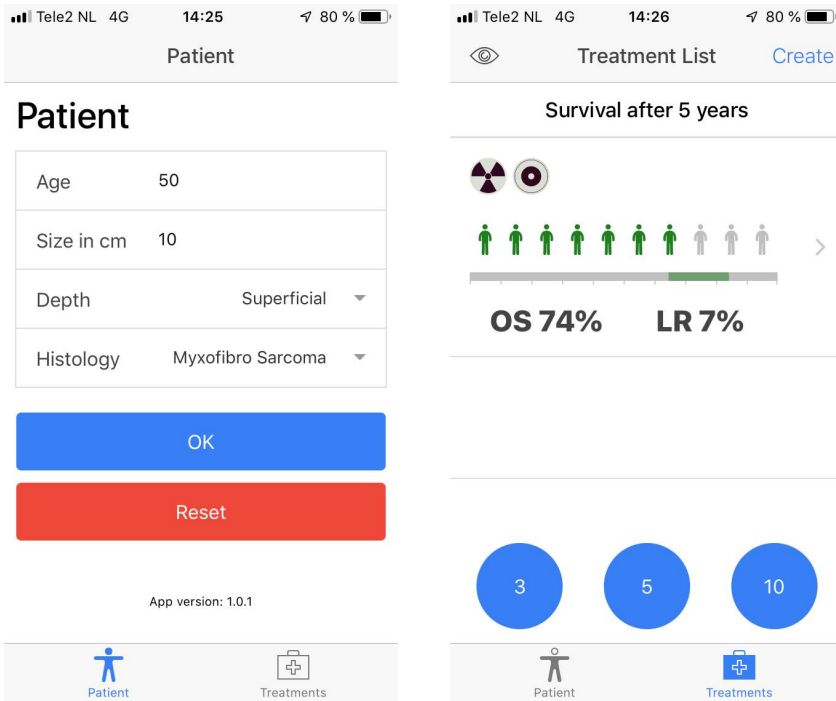


Figure 1.6: PERSARC app. Left panel: patient characteristics interface, with information of a 50 year old patient, with a 10 cm superficial tumor of type myxofibro sarcoma. Right panel: his prediction of survival and probability of recurrence within 5 years, with radiotherapy treatment and a margin of 0.1-2mm.

§1.10 Outline of thesis

The main objective of this thesis was to develop clinically relevant survival models for patients with high-grade STS of the extremities, in particular the development and validation of prediction models for use in clinical practice. The interdisciplinary collaboration between the Mathematical Institute of Leiden University and the Leiden University Medical Center resulted in important contributions to the care of STS patients. Each chapter is briefly summarized below.

In Chapter 2 [17] a novel frailty model is proposed for multi-center data with two competing risks. Frailty variables are used to model unobserved heterogeneity on the hospital level; they could be interpreted as the "hospital effect" on the competing events. The frailty model developed models the hospital effects on the competing events to be correlated within each hospital.

In Chapter 3, which is to be submitted to *Statistics in Medicine* [16], the effect of interval censoring is studied on the predicted accuracy of a binary disease marker. The motivation comes from cancer care. After surgery a patient is regularly screened for LR and DM. Once a recurrence is diagnosed, however, it is only known that it occurred between the last negative and the first positive screening. In this chapter we investigate through simulations how the assessment of predictive accuracy of recurrence is affected by the intermittent screening process.

Chapter 4 [21] was the first in a series of publications based on the growing soft tissue sarcoma data set. A data set of 687 patients was analysed with a multi-state model. The effect of risk factors on LR and DM/Death were studied, with particular interest in the effect of surgical margin.

Chapter 5 [20] is the continuation of the STS project, with a data set of 766 patients. Prediction models for survival and probability of local recurrence were developed and implemented in the PERSARC mobile application. The models are internally validated in terms of calibration and discrimination.

In Chapter 6 [19] a dynamic prediction model based on 2232 STS patients was developed. A landmark supermodel was used to provide predictions of additional 5-year survival from different prediction time points during follow-up. Disease related events, LR and DM, are used to update predictions over time and covariates were investigated for time-varying effects. The model was internally validated.

In Chapter 7, which is to be submitted to *Surgical Oncology* [18], the previously developed dynamic prediction model for STS patients is updated and externally validated. The updated model is based on 3826 patients and now includes grade as additional covariate in the model. It was externally validated using a cohort of 1111 patients and it is implemented in the updated PERSARC mobile application.

In Chapter 8 [14] a multi-state model was developed for 982 Ewing sarcoma patients. Adverse events in the multi-state model were LR, DM of the lungs, DM at other locations, and death. The effect of risk factors was studied with particular interest in surgical margins, histological response, and radiotherapy.

In Chapter 9 previous chapters are put in broader perspective and future research directions are suggested.

Investigating hospital heterogeneity with a competing risks frailty model

This chapter has been published in *Statistics in Medicine* 38(2) (2018) 269–288 as A.J. Rueten-Budde , H. Putter and M. Fiocco, "Investigating hospital heterogeneity with a competing risks frailty model" [17].

Abstract

Survival analysis is used in the medical field to identify the effect of predictive variables on time to a specific event. Generally, not all variation of survival time can be explained by observed covariates. The effect of unobserved variables on the risk of a patient is called frailty. In multicenter studies, the unobserved center effect can induce frailty on its patients, which can lead to selection bias over time when ignored. For this reason, it is common practice in multicenter studies to include a random frailty term modeling center effect. In a more complex event structure, more than one type of event is possible. Independent frailty variables representing center effect can be incorporated in the model for each competing event. However, in the medical context, events representing disease progression are likely related and correlation is missed when assuming frailties to be independent. In this work, an additive gamma frailty model to account for correlation between frailties in a competing risks model is proposed, to model frailties at center level. Correlation indicates a common center effect on both events and measures how closely the risks are related. Estimation of the model using the expectation-maximization algorithm is illustrated. The model is applied to a data set from a multicenter clinical trial on breast cancer from the European Organisation for Research and Treatment of Cancer (EORTC trial 10854). Hospitals are compared by employing empirical Bayes estimates methodology together with corresponding confidence intervals.

§2.1 Introduction

Survival data arises where interest lies on the time from a specific time origin until occurrence of an event of interest. Prominent applications are found in the medical field, where e. g. time from diagnosis of disease until death could be studied. What distinguishes survival analysis from other types of statistical analysis is the type of data it deals with: it is generally incomplete. Since it takes time to observe an event, it is usually not possible to collect complete information. A popular method to model the effect of covariates on risk of event occurrence is through the semi-parametric Cox proportional hazards model [44]. In some situations more than one type of end point are possible, when e. g. different causes of death are studied. Analogous to the single endpoint situation the Cox model can be used to model the effect of covariates on the cause-specific transition hazards [119] of each cause of failure. A more complicated event structure with intermediate states can be modeled by a multi-state model [119]. Dependence in survival data can be modeled by a random frailty term, which models heterogeneity between observations or between clusters of observations. The frailty term represents unobserved covariates on the individual or cluster level that act on the risk of event occurrence. The frailty variance can be interpreted as a measure of heterogeneity between clusters or individuals, however it can also be seen as a measure of dependence within a cluster.

Multicenter studies are a common strategy to collect sufficient data for a clinical study. Patients are clustered within treatment centers and possible correlation between patients within a center can be modeled by using a shared frailty model. Shared frailty models are able to model dependence, however these models limit the unobserved covariates modeled by the frailty to have the same effect within a cluster. In the presence of competing events the use of one frailty per center acting on all causes of failure is questionable. Similarly using J independent frailties per center one for each cause of failure does not yield a complete picture of the data structure. Frailties for different competing events within a center are likely to be correlated, since they represent the same unobserved covariates on cluster level. Yashin et al. [166] first introduced a correlated gamma frailty model to analyze twin survival data. They decompose a twin's frailty into a sum of two independent frailties, one of which is shared by both twins. Petersen et al. [116] use this idea of adding frailty components, which act multiplicatively on the individual hazard and describe more complex variance components models for survival data.

Clustered data in the presence of competing risks further complicate possible dependence structures and different approaches are taken. Extensions of Fine and Gray's subdistribution hazard model [57] incorporate a frailty term to model cluster dependence on the cumulative incidence function of the event of interest in the presence of competing events [89, 131, 51, 172]. Wienke et al. [161, 163] analyze correlated frailty models in the presence of competing risks, however assuming independence between risks. The assumption of independence is questionable since related events (e. g. events representing disease progression) might be influenced similarly by the same unobserved covariates. Wienke et al. [162] extend the bivariate correlated gamma frailty model of Yashin et al. [166] to model dependence among competing

risks based on parametric marginal survival functions. Gorfine and Hsu [72] combine frailty components multiplicatively to model dependence between competing risks for clustered survival data. Liquet et al. [99] analyze hospital heterogeneity in multi-state models using independent and joint frailty models to model dependence between transition intensities. Rotolo et al. [123] propose to incorporate correlated frailties in multi-state models acting on the transition-specific hazard functions. They construct frailties by combining a common cluster component and a transition-specific component multiplicatively.

In this paper we propose an additive gamma frailty model which acts multiplicatively on the cause-specific hazard to model dependence within clusters and between two competing events. The method can be used to investigate hospital heterogeneity in a competing risks setting. An elegant estimation procedure using the EM-algorithm is outlined as well as a strategy to calculate the standard error of the estimates. In contrast to Wienke et al. [162] who model dependence among competing risks by using a parametric approach our method is based on the semi-parametric Cox model [44]. Compared to methods suggested by Gorfine and Hsu [72] and Rotolo et al. [123] which combine frailty components multiplicatively in this article a gamma decomposition is proposed to model dependence between risks. The advantage of our method is its simplicity in construction and estimation, which is based on the mathematical properties of the gamma distribution. Additionally, estimation through the EM-algorithm provides empirical Bayes estimates for each center's frailty, which can be used to compare centers.

In Section 2.2 and 2.3 the cause-specific hazards model and frailty model will be reviewed briefly. The proposed competing risks frailty model is presented in Section 2.4. In Section 2.5 the method is applied to a data example and corresponding results are presented. A simulation study to investigate the performance of the correlated frailty model is discussed in Section 2.6. A discussion follows in Section 2.7.

§2.2 Competing risks model

Competing risks models are used when more than one type of failure is possible. An example is the study of different causes of death. A fundamental concept in competing risks is the cause-specific hazard. It is the hazard of failing from a particular cause given still event free at that time.

For right censored survival times the cause-specific hazard of cause j for a subject i with covariate vector \mathbf{X}_i is as follows:

$$\lambda_j(t|\mathbf{X}_i) = \lambda_{j0}(t)e^{\beta_j^T \mathbf{X}_i}, \quad (2.2.1)$$

where λ_{j0} is the cause-specific baseline hazard for cause j and β_j assesses the effect of the covariates \mathbf{X}_i on the progression rate to cause j [119]. Here the effects of covariates are quantified on the cause-specific hazard and not on the marginal hazard. Only if the censoring due to the competing risks is non-informative conditionally on the covariates in the model, the estimates can also be interpreted as effects on the marginal hazard.

§2.3 Frailty model

The concept of frailty introduces random effects in survival models, which represent the presence of unobserved heterogeneity. The variance of this random component is a measure used to quantify heterogeneity in the data. Vaupel et al. [157] discussed univariate frailty models with a gamma distribution and applied this concept to survival. Clayton [40] used frailties in the multivariate analysis of chronic disease incidence in families.

A frailty is an unobserved random factor varying over the population of individuals, which is assumed to have a multiplicative effect on the hazard of a single individual or a group or cluster of individuals. In univariate frailty models each individual has its own independent frailty, while in shared frailty models clustered individuals share a common frailty.

For subject i with covariate vector \mathbf{X}_i belonging to cluster k with frailty W_k the hazard is given as

$$\begin{aligned}\lambda(t|\mathbf{X}_i, W_k) &= W_k \lambda_0(t) e^{\beta^T \mathbf{X}_i} \\ &= \lambda_0(t) e^{\beta^T \mathbf{X}_i + \log(W_k)}.\end{aligned}\tag{2.3.1}$$

A convenient choice for the frailty distribution is the gamma distribution, since its posterior distribution given survival data, stays in the gamma family [116].

§2.4 Competing risks frailty model

Heterogeneity between centers in a competing risks setting can be modeled by assigning each center J frailties, one for each cause of failure. The J frailty terms within a center may be chosen to be independent, however the effects within a center are likely to be related which is ignored in such a model. In a more realistic model frailties within a center are correlated. A model for the dependence structure was first proposed by Yashin et al. [166] in a twin study, decomposing the frailty of each twin as a sum of two independent frailties one of which is shared. Petersen et al. [116] use an additive variance components structure on multiplicative gamma frailty models and outline its estimation. The correlated frailty model proposed in this article follows their approach.

§2.4.1 Frailty decomposition

In the following, let W_{k1}, W_{k2} denote the frailty variables corresponding to two causes of failure within hospital k ($k = 1, \dots, K$). Correlation between frailties is constructed by decomposing each frailty as the sum of two independent gamma distributed variables, one of which is common in both frailties [59, 58]. For cause j ($j = 1, 2$), frailties are given as

$$W_{kj} = \frac{Z_{k0} + Z_{kj}}{\nu_0 + \nu_j},\tag{2.4.1}$$

where

$$Z_{k0} \sim \Gamma(\nu_0, 1), \quad Z_{kj} \sim \Gamma(\nu_j, 1). \quad (2.4.2)$$

The random variables Z_{k0} , Z_{k1} , Z_{k2} are independent and from now on referred to as the independent frailty components of hospital k . This results in the following frailty distribution:

$$W_{kj} \sim \Gamma(\nu_0 + \nu_j, \nu_0 + \nu_j). \quad (2.4.3)$$

The expectation of the frailty variables is equal to one, which corresponds to no hospital effect or the average hospital effect. Their variance and correlation are given as

$$\text{Var}(W_{k1}) = \frac{1}{\nu_0 + \nu_1} = \xi_1, \quad \text{Var}(W_{k2}) = \frac{1}{\nu_0 + \nu_2} = \xi_2, \quad (2.4.4)$$

$$\text{Cor}(W_{k1}, W_{k2}) = \nu_0(\xi_1\xi_2)^{1/2} = \rho. \quad (2.4.5)$$

This construction allows for positive correlation only. In many practical situations however it may be justified to disregard negative correlation, e. g. when competing events describe disease progression. A further restriction is that not all variance correlation combinations are possible in this construction. A large correlation does not allow the variances to be too different, or equivalent, different frailty variances do not allow the correlation to be (almost) one:

$$\nu_1 = \frac{1}{\xi_1} - \frac{\rho}{\sqrt{\xi_1\xi_2}} > 0$$

$$\sqrt{\xi_2/\xi_1} > \rho \quad (2.4.6)$$

$$\nu_2 = \frac{1}{\xi_2} - \frac{\rho}{\sqrt{\xi_1\xi_2}} > 0$$

$$\sqrt{\xi_1/\xi_2} > \rho. \quad (2.4.7)$$

From (2.4.6) and (2.4.7) it follows that $\rho < \min(\sqrt{\xi_2/\xi_1}, \sqrt{\xi_1/\xi_2})$.

§2.4.2 Model estimation

Model parameters are obtained by maximizing the log-likelihood function based on the observed data. Since frailties associated to different centers and individuals across hospitals are independent, the likelihood is the product of hospital likelihoods. For simplicity only the log-likelihood and necessary quantities of a single center k are given in the following.

Denote by n_k and d_{kj} the number of patients and the number of patients that fail from cause j ($j = 1, 2$) in hospital k respectively. Let \mathbf{X}_{ki} , t_{ki} and δ_{ki} ($\delta_{ki} = 0, 1, 2$) be

the covariate vector for patient i treated at hospital k , the event or censoring time and the event or censoring indicator respectively. In the following let β_j be the vector of regression coefficients, λ_{j0} the baseline hazard and Λ_{j0} the cumulative baseline hazard for cause j ($j = 1, 2$). If the frailties were observed the complete data yields the following log-likelihood for hospital k

$$\begin{aligned} \ell_k(\beta_1, \beta_2, \lambda_{10}, \lambda_{20}) = & \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=1} \left\{ \log \left(\frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1} \right) + \log(\lambda_{10}(t_{ki})e^{\beta_1^T \mathbf{X}_{ki}}) \right\} \quad (2.4.8) \\ & + \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=2} \left\{ \log \left(\frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2} \right) + \log(\lambda_{20}(t_{ki})e^{\beta_2^T \mathbf{X}_{ki}}) \right\} \\ & - \frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1} \sum_{i=1}^{n_k} \Lambda_{10}(t_{ki})e^{\beta_1^T \mathbf{X}_{ki}} - \frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2} \sum_{i=1}^{n_k} \Lambda_{20}(t_{ki})e^{\beta_2^T \mathbf{X}_{ki}} \\ & + \log(f(Z_{k0}, Z_{k1}, Z_{k2})), \end{aligned}$$

where f is the probability density function of the independent and gamma distributed frailty components.

Integrating out all frailty components specific to each center in the log-likelihood yields the observed data log-likelihood, which is computationally challenging to maximize (see Appendix 2.A for details). Considering the unobserved frailties as missing information yields a typical application of the expectation maximization algorithm (EM-algorithm) [50].

§2.4.3 Implementation

For fixed parameter $\nu = (\nu_0, \nu_1, \nu_2)$, the estimation procedure uses the expectation maximization algorithm (EM-algorithm) to approximate the observed data log-likelihood to find optimal regression coefficients and baseline hazards [50]. The approximated observed data log-likelihood is then employed in a three dimensional search to a find maximum likelihood estimate for ν .

Since ν is fixed throughout the EM iterations, the estimation concerns the regression coefficients and baseline hazards only. The conditional expectations of the terms $\log((Z_{k0} + Z_{kj})/(\nu_0 + \nu_j))$, ($j = 1, 2$) and of $\log(f(Z_{k0}, Z_{k1}, Z_{k2}))$ given observed data are irrelevant to the estimation of the complete data case (2.4.8). Therefore the E-step reduces to the calculation of the conditional expectations of the frailties $W_{kj} = (Z_{k0} + Z_{kj})/(\nu_0 + \nu_j)$, ($j = 1, 2$) given observed data. As a result, defining $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki})e^{\beta_j^T \mathbf{X}_{ki}}$, ($j = 1, 2$), it is sufficient to consider

$$\begin{aligned} E(Z_{k0}|\text{data}_k) &= \int_{z_{k0}} z_{k0} f(z_{k0}|\text{data}_k) dz_{k0} & (2.4.9) \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{d_{k1} + d_{k2} + \nu_0 - l - m}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)} \end{aligned}$$

$$\begin{aligned} E(Z_{k1}|\text{data}_k) &= \int_{z_{k1}} z_{k1} f(z_{k1}|\text{data}_k) dz_{k1} & (2.4.10) \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{l + \nu_1}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1}\right)} \end{aligned}$$

$$\begin{aligned} E(Z_{k2}|\text{data}_k) &= \int_{z_{k2}} z_{k2} f(z_{k2}|\text{data}_k) dz_{k2} & (2.4.11) \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{m + \nu_2}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)}, \end{aligned}$$

where f is the conditional probability density function of a frailty component given data, and $c_k(l, m, \nu_0, \nu_1, \nu_2)$ is a function over the number of events of each type of failure for fixed frailty parameters. Details about the computations are outlined in the Appendix 2.A.

Since the conditional distributions of the frailty components Z_{k0}, Z_{k1}, Z_{k2} given observed data are mixtures of gamma distributions (see Appendix 2.A for details), it is straightforward to compute the quantities (2.4.9)-(2.4.11). Notably the factor $c_k(l, m, \nu_0, \nu_1, \nu_2)$ is the same in all three expectations.

The M-step consists of estimating the updated baseline hazards $\Lambda_{10}(t)$, $\Lambda_{20}(t)$ and coefficient vectors β_1 , β_2 , through maximization of the conditional log-likelihood, given frailties estimated in the E-step. This can be done with existing software, e.g. using `coxph()` from the R [122] package `survival` [142], incorporating the logarithm of the expected frailties as `offset` into the cause-specific hazards model. The algorithm iterates over these two steps and stops once the approximation of the observed data log-likelihood converged (e.g. change of smaller than 1e-06).

Until now, the frailty parameter $\nu = (\nu_0, \nu_1, \nu_2)$ was fixed throughout the EM iterations. Profile likelihood is used to obtain maximum likelihood estimates of $(\nu_0, \nu_1, \nu_2, \beta_1, \beta_2, \Lambda_{10}, \Lambda_{20})$; the function `optim()` is used to find the optimal ν , maximizing the observed data log-likelihood approximated with the EM-algorithm (see the supplementary material in this paper).

§2.4.4 Estimation of the standard error

Louis [101] discussed how to obtain the covariance matrix for the regression parameters, that stays within the EM-algorithm framework, using only derivatives of the complete data log-likelihood. This approach does not yet include the uncertainty caused by estimating the frailty parameters $\nu = (\nu_0, \nu_1, \nu_2)$ outside of the EM-algorithm.

Putter and van Houwelingen [120, supplementary material] proposed estimation as described in the following.

Let $\hat{\eta}(\boldsymbol{\nu}) = (\hat{\beta}_1^T(\boldsymbol{\nu}), \hat{\beta}_2^T(\boldsymbol{\nu}), \hat{\lambda}_{10}^T(\boldsymbol{\nu}), \hat{\lambda}_{20}^T(\boldsymbol{\nu}))^T$ denote the maximum likelihood estimates (MLE) of the regression coefficients and baseline hazards given frailty parameters $\boldsymbol{\nu}$, and denote by $\hat{\boldsymbol{\nu}}$ the MLE of $\boldsymbol{\nu}$ maximizing the observed data log-likelihood. The combined covariance matrix of $\hat{\boldsymbol{\nu}}, \hat{\eta}$ is given as

$$\begin{pmatrix} \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} & \Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}} \left(\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^T \\ \left(\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} + \left(\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} \left(\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^T \end{pmatrix}, \quad (2.4.12)$$

where $\Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}}$ and $\Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}$ are the covariance matrix of $\boldsymbol{\nu}$ and $\hat{\eta}$ respectively and the term $\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ are the partial derivatives of the regression parameters given $\boldsymbol{\nu}$. The term on the bottom right of (2.4.12) represents the covariance of $\hat{\eta}(\hat{\boldsymbol{\nu}})$ where the term $\hat{\eta}(\hat{\boldsymbol{\nu}})$ is obtained using a Taylor expansion of $\hat{\eta}(\boldsymbol{\nu})$ and the score functions of $\hat{\eta}(\boldsymbol{\nu})$ and $\hat{\boldsymbol{\nu}}$ around the MLEs. The off diagonal terms are covariance matrices of $(\hat{\boldsymbol{\nu}}, \hat{\eta}(\hat{\boldsymbol{\nu}}))$ and can be derived in a similarly way, see Appendix 2.B for details.

The term $\Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}}$ is computed from the Hessian matrix obtained using the `hessian()` function from the `numDeriv` package [69] around the point estimate of $\boldsymbol{\nu}$ found by the `optim()` function in R [122]. We proceed by inverting the negative of the Hessian matrix, since the inverse of the observed profile information equals the $\boldsymbol{\nu}$ component of the full observed inverse information evaluated at $(\boldsymbol{\nu}, \hat{\eta}(\boldsymbol{\nu}))$ [167, sec. 8.6.2].

The term $\frac{\partial \hat{\eta}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ is approximated numerically. The derivative around the MLE is estimated by calculating the slope between parameters for values of $\boldsymbol{\nu}$ close to the MLE.

The term $\Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}$ can be computed as described by Louis [101]. It requires the gradient vector and second derivative matrix of the complete data log-likelihood, but not the ones associated to the incomplete data case, see Appendix 2.B for details.

The standard error of the estimated regression parameters $\boldsymbol{\eta}$ can be calculated by taking the square root of the corresponding diagonal elements of the covariance matrix (2.4.12). To obtain the standard error of the frailty variances and correlation we apply the multivariate delta method on $\Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}}$ [39, sec. 5.6]. See the supplementary material for implementation in R.

§2.4.5 Empirical Bayes estimates

Heterogeneity between hospitals may raise the question of hospital ranking based on their frailty or relative performance. A popular method to compare institutions is the empirical Bayes approach introduced to this setting by Thomas et al. [143]. If many centers are involved, a crude center effect estimate may explode for small centers due to large variation and not due to a real center effect [153]. The empirical Bayes estimator helps distinguish observations that are "extreme by nature" and those that are "extreme by chance" and is very well suited for the analysis of quality comparison

data [152]. The empirical Bayes approach not only uses information on a particular center to quantify its performance, but uses information on all centers to help improve the estimate.

Following van Houwelingen [152], the empirical Bayes principle will be outlined. Let X_1, \dots, X_K be independent outcomes with densities $f(x_k, \theta_k)$ and $\theta_1, \dots, \theta_K$ iid with distribution G . The optimal estimator under mean squared error loss for each θ_k is given by the Bayes estimator $d(x_k|G) = E(\theta_k|x_k, G)$, when G is known. When G is unknown, one can estimate $E(\theta_k|x_k, G)$ through an estimate of the distribution G . The resulting estimator is shrunk towards the mean, where the amount of shrinkage depends on the variance of the underlying distribution. In the context of center performance X_k represents the outcome and θ_k the true unobserved performance of center k .

The E-step of the EM-algorithm estimates the empirical Bayes estimate of the center frailties given current model parameters and ν . Hence computing a last E-step based on the MLE of regression parameters and ν after convergence of the algorithm will give the empirical Bayes estimate of center frailties.

Even though empirical Bayes estimates are preferred to crude performance estimates when analyzing quality comparison data, interpretation of results should be made with caution as reasons for different outcome may lie outside a center's responsibility. Statistical issues in comparing institutions are discussed in more detail in Goldstein and Spiegelhalter [71].

The conditional distribution of Z_{k0}, Z_{k1} and Z_{k2} given data is the weighted sum of gamma distributions depending on the number of events of each type (see Appendix 2.A for details). To obtain prediction intervals for the empirical Bayes estimates a simplified sampling procedure is applied.

- 1) Sample from set of tuples (l, m) from $\{(0, 0), \dots, (d_{k1}, d_{k2})\}$, where d_{k1} and d_{k2} are the number of events of type 1 and 2 respectively.
- 2) Sample Z_{k0}, Z_{k1} and Z_{k2} from gamma distributions $\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m, 1 + \Lambda_{k1} + \Lambda_{k2})$, $\Gamma(l + \nu_1, 1 + \Lambda_{k1})$ and $\Gamma(m + \nu_2, 1 + \Lambda_{k2})$ respectively.

Repeating this sampling procedure many times lower and upper confidence limits can be found by taking the 2.5% and 97.5% quantile.

§2.5 Data application

§2.5.1 Data description

The data used in this work originates from the European Organization for Research and Treatment of Cancer (EORTC) trial 10854, which studied the effect of one course of perioperative chemotherapy given directly after surgery on survival [147]. The data set includes 2795 women treated for invasive stage I or II breast cancer, randomized for treatment in 15 different centers. Breast cancer is one of the most common types of cancer in women. The standard treatment for breast cancer is surgery, which may

be followed by chemotherapy, radiotherapy or both. Disease progression after surgery can be described in terms of events a patient might experience. A patient can develop local recurrence (LR), which means that the tumor grows back at the site of surgery and/or might develop distant metastasis (DM), which corresponds to a tumor growth not at the site of surgery and/or she might die.

Patients were excluded from this analysis following exclusion criteria of the trial ($n = 41$) and if information on relevant covariates was missing ($n = 91$). Furthermore, all 5 patients from a particular center were excluded, because of the small amount of patients treated at this center, leaving a total of 2658 patients from 14 different centers for analysis.

The competing risks model for this data is illustrated in Figure 2.1. Two competing events are considered, recurrence of disease (LR or DM) and death. The starting state is the state a patient enters after surgery, being alive with no evidence of disease after surgical removal of the primary tumor (ANED).

Table 2.1: Characteristics of 2658 patients.

Variable	N	(%)
Age		
≥ 50	1602	(60.3)
40–50	762	(28.7)
< 40	294	(11.1)
Tumor size		
$< 2\text{cm}$	798	(30.0)
$\geq 2\text{cm}$	1860	(70.0)
Nodal status		
negative	1407	(52.9)
positive	1251	(47.1)
Surgery		
mastectomy	1164	(43.8)
breast conserving	1494	(56.2)
Perioperative chemotherapy		
yes	1325	(49.8)
no	1333	(50.2)
Adjuvant chemotherapy		
no	2173	(81.8)
yes	485	(18.2)
Adjuvant radiotherapy		
no	54	(2.0)
yes	2604	(98.0)

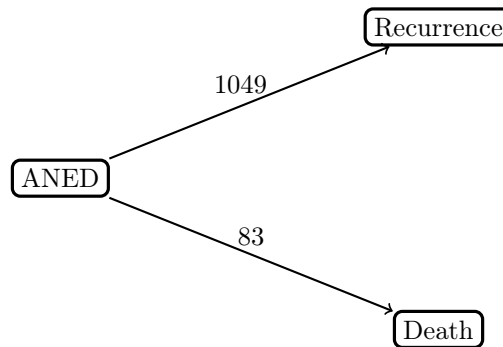


Figure 2.1: Initially, 2658 patients are alive with no evidence of disease (ANED).

The choice of covariates to analyze is based on a previous study on the same data [46]. The following prognostic factors are considered in the analysis: age (≥ 50 , $40 - 50$, < 40), tumor size ($< 2\text{cm}$, $\geq 2\text{cm}$), nodal status (negative, positive), type of surgery (mastectomy, breast conserving), perioperative chemotherapy (yes, no), adjuvant chemotherapy (yes, no), adjuvant radiotherapy (yes, no). Patients' characteristics are provided in Table 2.1.

§2.5.2 Competing risks model with independent frailties

To account for center effect in a cause-specific regression model each cause of failure within a hospital is assigned its own independent frailty.

The model can be estimated similarly to the classical competing risks model, by using `coxph()` together with the `frailty()` function from the R package `survival` [142] or the `emfrail()` function from the `frailtyEM` [27] package. The results of the estimated model with independent gamma frailties are shown in Table 2.2.

A young age (< 40) significantly increases the risks of experiencing recurrence (HR: 1.43; CI: 1.16-1.76), as well as a larger tumor size (HR: 1.41; CI: 1.22-1.64), a positive nodal status (HR: 1.55; CI: 1.34-1.79) and whether or not perioperative chemotherapy and adjuvant chemotherapy was administered (HR: 1.15; CI: 1.02-1.30 and HR: 0.79; CI: 0.64-0.97 respectively). The frailty variance for transition 1 is estimated to be equal to 0.05.

A larger tumor size and a positive nodal status also have a significant effect on death before recurrence with HR: 1.46 (CI: 1.21-1.76) and HR: 2.22 (CI: 1.87-2.63). For death also type of surgery has a significant effect with HR equal to 0.82 (CI: 0.70-0.97) for breast conserving therapy compared to mastectomy. This finding is unexpected and should probably be ascribed to insufficient adjustment for factors relates to choice of primary surgical treatment. The frailty variance for this transition is estimated to be equal to 0.13.

A different frailty model assigns to each hospital a shared frailty term for both causes of failure. Both the independent and shared frailty model are not realistic. The former assumes an independent effect of the unobserved covariates on the two events and the latter assumes them to have the same effect on both events. A model allowing

Table 2.2: Cause-specific hazards model with independent frailties.

	ANED \rightarrow Recurrence		ANED \rightarrow Death	
	HR	0.95 CI	HR	0.95 CI
Age				
≥ 50	1.00		1.00	
40–50	1.00	0.85-1.19	0.84	0.68-1.04
< 40	1.43	1.16-1.76	1.03	0.79-1.34
Tumor size (≥ 2 vs < 2 cm)	1.41	1.22-1.64	1.46	1.21-1.76
NodST (pos. vs neg.)	1.55	1.34-1.79	2.22	1.87-2.63
Surgery (cons. vs mast.)	0.92	0.80-1.05	0.82	0.70-0.97
PeriCT (no vs yes)	1.15	1.02-1.30	1.11	0.96-1.29
AdjCT (yes vs no)	0.79	0.64-0.97	0.82	0.64-1.05
AdjRT (yes vs no)	1.20	0.73-1.98	1.12	0.62-2.00
	Variance	SE	Variance	SE
Frailty	0.05	0.03	0.13	0.06

Abbreviations: NodST (pos. vs neg.), Nodal status (positive vs negative); Surgery (cons. vs mast.), Surgery (breast conserving vs mastectomy); PeriCT, Perioperative chemotherapy; AdjCT, Adjuvant chemotherapy; AdjRT, Adjuvant radiotherapy; ANED, alive with no evidence of disease; CI, confidence interval; HR, hazard ratio; SE, standard error.

for possible correlation between frailties is probably a more accurate representation of reality.

§2.5.3 Competing risks model with correlated frailties

In Table 2.3 the results for the competing risks frailty model with correlated frailties are shown.

The hazard ratios for recurrence are almost unchanged compared to the independent frailty model. However, in the correlated frailty model nodal status and size are the only significant factors. The hazard ratios for death without recurrence are very different from the independent frailty model. This can be explained by the small number of deaths without recurrence in the data set. The variation added by additionally estimating the frailties, increased the standard errors and fewer variables are significant.

The variance of the frailty for transition 1 (ANED \rightarrow Recurrence) is equal to 0.05 with a standard error of 0.03. For transition 2 (ANED \rightarrow Death) the frailty variance is equal to 0.27 with a standard error of 0.22. The correlation of the frailties is estimated to be equal to 0.37 with a standard error of 0.18. Given these frailty variances the maximum correlation between frailties in this model is 0.43 resulting from inequalities (2.4.6) and (2.4.7).

Table 2.3: Cause-specific hazards model with correlated frailties.

	ANED → Recurrence		ANED → Death	
	HR	0.95 CI	HR	0.95 CI
Age				
≥50	1.00		1.00	
40–50	1.00	0.69-1.44	0.35	0.05-2.76
<40	1.42	0.92-2.18	0.62	0.06-6.48
Tumor size (≥ 2 vs <2 cm)	1.41	1.05-1.89	0.96	0.25-3.73
NodST (pos. vs neg.)	1.55	1.15-2.08	1.72	0.47-6.27
Surgery (cons. vs mast.)	0.92	0.70-1.22	0.65	0.18-2.31
PeriCT (no vs yes)	1.15	0.89-1.48	1.14	0.35-3.70
AdjCT (yes vs no)	0.79	0.50-1.27	0.80	0.06-10.08
AdjRT (yes vs no)	1.18	0.81-1.71	0.66	0.12-3.72
	Variance	SE	Variance	SE
Frailty	0.05	0.03	0.27	0.22
	Correlation		SE	
Correlation	0.37		0.18	

Abbreviations: NodST (pos. vs neg.), Nodal status (positive vs negative); Surgery (cons. vs mast.), Surgery (breast conserving vs mastectomy); PeriCT, Perioperative chemotherapy; AdjCT, Adjuvant chemotherapy; AdjRT, Adjuvant radiotherapy; ANED, alive with no evidence of disease; CI, confidence interval; HR, hazard ratio; SE, standard error.

§2.5.4 Empirical Bayes estimates

Figure 2.2 shows the empirical Bayes estimates of the frailties of each center together with 95% prediction intervals, for event recurrence and death. A value equal to 1 implies that there is no center effect. Centers are ordered by number of patients treated. The prediction intervals are computed by sampling from the gamma mixture distribution of frailties and taking 2.5% and 97.5% quantiles as lower and upper limit.

The left panel of Figure 2.2 shows the frailties for the event recurrence for 14 hospitals ordered by number of patients treated. Two hospitals (9 and 11) have a significantly increased risk for their patients to develop recurrence. One hospital (12) has a significantly decreased risk for its patients to develop recurrence. Further one can see that the width of the prediction intervals decrease with a growing number of patients in the hospital.

The right panel of Figure 2.2 shows that one hospital (11) has an increased risk for its patients to move to the state death. One hospital (14) has a marginally significant decreased risk for its patients to die.

To visualize the relation of the frailties within a hospital the empirical Bayes estimates of the two frailties for each center are plotted against each other in Figure 2.3, together with the joint empirical distribution of the frailties for two centers with index 11 and 12.

The hospital effects on a patient can be investigated by looking at the difference in cumulative hazard and cumulative incidence between the hospitals for a particular

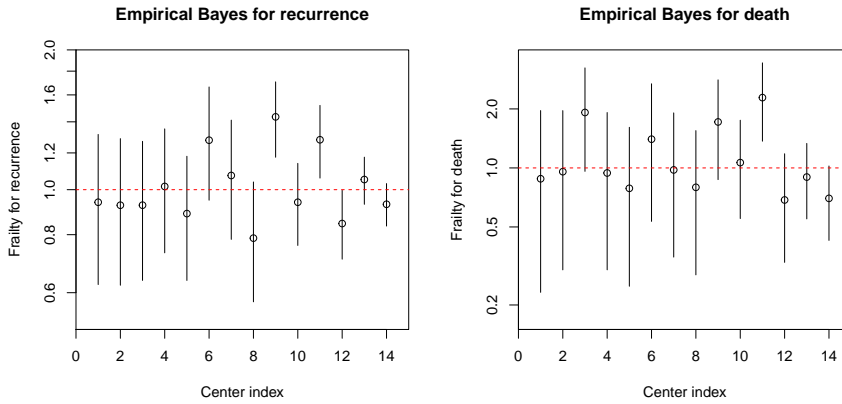


Figure 2.2: Empirical Bayes estimates of frailties and 95% prediction intervals for event recurrence and death of 14 centers, sorted by number of patients.

patient. This is shown in Figure 2.4, for a patient whose covariate values correspond to the mean covariate values in the data.

A pairwise comparison of cumulative incidence curves for an average patient treated in two hospitals further illustrates the difference in effects. This is depicted in Figure 2.5, which shows the stacked cumulative incidence curves for an average patient treated in the two hospitals with the lowest and highest frailty for recurrence. The prognosis shown in the left panel estimates a lower risk for both events, compared to the right panel. This is explained by the estimated correlation between frailties (Table 2.3) and the empirical Bayes estimates of the hospitals (Figure 2.3), which indicate that a hospital with a decreased risk for one cause also has a decreased risk for the other cause. This makes the hospital corresponding to the left panel more appealing.

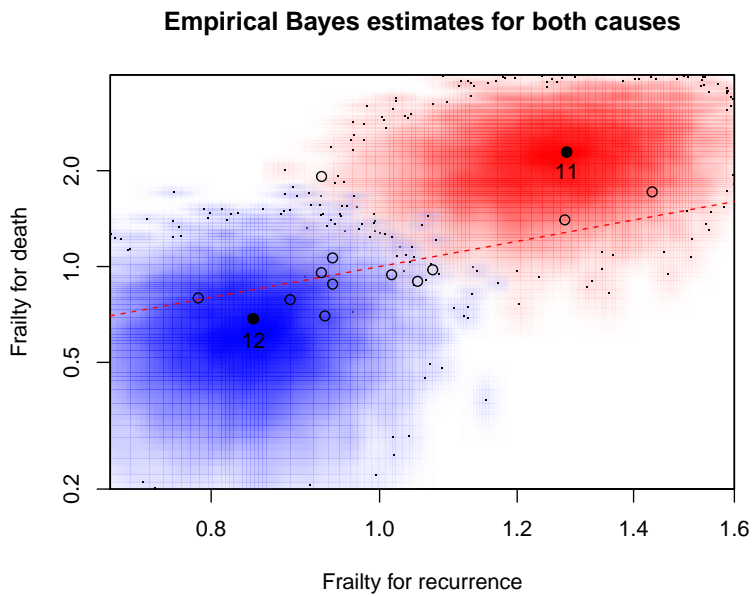


Figure 2.3: Empirical Bayes estimates of frailties for two causes of failure plotted together for 14 centers. For centers with index 11 and 12 the joint empirical distribution of the frailties is shown in red and blue respectively.

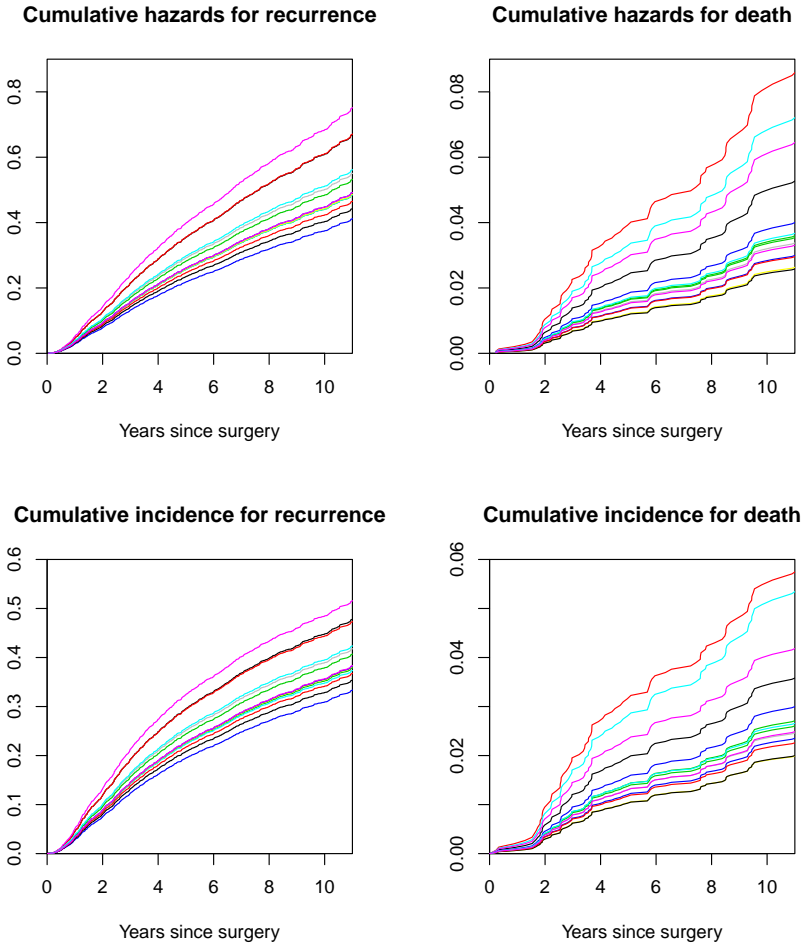


Figure 2.4: Upper panels: cumulative hazards for an average patient for recurrence (on the left) and death (on the right); each line represents a hospital. Lower panels: cumulative incidence of an average patient for recurrence (on the left) and death (on the right); each line represents a hospital.

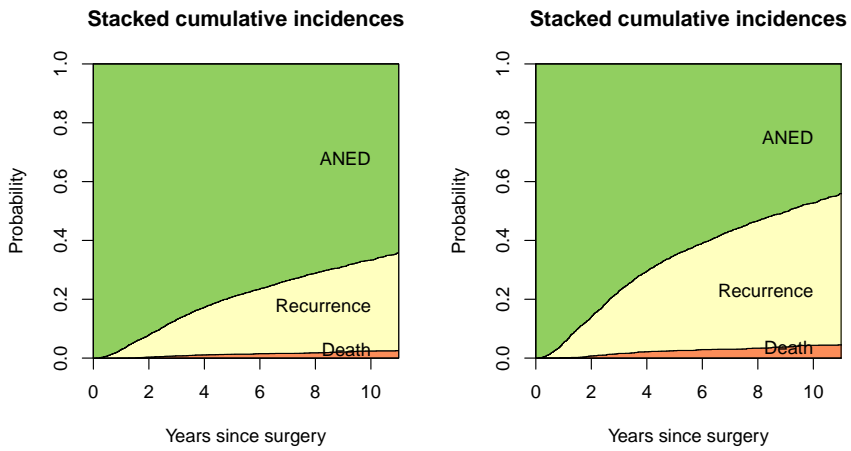


Figure 2.5: Left panel: stacked cumulative incidence curves for an average patient treated in hospital with lowest estimated frailty for recurrence. Right panel: cumulative incidence curves for an average patient treated in hospital with highest estimated frailty for recurrence.

§2.6 Simulation

To investigate the performance of the correlated frailty model a simulation study is conducted. Multiple data scenarios are simulated and the results of the independent and correlated frailty model are compared. Motivated by the data example from Section 2.5 a similar scenario with 2700 patients distributed equally over 15 centers is used for simulation. To study how the number of centers affects the estimation, different scenarios with 5, 30 and 50 centers are considered, while keeping the total number of patients fixed to 2700 (see Table 2.4).

Table 2.4: Scenarios for simulation.

Scenario	n	K	n_k	$\text{Var}(W_{k1})$	$\text{Var}(W_{k2})$	$\text{Cor}(W_{k1}, W_{k2})$	Correlation Bounds
A	2700	5	540	0.25	0.25	0.3	(0, 1)
B	2700	15	180	0.25	0.25	0.3	(0, 1)
C	2700	30	90	0.25	0.25	0.3	(0, 1)
D	2700	50	54	0.25	0.25	0.3	(0, 1)
E	2700	15	180	0.1	0.3	0.8	(0, 0.58)
F	2700	15	180	0.25	0.25	-0.3	(0, 1)

Notation: n , total number of patients; K , number of centers; n_k , number of patients per center; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Survival times are generated by using two Weibull baseline hazards with a common shape parameter a and rate parameters b_1 and b_2 for the two causes of failure respectively. Weibull parameters are fixed throughout the data scenarios and are estimated from the data example of Section 2.5 ($a = 1.01, b_1 = 0.05, b_2 = 0.03$).

Different frailty variance structures are simulated in the different scenarios. Using an additive gamma model as presented in Section 2.4 correlated frailties are sampled with variances equal to 0.25 and correlation equal to 0.3 for scenarios A, B, C and D. As discussed in Section 2.4 different frailty variances by construction do not allow too large correlations; in addition correlation is assumed to be positive to use the proposed method. To study the performance of the method proposed in this article data scenarios E and F which violate these assumptions are simulated. Center and patient distribution are set closest to the data example (15 centers with 180 patients each). Frailties for scenarios E and F in Table 2.4 are sampled from a multivariate lognormal distribution. Scenario E considers a situation in which the correlation is too large to be modeled: frailty variances are equal to 0.1 and 0.3 for cause 1 and 2 respectively while correlation is equal to 0.8. Scenario F represents a situation in which negative correlation is present, with frailty variances equal to 0.25 and correlation equal to -0.3.

Table 2.4 summarizes all scenarios simulated. Censoring times are simulated from a uniform distribution between 9 and 14 years, motivated by the data example.

For each scenario, 1000 data sets are simulated for which two models are estimated: a model with independent frailties for the two causes and the proposed correlated frailty model. Results for frailty variance and empirical Bayes estimates are shown in Table 2.5 and Table 2.6, respectively.

Table 2.5 shows that the independent frailty model generally estimates the frailty

variances to be too high with a large bias and large root-mean-square error (RSME). This seems to be more apparent in data sets with fewer centers.

The correlated frailty model estimates on average results that are closer to the true parameter values with bias of less than half the empirical standard error apart from scenarios E and F. Empirical standard errors are smaller compared to the independent model and are comparable to the average standard error, which even though close, is consistently smaller than the empirical standard error. Root-mean-square errors are generally smaller for the correlated frailty model compared to the independent model.

For scenarios with a larger number of centers better estimation results are obtained. Scenario D with 50 centers per data set shows the best estimation results. Average standard errors are close to the empirical standard errors and RSMEs are small. Scenario E showcases a situation in which the correlation is too large to be modeled with the additive gamma construction. Given frailty variances correlation is restricted to $\rho < \sqrt{0.1/0.3} = 0.58$ (see equations (2.4.6-2.4.7)). The method in this case finds a middle ground and underestimates the frailty variance for cause 2 to allow for a larger correlation. Scenario F considers negative correlation. In this case frailty variances are underestimated, however they are still closer to the true values compared to estimates of the independent model and the correlation estimate is very close to 0.

Table 2.6 shows summary measures of empirical Bayes estimates over the different data scenarios. Bias as well as RMSEs are reported together with coverage probabilities of prediction intervals acquired using the sampling method described in Section 2.4 and studied for each scenario. The number of centers has a stronger effect on the empirical Bayes estimates compared to the frailty variance estimates. Scenario A with only 5 centers shows very poor coverage of the 95% prediction intervals with probabilities of 0.394 and 0.492 for empirical Bayes estimates corresponding to cause 1 and 2 respectively. Scenarios with 15 centers (B, E and F) achieved coverage probabilities between 0.749 and 0.864 and scenarios with more centers (C and D) achieved values between 0.877 and 0.930. Bias and RSME of empirical Bayes estimates appear consistent over different scenarios. To quantify the performance of the method on the estimation of the center-specific cumulative incidence, its bias and RMSE are estimated at quartiles of the theoretical overall event time distribution ($t_1 = 3.55$, $t_2 = 8.48$, $t_3 = 16.85$). The estimates appear unbiased but worsen for the later time t_3 . Interestingly, the bias and RSMEs appear not to be influenced much by the amount of centers and it even appears to become slightly worse if more centers are present in the data. An explanation could be that the estimation of the cumulative incidence becomes more challenging due to the data being generated from many different hazard rates.

For some of the simulated data sets the standard error of the frailty variance and correlation estimate could not be obtained because the hessian matrix obtained during optimization was not positive definite. In this case another attempt was made by starting the optimization of the frailty components from another starting value. This procedure was able to compute results in some cases (see Table 2.7). In case the hessian was not positive definite the data set was discarded. The amount of failed estimation was strongly dependent on the amount of centers in the data set.

Percentages of second attempts and discarded data sets are given illustrated in Table 2.7.

Table 2.5: Frailty variance results of simulation study for 6 different data scenarios.

Scenario	Parameter	Truth	Correlated frailty model			Independent frailty model				
			Mean	(avSE; empSE)	Bias	RMSE	Mean (empSE)	Bias	RMSE	
A	Var (W_{k1})	0.25	0.20	(0.13; 0.15)	-0.05	0.16	0.72	(0.28)	0.47	0.55
	Var (W_{k2})	0.25	0.20	(0.13; 0.15)	-0.05	0.16	0.59	(0.29)	0.34	0.45
	Cor (W_{k1}, W_{k2})	0.3	0.29	(0.20; 0.29)	-0.01	0.29				
B	Var (W_{k1})	0.25	0.24	(0.10; 0.10)	-0.01	0.10	0.57	(0.18)	0.32	0.37
	Var (W_{k2})	0.25	0.24	(0.10; 0.10)	-0.01	0.10	0.42	(0.18)	0.17	0.24
	Cor (W_{k1}, W_{k2})	0.3	0.3	(0.21; 0.22)	0.00	0.22				
C	Var (W_{k1})	0.25	0.25	(0.07; 0.08)	0.00	0.08	0.39	(0.13)	0.14	0.19
	Var (W_{k2})	0.25	0.25	(0.08; 0.08)	0.00	0.08	0.30	(0.12)	0.05	0.12
	Cor (W_{k1}, W_{k2})	0.30	0.33	(0.19; 0.17)	0.03	0.17				
D	Var (W_{k1})	0.25	0.25	(0.06; 0.06)	0.00	0.06	0.29	(0.09)	0.04	0.10
	Var (W_{k2})	0.25	0.24	(0.07; 0.07)	-0.01	0.07	0.25	(0.08)	0.00	0.08
	Cor (W_{k1}, W_{k2})	0.3	0.33	(0.16; 0.15)	0.03	0.16				
E	Var (W_{k1})	0.1	0.09	(0.04; 0.04)	-0.01	0.04	0.18	(0.12)	0.08	0.15
	Var (W_{k2})	0.3	0.19	(0.08; 0.08)	-0.11	0.13	0.41	(0.18)	0.11	0.22
	Cor (W_{k1}, W_{k2})	0.80	0.67	(0.17; 0.15)	-0.13	0.20				
F	Var (W_{k1})	0.25	0.20	(0.08; 0.08)	-0.05	0.20	0.51	(0.18)	0.26	0.32
	Var (W_{k2})	0.25	0.20	(0.08; 0.09)	-0.05	0.10	0.37	(0.19)	0.12	0.22
	Cor (W_{k1}, W_{k2})	-0.3	0.02	(0.05; 0.06)	0.32	0.32				

Abbreviations and notation: empSE, empirical standard error; avSE, average standard error; RMSE, root-mean-square error; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Table 2.6: Empirical Bayes results of simulation study for 6 different data scenarios.

Scenario	Parameter	Bias	RMSE	Coverage	Bias($F(t_1)$)	Bias($F(t_2)$)	Bias($F(t_3)$)	RMSE($F(t_1)$)	RMSE($F(t_2)$)	RMSE($F(t_3)$)
A	W_{k1}	0.10	0.30	0.39	0.00	0.00	-0.04	0.01	0.02	0.05
	W_{k2}	0.10	0.30	0.49	0.00	0.00	-0.02	0.01	0.01	0.03
B	W_{k1}	0.12	0.23	0.75	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.11	0.24	0.82	0.00	0.00	-0.02	0.01	0.03	0.04
C	W_{k1}	0.10	0.23	0.88	0.00	0.00	-0.04	0.02	0.04	0.07
	W_{k2}	0.09	0.25	0.91	0.00	0.00	-0.03	0.02	0.03	0.05
D	W_{k1}	0.09	0.25	0.92	0.00	0.00	-0.04	0.03	0.05	0.07
	W_{k2}	0.08	0.28	0.93	0.00	0.00	-0.02	0.02	0.04	0.06
E	W_{k1}	0.04	0.14	0.86	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.10	0.23	0.80	0.00	0.00	-0.02	0.01	0.02	0.04
F	W_{k1}	0.09	0.20	0.79	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.08	0.22	0.85	0.00	0.00	-0.03	0.01	0.03	0.05

Abbreviations and notation: RMSE, root-mean-square error; Coverage, coverage of 95% prediction intervals; F , cause-specific cumulative incidence; t_1, t_2, t_3 , quartiles of overall event time distribution; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Table 2.7: Failed estimation of standard error.

Scenario	Nbr. of data sets	Successful runs	Success attempt	second	Failed estimation	Evaluated
A	1200	1062	80		138	1000
B	1200	1165	17		35	1000
C	1200	1117	1		83	1000
D	1200	1142	0		58	1000
E	1200	1046	232		154	1000
F	1200	1196	0		4	1000

§2.7 Discussion and Conclusion

Using shared frailty models to account for unobserved covariates in multicenter studies is common practice to avoid bias and to measure the amount of heterogeneity between centers. Correlated frailty models extend the shared frailty model by incorporating dependence structures between related individuals. Dependence among transition intensities of competing risks have come of interest [162, 99, 123].

The model presented uses correlated gamma frailties to model dependence within hospitals and between two competing risks. The mathematical properties of the gamma distribution are exploited to construct and estimate correlated frailties. An estimation procedure using the EM-algorithm is outlined and estimation of the standard error is illustrated. The estimation procedure provides empirical Bayes estimates for hospital frailties, which together with their prediction intervals can be used to compare hospital effects. The model is applied to breast cancer data and a moderate correlation between the frailties of the competing events recurrence and death is estimated. A simulation study is conducted to investigate performance of the method in different situations. Data scenarios with differing number of centers and correlation structures are considered and estimates of a model with independent frailties are compared to the proposed correlated frailty model. The performance of the empirical Bayes estimates obtained by the method was studied under different conditions.

The independent frailty model showed that it is not capable of accounting for center frailty in case of correlation between frailties. The correlated frailty model outperformed it in all data scenarios, concerning estimates as well as size of empirical standard errors. Its estimation benefits from a larger number of centers in the data. In data scenarios with unattainable correlation structures it still performed reasonably well and behaved in an expectable way.

The method is well suited to investigate hospital heterogeneity in the presence of competing risks. It distinguishes between common and separate effects of a hospital on two competing events and performed well in a simulation study. The proposed model can be extended to the case of more than two competing events. Dependence between risks can be modeled by adding frailty components, where shared components induce dependence between risks. However, the model is limited to positive correlation between frailties.

Wienke et al.[162] pointed out that in the case of cause-specific mortality the presence of risk factors might increase the risk of death with respect to all disease, making the case for positive dependence between risks. At the same time he argues that everyone dies eventually, so if the risk of death from one cause is decreased the risk from another cause must be increased, which suggests negative correlation between risks. Further study should be dedicated to the nature of dependencies among competing risks.

Putter and van Houwelingen [121] compare a two-point frailty distribution to a gamma distribution to model association between transition times in multi-state models. An advantage of the two-point frailty model is that it allows the two frailty terms to operate on different scale and that, in contrast to the gamma distribution, it allows negative association. In their simulation study the two-point frailty outperforms the

gamma distribution. A similar model could be used in the competing risks setting modelling dependence between risks, possibly with three or four points.

Acknowledgement

Research leading to this article was supported by the KWF Kankerbestrijding grant UL2015-8028. The European Organisation for Research and Treatment of Cancer (EORTC) is gratefully acknowledged for making data from EORTC trial 10854 available for this analysis.

Appendix

§2.A Probabilities for E-step

Let z_{k0} , z_{k1} , z_{k2} be the independent gamma distributed frailty components and let d_{kj} ($j = 1, 2$) be the number of failures of type j in hospital k ($k = 1, \dots, K$). Defining $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki}) e^{\beta_j^T \mathbf{X}_{ki}}$ ($j = 1, 2$) the conditional probability of the data given frailty components is given as

$$\begin{aligned} f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) &= \prod_{i=1}^{n_k} \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}) \right)^{1_{\{\delta_{ki}=1\}}} \left(\frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}) \right)^{1_{\{\delta_{ki}=2\}}} \\ &\quad \exp \left(- \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \Lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}) + \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \Lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}) \right) \right) \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} z_{k1}^l \right) \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} z_{k2}^m \right) \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right) \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right). \end{aligned}$$

Integrating over the frailty components yields the following conditional probabilities

$$\begin{aligned} f(\text{data}_k | z_{k0}, z_{k2}) &= \int_{z_{k1}} f(z_{k1}) f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) dz_{k1} \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} z_{k2}^m \right) \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \\ &\quad \frac{1}{\Gamma(\nu_1)} \sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \end{aligned}$$

$$\begin{aligned} f(\text{data}_k | z_{k0}) &= \int_{z_{k2}} f(z_{k2}) f(\text{data}_k | z_{k0}, z_{k2}) dz_{k2} \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_1) \Gamma(\nu_2)} \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \\ &\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} z_{k0}^{d_{k1} + d_{k2} - l - m} \end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k0}, z_{k1}) &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} z_{k1}^l \right) \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right) \\
&\quad \frac{1}{\Gamma(\nu_2)} \sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}}
\end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k1}) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}, z_{k1}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_2)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m} \right)} \\
&\quad z_{k1}^l \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)
\end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k2}) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}, z_{k2}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_1)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m} \right)} \\
&\quad z_{k2}^m \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)
\end{aligned}$$

The observed data likelihood is given as

$$\begin{aligned}
f(\text{data}_k) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_1) \Gamma(\nu_2)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} \\
&\quad \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m}}
\end{aligned}$$

The conditional probabilities of the frailty components given the data necessary for the E-step are given as

$$\begin{aligned} f(z_{k0}|\text{data}_k) &= \frac{f(\text{data}_k|z_{k0})f(z_{k0})}{f(\text{data}_k)} \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{d_{k1} + d_{k2} + \nu_0 - l - m}}{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)} z_{k0}^{d_{k1} + d_{k2} + \nu_0 - l - m - 1} \\ &\quad \exp\left(-z_{k0} \left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)\right) \end{aligned}$$

$$\begin{aligned} f(z_{k1}|\text{data}_k) &= \frac{f(\text{data}_k|z_{k1})f(z_{k1})}{f(\text{data}_k)} \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1}\right)^{l + \nu_1}}{\Gamma(l + \nu_1)} z_{k1}^{l + \nu_1 - 1} \exp\left(-z_{k1} \left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1}\right)\right) \end{aligned}$$

$$\begin{aligned} f(z_{k2}|\text{data}_k) &= \frac{f(\text{data}_k|z_{k2})f(z_{k2})}{f(\text{data}_k)} \\ &= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c_k(l, m, \nu_0, \nu_1, \nu_2) \frac{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{m + \nu_2}}{\Gamma(m + \nu_2)} z_{k2}^{m + \nu_2 - 1} \exp\left(-z_{k2} \left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)\right) \end{aligned}$$

where

$$\begin{aligned} \tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2) &= \\ &\binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1}\right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{m + \nu_2}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{d_{k1} + d_{k2} + \nu_0 - l - m}} \\ c_k(l, m, \nu_0, \nu_1, \nu_2) &= \frac{\tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2)}{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2)}. \end{aligned}$$

§2.B Observed information of regression parameters

The term $\Sigma_{\eta\eta} = I_{\eta\eta}^{-1}$ can be computed as described by Louis [101].

Let ℓ^* and ℓ be the log-likelihood and the conditional log-likelihood given frailties. The Fisher information for $\hat{\eta}$ can be rewritten in terms of the conditional log-likelihood given as

$$\begin{aligned} I_{\eta\eta}(\nu) &= \mathbb{E}_{\nu} \left(-\frac{\partial^2}{\partial \eta \partial \eta} \ell^*(\eta) \right) \tag{2.B.1} \\ &= \mathbb{E}_{\nu} \left(-\frac{\partial^2}{\partial \eta \partial \eta} \ell(\eta|\mathbf{W}) | \mathbf{W} \in R \right) \\ &\quad - \mathbb{E}_{\nu} \left(\frac{\partial}{\partial \eta} \ell(\eta|\mathbf{W}) \frac{\partial}{\partial \eta} \ell^T(\eta|\mathbf{W}) | \mathbf{W} \in R \right) \\ &\quad + \frac{\partial}{\partial \eta} \ell^*(\eta) \frac{\partial}{\partial \eta} \ell^{*T}(\eta), \end{aligned}$$

where \mathbf{W} are the unobserved frailties and R is the set of possible frailties given the data. Notably the last term is zero at the MLE and thus a simplified notation for the Fisher information at the MLE is given as

$$I_{\eta\eta} = I_{\eta\eta}^{(full)} - I_{\eta\eta}^{(loss)}$$

where the first term represents the full information and the second term represents the loss of information due to the unobserved frailties.

Let

- d_{k1}, d_{k2} : number of failures of cause 1 and 2 in hospital k respectively
- d_1, d_2 : number of failures of cause 1 and 2 in total respectively
- $d_{kl'}$: number of failures of cause 1 at time $t_{l'}$ in hospital k
- $d_{km'}$: number of failures of cause 2 at time $t_{m'}$ in hospital k
- $d_{1l'}$: number of failures of cause 1 at time $t_{l'}$
- $d_{2m'}$: number of failures of cause 2 at time $t_{m'}$
- $t_{kl}, l = 1, \dots, d_{k1}$: ordered event times for cause 1 in hospital k
- $t_{km}, m = 1, \dots, d_{k2}$: ordered event times for cause 2 in hospital k
- $t_{l'}, (l' = 1, \dots, d_1)$: ordered event times for cause 1
- $t_{m'}, (m' = 1, \dots, d_2)$: ordered event times for cause 2
- $\Lambda_{10}(t) = \sum_{t_{l'} \leq t} \lambda_{10}(t_{l'})$
- $\Lambda_{20}(t) = \sum_{t_{m'} \leq t} \lambda_{20}(t_{m'})$
- $\sum_{i=1}^{n_k} e^{\beta_1^T \mathbf{X}_{ki}} \Lambda_{10}(t_{ki}) = \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i:t_{ki} \geq t_{l'}} e^{\beta_1^T \mathbf{X}_{ki}}$
- $\sum_{i=1}^{n_k} e^{\beta_2^T \mathbf{X}_{ki}} \Lambda_{20}(t_{ki}) = \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i:t_{ki} \geq t_{m'}} e^{\beta_2^T \mathbf{X}_{ki}}$
- $R_k(t) = \{i : t_{ki} \geq t\}$: risk set at time t for hospital k

The conditional log-likelihood given frailties can be expressed as

$$\begin{aligned} \ell = & \sum_k d_{k1} \log\left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1}\right) + \sum_{l=1}^{d_{k1}} \log(\lambda_{10}(t_{kl})) + \sum_{l=1}^{d_{k1}} \beta_1^T \mathbf{X}_{kl} \\ & - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \\ & + d_{k2} \log\left(\frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2}\right) + \sum_{m=1}^{d_{k2}} \log(\lambda_{20}(t_{km})) + \sum_{m=1}^{d_{k2}} \beta_2^T \mathbf{X}_{km} \\ & - \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} e^{\beta_2^T \mathbf{X}_{ki}}. \end{aligned}$$

The term $I_{\eta\eta}^{(loss)}$ is the product of the gradient vector of the conditional log-likelihood with itself. The elements of the gradient vector are:

$$\begin{aligned} \frac{\partial}{\partial \beta_{1j}} \ell &= \sum_k \left\{ \sum_{l=1}^{d_{k1}} \mathbf{X}_{klj} - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} e^{\beta_1^T \mathbf{X}_{ki}} \right\} \\ \frac{\partial}{\partial \beta_{2j}} \ell &= \sum_k \left\{ \sum_{m=1}^{d_{k2}} \mathbf{X}_{kmj} - \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} e^{\beta_2^T \mathbf{X}_{ki}} \right\} \\ \frac{\partial}{\partial \lambda_{10l'}} \ell &= \sum_k \left\{ \frac{d_{kl'}}{\lambda_{10l'}(t_{l'})} - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \right\} \\ &= \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial}{\partial \lambda_{20m'}} \ell &= \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{i \in R_k(t_{m'})} e^{\beta_2^T \mathbf{X}_{ki}} \end{aligned}$$

The second order derivatives to calculate the full information matrix $I^{(full)}$ are:

$$\begin{aligned} \frac{\partial^2}{\partial \beta_{1j} \partial \beta_{1h}} \ell &= - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} \mathbf{X}_{kih} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{1j} \partial \beta_{2h}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{1j} \partial \lambda_{10l'}} \ell &= - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{1j} \partial \lambda_{20m'}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{2j} \partial \beta_{2h}} \ell &= - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} \mathbf{X}_{kih} e^{\beta_2^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{2j} \partial \lambda_{10l'}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{2j} \partial \lambda_{20m'}} \ell &= - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} e^{\beta_2^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \lambda_{10p'} \partial \lambda_{10l'}} \ell &= 0, \quad \frac{\partial^2}{\partial \lambda_{10l'} \partial \lambda_{10l'}} \ell = - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})^2} \\ \frac{\partial^2}{\partial \lambda_{10p'} \partial \lambda_{20m'}} \ell &= 0 \\ \frac{\partial^2}{\partial \lambda_{20p'} \partial \lambda_{20m'}} \ell &= 0, \quad \frac{\partial^2}{\partial \lambda_{20m'} \partial \lambda_{20m'}} \ell = - \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})^2} \end{aligned}$$

Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

This chapter is based on joint work with Hein Putter and Marta Fiocco.

Abstract

Following tumor removal surgery soft tissue sarcoma patients are at risk for disease recurrence, which can indicate an increased risk of death. The predictive value of this time-dependent variable can be summarized by the time-specific Area Under the receiver operating characteristics Curve (AUC). However, the fact that recurrence is often diagnosed in an interval-censored fashion is frequently ignored when modelling its effect on survival. Follow-up schemes determine the times at which a patient is diagnosed with recurrence. The effect that ignoring the interval-censored nature of the observation time has on the time-specific AUC in both incident/dynamic and cumulative/dynamic definition is studied.[80, 171] AUC estimates derived from different methods for fitting two types of models are compared: the Cox model with time-dependent covariate and the illness-death model for interval-censored data. Data is simulated from an illness-death model with Weibull transition hazards and the disease state is censored at regular observation intervals. The true AUC is determined by transition probabilities, derived from the Weibull transition hazards. The method is applied to a data set of 2232 patients with high-grade soft tissue sarcoma and results are discussed.

§3.1 Introduction

Survival analysis studies the distribution of time from a time origin to an event of interest. It is often applied in the medical field where for example the time from diagnosis to death is studied. The intrinsic particularity of survival data is that it is generally incomplete: the event of interest cannot always be observed because it takes time to observe it. Data of individuals who did not experience the event of interest within a specific time window are right-censored. A frequently used method to study the effect of covariates on survival time is the Cox proportional hazards model.[44] In the medical field it is often applied to study the effect of risk factors on a single event such as death or disease progression. However, in practice disease progression may be described by more than one type of event. These more complicated event structures can be modeled simultaneously using multi-state models.[119] The most simple of such models is the illness-death model, which is described by three states (see Figure 3.1): an individual is initially disease-free (state 0), he may then develop disease (state 1) and die (state 2) or he may die without disease. Like in the single event situation the Cox model can be used to model the effect of covariates on the transitions between states.

The illness-death model is applicable to a variety of disease settings; a problem arises, however, if the time of disease cannot be observed exactly. Often, disease can only be diagnosed at pre-specified follow-up times. An example lies in the care of patients with soft tissue sarcoma. After initial treatment by tumor removal surgery a patient may develop distant metastases and then die. Metastases are diagnosed at pre-specified follow-up visits at which an X-ray of the patient is screened. If metastases are found, it is therefore only known that they appeared between the last negative screening and the first positive screening; the data is interval-censored. This type of data contains two types of missing information: (1) the time of disease is only known to have happened between two visits, it is interval-censored. (2) If the last disease screening prior to death or last recorded follow-up was negative the disease status of a patient between last screening and death or last recorded follow-up is unknown.

The illness-death model for interval-censored data has been previously studied and it was found that ignoring the observation scheme of the data leads to biased estimates of regression coefficients, baseline hazards, and survival.[64, 86, 65, 168, 97] A prominent motivation comes from the study of dementia data.[86, 168, 97] Dementia is diagnosed at infrequent follow-up visits which results in the time to dementia being interval-censored. Further, if a patient's last dementia test was negative and he dies it is not known if he acquired dementia prior to death. Frydman (1995)[64] developed a non-parametric maximum likelihood procedure for the estimation of the cumulative transition hazards when times of disease are interval-censored. He does not address the second form of incompleteness however, i.e. it is assumed that the disease state is known before death or right-censoring time. Joly et al. (2002)[86] proposed a non-parametric penalized likelihood method to estimate transition intensities in an illness-death model with an intermittently observed disease state. Simulations showed that not adjusting for the interval-censored nature of the data leads to a systematic bias in the estimation of transition intensities. Frydman and Szarek (2009)[65] ex-

tended the methodology of Frydman (1995)[64] to incorporate the observations with unknown intermediate event status. They estimated the distribution of the time to the first occurrence of disease or death and showed that their method corrects bias. Yu et al. (2010)[168] used multiple imputation to analyze two aspects concerning the risk of dementia: the risk of developing dementia and the impact of dementia on survival. Leffondré et al. (2013)[97] performed simulation studies to show how interval-censoring affects the estimation of the effect of risk factors.

If event times are observed exactly the illness-death model can be estimated with several R-packages, such as the `survival` and the `mstate` package.[141, 49, 48] The number of packages that can deal with an interval-censored disease state however is limited. The `msm` and the `SmoothHazard` package can fit an illness-death model for interval-censored disease times and exact death times.[85, 144] In the `msm` package piece-wise constant hazards need to be assumed and in the `SmoothHazard` package the user is able to choose between Weibull transition hazards and M-splines. The `coxinterval` package can estimate the illness-death model for data with interval-censored disease times as long as some disease times are observed exactly.[31] While the effect of ignoring the interval-censored nature of the data on regression coefficients and baseline hazards has been studied, the effect on the assessment of predictive accuracy has been neglected so far.

The aim of this article is to study the predictive accuracy of an interval-censored binary disease marker on survival. How much does the occurrence or absence of disease contribute to survival predictions over time? The illness-death model for data in which the disease state is interval-censored is considered. The effect of interval-censoring on the time-specific Area Under the receiver operating characteristics Curve (AUC) in both incident/dynamic and cumulative/dynamic definition is evaluated.[80, 171] Several estimation approaches are compared for two types of models: the Cox model with time-dependent disease marker and the illness-death model for interval-censored data as implemented in the `msm` and `SmoothHazard` R-packages.[85, 144] For this purpose a simulation study is conducted where data is simulated from an illness-death model with Weibull transition hazards.

The remainder of this article is organized as follows. Section 3.2 introduces the definitions of time-specific AUC for a binary time-dependent marker and the theoretical AUC values for a Weibull illness-death model. In Section 3.3 the different models considered in this work are illustrated. A simulation study is presented in Section 3.4. In Section 3.5 the different methods are applied to data of soft tissue sarcoma patients. A discussion follows in Section 3.6.

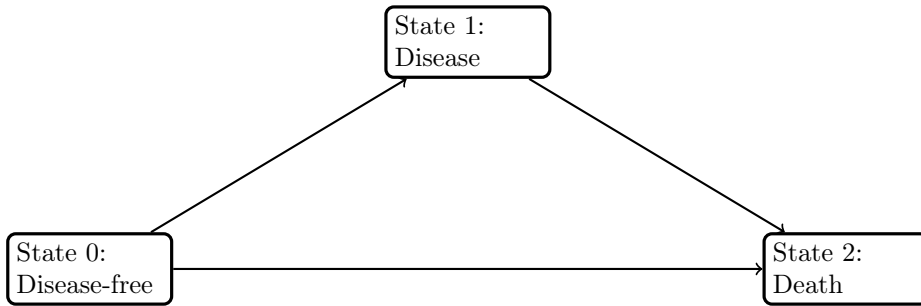


Figure 3.1: Illness-death model.

§3.2 Time-specific AUC for binary marker

Several measures of predictive accuracy have been introduced in the field of survival analysis. In this article predictive accuracy is assessed using the time-specific definitions of sensitivity and specificity which allow for censoring proposed by Heagerty et al. (2000)[79], Heagerty and Zheng (2005)[80], and Zheng and Heagerty (2007)[171].

Originally, sensitivity and specificity were defined considering a binary outcome B . Individuals with outcome $B = 1$ were considered to be ‘cases’ and individuals with outcome $B = 0$ were considered ‘controls’. A covariate X together with a classification criterion c can then be used as a classification rule: a subject is predicted to be a ‘case’ if the value of the covariate is bigger than c and it is predicted to be a ‘control’ otherwise. The accuracy of this classification rule can be summarized by the correct classification rates; sensitivity(c) = $P(X > c|B = 1)$ and specificity(c) = $P(X \leq c|B = 0)$. The full range of sensitivity and specificity for different classification criteria c can be graphically summarized by the Receiver Operation Characteristic (ROC) curve which plots sensitivity against 1-specificity. The ROC curve illustrates the difference of the marker distribution between cases and controls. If the distributions are the same, which means that the marker is useless to distinguish cases from controls, then the ROC curve lies on the 45 degree line. The Area Under the Curve (AUC) is a measure of concordance between the marker and the outcome and can be used to summarize the predictive accuracy of the marker X . It is defined by

$$\text{AUC}(X) = P(X_1 > X_0) + 0.5 \cdot P(X_1 = X_0),$$

where X_1 is the value of a covariate drawn from the distribution of cases ($B = 1$) and X_0 is the value of a covariate drawn from the distribution of controls ($B = 0$). To extend the concept of sensitivity and specificity to allow for censored data several definitions for cases and controls were studied[79, 80, 171].

In this article a time-dependent binary covariate $X(t)$ representing disease is considered. The covariate $X(t)$ can take values 0 and 1 which correspond to not having disease and having disease at time t , respectively. The Markov assumption is assumed for the studied illness-death model throughout the article.

§3.2.1 Incident cases and dynamic controls

Heagerty and Zheng (2005)[80] define incident sensitivity and dynamic specificity at time t as

$$\begin{aligned}\text{sensitivity}^I(c, t) &= P(X(t) > c \mid T = t), \\ \text{specificity}^D(c, t) &= P(X(t) \leq c \mid T > t),\end{aligned}$$

where c is a classification criterion, T is time of death and $X(t)$ is the time-dependent disease marker evaluated at time t . In this definition the individuals who die at time t are considered cases and individuals who survive beyond time t are considered controls. Let i, j be individuals, $X_i(t), X_j(t)$ their marker values at time t , and T_i and T_j their death times. The incident/dynamic AUC is then defined by[80]

$$\begin{aligned}\text{AUC}^{I/D}(t) &= P(X_i(t) > X_j(t) \mid T_i = t, T_j > t) \\ &+ 0.5P(X_i(t) = X_j(t) \mid T_i = t, T_j > t).\end{aligned}$$

In case $X_i(t)$ and $X_j(t)$ are binary covariates the $\text{AUC}^{I/D}(t)$ can be rewritten as

$$\text{AUC}^{I/D}(t) = 0.5 + 0.5(p(t) - \pi_1(t)), \quad (3.2.1)$$

where $\pi_1(t)$ is the probability that a person alive at time t has experienced disease (prevalence of disease) and $p(t)$ is the probability that a person who dies at time t has a history of disease (see Appendix 3.A). The disease marker $X(t)$ is related to the illness-death model of Figure 3.1 in the following way: $X(t) = 0$ if a patient did not move to state 1 (disease) before time t (in state 0 or 2 at time t) and $X(t) = 1$ if a patient moved to state 1 (disease) before time t (in state 1 or 2 at time t). The terms $\pi_1(t)$ and $p(t)$ can be expressed by transition probabilities in a multi-state model with states 0, 1, 2 (Figure 3.1),

$$\pi_1(t) = P(X_i(t) = 1 \mid T_i > t) = \frac{P_{01}(t)}{P_{00}(t) + P_{01}(t)}, \quad (3.2.2)$$

$$p(t) = P(X_i(t-) = 1 \mid T_i = t) = \frac{\frac{\lambda_{12}(t)}{\lambda_{02}(t)} P_{01}(t-)}{P_{00}(t-) + \frac{\lambda_{12}(t)}{\lambda_{02}(t)} P_{01}(t-)}, \quad (3.2.3)$$

where $t-$ means just before time t , $P_{0l}(t)$ is the conditional probability of being in state l , ($l = 0, 1$) at time t given in state 0 at time 0 and $\lambda_{k2}(t)$ is the transition hazard at time t for moving from state k , ($k = 0, 1$) to state 2.

The incident/dynamic AUC at a specific time t measures how well the disease marker evaluated at time t separates those who die at t from those who survive.

The difference between $p(t)$ and $\pi_1(t)$ is equal to

$$\begin{aligned}
 p(t) - \pi_1(t) &= \frac{\gamma(t)P_{01}(t)}{P_{00}(t) + \gamma(t)P_{01}(t)} - \frac{P_{01}(t)}{P_{00}(t) + P_{01}(t)} \\
 &= (\gamma(t) - 1) \frac{P_{01}(t)P_{00}(t)}{(P_{00}(t) + \gamma(t)P_{01}(t))(P_{00}(t) + P_{01}(t))}, \quad (3.2.4) \\
 &= (\gamma(t) - 1) \frac{1}{(1 + \gamma(t)P_{01}(t)/P_{00}(t))(1 + P_{00}(t)/P_{01}(t))},
 \end{aligned}$$

where $\gamma(t) = \frac{\lambda_{12}(t)}{\lambda_{02}(t)}$.

From (3.2.1) and (3.2.4) follows that if $\gamma(t) \equiv 1$ then $AUC^{I/D}(t) = 0.5$, if $\gamma(t) > 1$ then $AUC^{I/D}(t) \geq 0.5$ and if $\gamma(t) < 1$ then $AUC^{I/D}(t) \leq 0.5$.

§3.2.2 Cumulative cases and dynamic controls

Zheng and Heagerty (2007)[171] define cumulative sensitivity and dynamic specificity at time t for a time-dependent covariate evaluated at time s as

$$\begin{aligned}
 \text{sensitivity}^C(c \mid \text{start} = s, \text{stop} = t) &= P(X(s) > c \mid T \geq s, T \leq t), \\
 \text{specificity}^D(c \mid \text{start} = s, \text{stop} = t) &= P(X(s) \leq c \mid T \geq s, T > t),
 \end{aligned}$$

where T is time of death, $X(s)$ is marker measurement at time s . Cases are individuals who die within a time window $(t - s)$ from s and controls are individuals who survive the time window. The cumulative/dynamic AUC is then defined by

$$\begin{aligned}
 AUC^{C/D}(s, t) &= P(X_i(s) > X_j(s) \mid T_i > s, T_i \leq t, T_j > s, T_j > t) \\
 &\quad + 0.5P(X_i(s) = X_j(s) \mid T_i > s, T_i \leq t, T_j > s, T_j > t)
 \end{aligned}$$

where i, j are individuals, $X_i(s), X_j(s)$ their marker values at time s , and T_i, T_j their death times. For binary $X_i(s)$ and $X_j(s)$ the $AUC^{C/D}$ can be rewritten as

$$AUC^{C/D}(s, t) = 0.5 + 0.5(p(s, t) - \pi_1(s, t)), \quad (3.2.5)$$

where $\pi_1(s, t)$ is the probability that a person alive at time t had experienced disease by time s and $p(s, t)$ is the probability that a person that dies in the time interval $(s, t]$ had experienced disease by time s (see Appendix 3.A). The quantities $\pi_1(s, t)$ and $p(s, t)$ can be written in terms of transition probabilities, in the same multi-state model of Figure 3.1:

$$\pi_1(s, t) = P(X_j(s) = 1 | T_j > t) = \frac{P_{11}(s, t)P_{01}(0, s)}{P_{00}(0, t) + P_{01}(0, t)}, \quad (3.2.6)$$

$$p(s, t) = P(X_i(s) = 1 | T_i > s, T_i \leq t) = \frac{P_{12}(s, t)P_{01}(0, s)}{P_{02}(s, t)P_{00}(0, s) + P_{12}(s, t)P_{01}(0, s)}, \quad (3.2.7)$$

where $P_{kl}(u, v)$ is the conditional probability of being in state l at time v given in state k at time u .

The cumulative/dynamic AUC at time s measures how well the disease marker evaluated at time s separates those who die before time t from those who survive until t .

§3.2.3 AUC for Weibull illness-death model

In this article an illness-death model with Weibull distributed transition hazards is studied because of its simple transition probabilities. The transition hazards from state i to state j are defined by

$$\lambda_{ij}(t) = \alpha_{ij}kt^{k-1}, \quad (3.2.8)$$

where k is the common shape parameter and α_{ij} are transition-specific rate parameters. Let

$$\begin{aligned} S_0(t) &= \exp(-(\alpha_{01} + \alpha_{02})t^k), \\ S_1(t) &= \exp(-\alpha_{12}t^k). \end{aligned}$$

The transition probabilities are then equal to [119]

$$P_{00}(u, t) = \frac{S_0(t)}{S_0(u)},$$

$$P_{11}(u, t) = \frac{S_1(t)}{S_1(u)},$$

$$P_{01}(u, t) = \begin{cases} \frac{\alpha_{01}}{\alpha_{01} + \alpha_{02} - \alpha_{12}} \left(\frac{S_1(t)}{S_1(u)} - \frac{S_0(t)}{S_0(u)} \right) & , \text{if } \alpha_{01} + \alpha_{02} - \alpha_{12} \neq 0 \\ \alpha_{01} \left(\frac{S_1(t)}{S_1(u)} t^k - \frac{S_0(t)}{S_0(u)} u^k \right) & , \text{otherwise (note: } S_1(t) = S_0(t)), \end{cases}$$

$$P_{02}^0(u, t) = \frac{\alpha_{02}}{\alpha_{01} + \alpha_{02}} \left(1 - \frac{S_0(t)}{S_0(u)} \right),$$

$$P_{02}^1(u, t) = \begin{cases} \frac{\alpha_{01}}{\alpha_{01} + \alpha_{02}} \left(1 - \frac{S_0(t)}{S_0(u)}\right) - \frac{\alpha_{01}}{\alpha_{01} + \alpha_{02} - \alpha_{12}} \left(\frac{S_1(t)}{S_1(u)} - \frac{S_0(t)}{S_0(u)}\right), & \text{if } \alpha_{01} + \alpha_{02} - \alpha_{12} \neq 0 \\ \frac{\alpha_{01}}{\alpha_{01} + \alpha_{02}} \left(1 - \frac{S_0(t)}{S_0(u)}\right) - \alpha_{01} \frac{S_0(t)}{S_0(u)} (t^k - u^k), & \text{otherwise} \end{cases},$$

$$P_{02}(u, t) = P_{02}^0(u, t) + P_{02}^1(u, t) = 1 - \frac{\alpha_{02} - \alpha_{12}}{\alpha_{01} + \alpha_{02} - \alpha_{12}} \frac{S_0(t)}{S_0(u)} - \frac{\alpha_{01}}{\alpha_{01} + \alpha_{02} - \alpha_{12}} \frac{S_1(t)}{S_1(u)},$$

$$P_{12}(u, t) = 1 - \frac{S_1(t)}{S_1(u)}.$$

These transition probabilities can be used to calculate the time-specific incident/dynamic and cumulative/dynamic AUC using Equations (3.2.1) and (3.2.5), respectively.

§3.2.4 Estimation

Equations (3.2.1)–(3.2.3) and (3.2.5)–(3.2.7) relate, respectively, the incident/dynamic and cumulative/dynamic AUC to transition probabilities and hazards. Estimates for the AUCs can be obtained by replacing transition probabilities and hazards by their estimated counterparts. Such estimates may be obtained from software packages for multi-state models, such as the R-packages `mstate`, `msm`, and `SmoothHazard` discussed in Section 3.3.[49, 48, 85, 144]

§3.2.5 Estimation of incident/dynamic AUC

Equations (3.2.1)–(3.2.3) are used to estimate the incident/dynamic AUC,

$$\widehat{\text{AUC}}^{I/D}(t) = 0.5 + 0.5(\hat{p}(t) - \hat{\pi}_1(t)), \quad (3.2.9)$$

where

$$\hat{\pi}_1(t) = \frac{\hat{P}_{01}(t)}{\hat{P}_{00}(t) + \hat{P}_{01}(t)},$$

$$\hat{p}(t) = \frac{\frac{\hat{\lambda}_{12}(t)}{\hat{\lambda}_{02}(t)} \hat{P}_{01}(t-)}{\hat{P}_{00}(t-) + \frac{\hat{\lambda}_{12}(t)}{\hat{\lambda}_{02}(t)} \hat{P}_{01}(t-)},$$

where $t-$ means just before time t , $\hat{P}_{0l}(t)$ is an estimate of the conditional probability of being in state l , ($l = 0, 1$) at time t given in state 0 at time 0 and $\hat{\lambda}_{k2}(t)$ is an estimate of the transition hazard at time t for moving from state k , ($k = 0, 1$) to state 2.

§3.2.6 Estimation of cumulative/dynamic AUC

Equations (3.2.5)–(3.2.7) are used to estimate the cumulative/dynamic AUC,

$$\widehat{\text{AUC}}^{C/D}(s, t) = 0.5 + 0.5(\hat{p}(s, t) - \hat{\pi}_1(s, t)),$$

where

$$\hat{\pi}_1(s, t) = \frac{\hat{P}_{11}(s, t)\hat{P}_{01}(0, s)}{\hat{P}_{00}(0, t) + \hat{P}_{01}(0, t)},$$

$$\hat{p}(s, t) = \frac{\hat{P}_{12}(s, t)\hat{P}_{01}(0, s)}{\hat{P}_{02}(s, t)\hat{P}_{00}(0, s) + \hat{P}_{12}(s, t)\hat{P}_{01}(0, s)},$$

where $\hat{P}_{kl}(u, v)$ is an estimate of the conditional probability of being in state l at time v given in state k at time u .

§3.3 Illness-death models

Four different methods to estimate the illness-death model for interval-censored data were compared: (1) the Cox model with disease state as time-dependent covariate (ignoring the interval-censored nature of the time-dependent covariate), (2) the piecewise-constant model accounting for interval-censoring using the `msm` function from the `msm` package, (3) the Weibull model accounting for interval-censoring using the `idm` function from the `SmoothHazard` package, and (4) the M-spline model accounting for interval-censoring using the `idm` function from the `SmoothHazard` package.[85, 144] A sieve estimator for a Cox based multi-state model that accounts for interval-censoring is implemented in the `coxduel` function from the `coxinterval` package, however, at least some disease times need to be observed exactly for the estimation procedure to work.[31] Since this is not the case in the motivation for this study the `coxinterval` package was not further considered. In the simulation study presented in Section 3.4 all methods are used and from their transition probabilities the AUC is estimated.

§3.3.1 Cox model with time-dependent covariate

The Cox model with a binary time-dependent covariate is defined by the following hazard function:

$$\lambda(t|X(t)) = \lambda_0(t) \exp(\beta X(t)),$$

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

where $\lambda_0(t)$ is the baseline hazard, $X(t)$ is the binary disease marker at time t and β its effect. This model can be estimated by e.g. the `coxph` R-function from the `survival` package[141], however, ignoring the interval-censored nature of the time-dependent covariate. Disease time is assumed to be the time of diagnosis of disease: $X(t) = 0$ if a patient was not diagnosed with disease yet at time t and $X(t) = 1$ if a patient was diagnosed with disease by time t . The Cox model with time-dependent covariate corresponds to an illness-death model in which the transition hazards to the state death are proportional. This allows for the estimation of the effect of disease on death in form of a hazard ratio (HR). Transition probabilities can be retrieved from the model using `msfit` and `probtrans` functions from the `mstate` package.[49, 48] The `risksetAUC` R-function from the `risksetROC` package[80] estimates the incident/dynamic AUC for a Cox model with time-dependent covariate. Additionally to estimating the AUC using transition probabilities this function is also used in the simulation study in Section 3.4.

§3.3.2 Piecewise-constant model accounting for interval-censoring

This Markov model is described in Figure 3.1. Interval-censored data from an illness-death process are a special case of panel data, in which the state of an individual is observed at a finite series of times. The likelihood for panel data can be calculated in closed form if the transition hazards are constant or piece-wise constant.[85] A model with piecewise-constant hazards given by

$$\lambda_{ij}(t) = \begin{cases} \lambda_{ij1} & , \text{if } t \leq c_1 \\ \lambda_{ij2} & , \text{if } c_1 < t \leq c_2 , \\ \vdots & \end{cases}$$

where c_k are the times at which the hazard may change is considered. This model is implemented in the `msm` package and can account for the interval-censored disease state.[85] In the simulation study of Section 3.4 the hazards towards the death state are assumed to be proportional so that an effect of disease on survival can be estimated.

§3.3.3 Weibull model accounting for interval-censoring

This model is a Markov illness-death model (see Figure 3.1) which assumes a Weibull distribution for the transition hazards given by

$$\lambda_{ij}(t) = \alpha_{ij} k_{ij} t^{k_{ij}-1},$$

where α_{ij} and k_{ij} are rate and shape parameters for the transition from state i to state j , respectively. This model is implemented in the `SmoothHazard` R-package.[144]

It accounts for interval-censoring and the probability of developing disease between last disease scan and death or lost to follow-up and it is estimated by maximizing the likelihood with the `idm` function. The function does not allow for the transition hazards to the death state to be set proportional and therefore no effect of disease on death can be estimated. The package provides prediction of transition probabilities based on estimated transition hazards.

§3.3.4 M-spline model accounting for interval-censoring

This Markov illness-death model is described in Figure 3.1. The model is estimated using a penalized likelihood approach with non-parametric transition hazards $\lambda_{01}(t)$, $\lambda_{02}(t)$, and $\lambda_{12}(t)$, approximated by M-splines and it is implemented in the `SmoothHazard` R-package.[86, 144] This model as the previous two, accounts for interval-censoring of the disease state as well as the probability of developing disease between the last disease scan and death or lost to follow-up. It is estimated by the `idm` function from the `SmoothHazard` R-package in which the option `method = "Splines"` is set.[144] By default 7 knots per transition are estimated. As for the Weibull model, the transition hazards towards the death state can not be set proportional and therefore no HR for disease can be estimated. Transition probabilities can be obtained using functions provided in the package.

§3.4 Simulation

To study the predictive accuracy of an interval-censored disease marker on survival a simulation study was conducted. Incident/dynamic and cumulative/dynamic AUC were computed to quantify the predictive accuracy of the disease marker for different estimation procedures of the illness-death model. The methods compared were the Cox model with time-dependent disease marker, which ignores interval-censoring, and the illness-death model for interval-censored data estimated with three different implementations: the piecewise-constant model implemented in the `msm` package, the Weibull model, and the M-spline model which are both implemented in the `SmoothHazard` package (see Section 3.3). The piecewise-constant model needs as input pre-specified change points at which the hazard may change. For the simulation study 4 change points were considered 6, 30, 60, and 90 months. For the M-spline model the default of 7 knots per transition was used.

Motivated by the clinical data discussed in Section 3.5 multiple data scenarios were simulated and results from the different methods were compared. The number of individuals per data set was either equal to 1000 or equal to 2000. Data were generated from Weibull transition hazards with a common shape parameter k and different rate parameters α_{01} , α_{02} and α_{12} (see Equation (3.2.8)). The Weibull parameters were based on the data discussed in Section 3.5 and were fixed throughout the simulated scenarios ($\alpha_{01} = 0.05$, $\alpha_{02} = 0.05$, $\alpha_{12} = 0.56$, $k = 0.5$).

The survival time was censored according to two different censoring schemes: either it was censored administratively at 10 years follow-up or censoring times were sampled from a uniform distribution between 5 and 10 years. The disease state was

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

Table 3.1: Simulated scenarios.

Scenario	N	Censoring	Follow-up
A	1000	unif(5, 10)	3
B	1000	unif(5, 10)	6
C	1000	unif(5, 10)	12
D	1000	10	3
E	1000	10	6
F	1000	10	12
G	2000	unif(5, 10)	3
H	2000	unif(5, 10)	6
I	2000	unif(5, 10)	12
J	2000	10	3
K	2000	10	6
L	2000	10	12

Abbreviations: N, total number of patients; Censoring, type of death censoring, unif(5, 10) means censoring was uniformly sampled between 5 and 10 years and 10 means that administrative censoring occurred at 10 years; Follow-up, time between disease observations in months.

observed only at pre-specified follow-up visits. The scenarios cover three different follow-up schemes in which the disease state was observed every 3, 6, or 12 months. Table 3.1 summarizes the simulated scenarios. Each scenario was simulated 1000 times.

Table 3.2 shows the estimated coefficients and hazard ratios of disease for the piecewise-constant and the Cox model. For the Weibull and M-spline model no effect could be estimated, since the `idm` function does not allow transition hazards to be proportional. The coefficients from the Cox model were consistently more biased than from the piecewise-constant model. The Cox model underestimated the true coefficient and the bias increased for larger follow-up intervals. These results are in line with Leffondré et al. (2013)[97] who showed that the effect estimates of the Cox model were biased if the covariate affected both the risk of disease and death.

Simulation results show that the coefficients from the piecewise-constant model had smaller bias and smaller root mean square error.

AUC results obtained from different methods for scenarios A-F are summarized in Tables 3.3 and 3.4. For results concerning other scenarios, see Appendix 3.B. The cumulative/dynamic AUC was estimated every month and the incident/dynamic AUC was estimated at each event time, because it depends on the transition hazard evaluated at that time. In Table 3.3 where the AUC at specific time points was investigated, the AUC estimate just before that time was considered.

The M-spline model did not converge for many data sets. In some of these cases this prevented the estimation of the incident/dynamic and cumulative/dynamic AUC.

The number of invalid estimations is shown in Appendix 3.B, Table 3.B.4. The results of the M-spline model in Tables 3.3 and 3.4 are based only on valid estimations. Additionally, for the M-spline model it is not possible to obtain transition probabilities for a time after the last observation time. This restricts the estimation of the cumulative/dynamic AUC (with prediction window of 5 years) to be estimated only until 5 years prior to the last observation time, see Figure 3.2.

Table 3.3 shows the bias, empirical standard error, and root mean square error for estimates of the incident/dynamic AUC at different years. The Weibull model outperformed the other models in every scenario. This is not surprising since data were generated according to Weibull distributions. The M-spline model consistently had the largest standard error as well as the second smallest bias overall. The piecewise-constant model was slightly less biased than the Cox model for scenarios with 6 and 12 months in between follow-up visits (scenarios B, C, E, F). For the scenarios with 3 months in between follow-up visits the Cox model outperformed the piecewise-constant model (scenarios A, D) in terms of bias. The incident/dynamic AUC for the Cox model was estimated by two different approaches. The first approach computes the AUC from the ROC curve derived from the estimated sensitivity and specificity and is implemented in the `risksetAUC` function from the `risksetROC` R-package[80]. The second approach computes the AUC from estimated transition probabilities as described in Equation (3.2.9). Since the two estimation procedures for the Cox model's AUC gave similar result, only results for the transition probability based AUC are presented in Table 3.3 (see Appendix 3.B, Table 3.B.2 for all results).

Table 3.4 shows the bias, empirical standard error, and root mean square error for estimates of the cumulative/dynamic AUC. The piecewise-constant model showed the worst performance and underestimated the true AUC. The Weibull model, M-spline and the Cox model provided good results.

In Table 3.3 and 3.4 the AUC estimates were investigated at 1, 3, and 5 years which coincide with the times of follow-up visits for every scenario. At these times the Cox model displays less bias compared to times in between follow-up visits (see, Figure 3.1 and 3.2).

The censoring scheme did not have a large effect on the incident/dynamic and cumulative/dynamic AUC estimates for the Cox, piecewise-constant and Weibull model. It did however, have an effect on the estimates of the M-spline model. Earlier censoring according to the uniform distribution between 5 and 10 years (scenarios A-C, G-I) resulted in a larger percentage of invalid estimations (see Appendix 3.B, Table 3.B.4), compared to administrative censoring at 10 years (scenarios D-F, J-L).

The number of individuals per data set did not have a large effect on the mean HRs for disease, it did however reduce the empirical standard error (Table 3.2). Average AUC estimates were nearly identical between scenarios where only the size differed and therefore only results for $n = 1000$ are shown in Table 3.3 and 3.4 (see Appendix 3.B for results of all scenarios). The number of patients per data set did have an effect on the percentage of converged M-spline models (see Appendix 3.B, Table 3.B.4).

The follow-up schemes with larger intervals resulted in larger bias of the incident/dynamic AUC estimates, particularly for the Cox model. The follow-up scheme with larger intervals resulted in consistently more biased estimates of the cumulat-

ive/dynamic AUC for the piecewise-constant model. The Cox, Weibull and M-spline model based estimates were of limited bias for the different follow-up schemes.

Figure 3.1 and 3.2 show incident/dynamic and cumulative/dynamic AUC estimates respectively for data scenarios A, B and C with follow-up visits every 3, 6, and 12 months, respectively. Each plot depicts the true AUC in blue and 1000 green lines which correspond to the AUC estimates of each simulated data set. The Cox model's AUC displays jumps at the observation time points. The reason is that at those time points the proportion of diseased individuals is increased in the risk set. Before the first observation time point the curve is equal to 0.5, because no disease was observed yet.

For the incident/dynamic AUC in Figure 3.1 the M-spline model shows a similar behaviour to the Cox model. No distinct jumps are observed but waves can be seen that are most defined at the beginning of follow-up time. Since the piecewise-constant model and the Weibull model make assumptions about the hazard function, the AUC estimates do not display jumps or waves, like for the Cox and M-spline model.

The variation between curves is much larger for the incident/dynamic AUC estimates in Figure 3.1 compared to the cumulative/dynamic estimates in Figure 3.2. Results indicate that the piecewise-constant model is not flexible enough to follow the shape of the true AUC curve, particularly in the incident/dynamic case. The M-spline model displays a larger variance in the incident/dynamic case and shows a better performance in the cumulative/dynamic case. Its cumulative/dynamic curves underestimated the true AUC initially but recovered later on. The Weibull model outperformed the other models, but again one should keep in mind that data were generated from Weibull distributions.

Table 3.2: Effect of disease.

Scenario	N	Censoring	Follow-up	Model	Mean(coef)	exp(mean(coef))	SE(coef)	Bias(coef)	RMSE(coef)
Truth					2.42	11.20			
A	1000	unif(5, 10)	3	Cox	2.35	10.44	0.09	-0.07	0.11
A	1000	unif(5, 10)	3	Piecewise-constant	2.43	11.32	0.09	0.01	0.09
B	1000	unif(5, 10)	6	Cox	2.29	9.91	0.10	-0.12	0.16
B	1000	unif(5, 10)	6	Piecewise-constant	2.48	11.90	0.10	0.06	0.12
C	1000	unif(5, 10)	12	Cox	2.24	9.37	0.11	-0.18	0.21
C	1000	unif(5, 10)	12	Piecewise-constant	2.41	11.11	0.12	-0.01	0.12
D	1000	10	3	Cox	2.35	10.45	0.09	-0.07	0.11
D	1000	10	3	Piecewise-constant	2.45	11.59	0.08	0.03	0.09
E	1000	10	6	Cox	2.31	10.05	0.09	-0.11	0.14
E	1000	10	6	Piecewise-constant	2.50	12.19	0.10	0.08	0.13
F	1000	10	12	Cox	2.25	9.47	0.10	-0.17	0.20
F	1000	10	12	Piecewise-constant	2.43	11.34	0.11	0.01	0.11
G	2000	unif(5, 10)	3	Cox	2.34	10.43	0.07	-0.07	0.10
G	2000	unif(5, 10)	3	Piecewise-constant	2.42	11.30	0.06	0.01	0.06
H	2000	unif(5, 10)	6	Cox	2.29	9.91	0.07	-0.12	0.14
H	2000	unif(5, 10)	6	Piecewise-constant	2.47	11.88	0.07	0.06	0.09
I	2000	unif(5, 10)	12	Cox	2.24	9.38	0.08	-0.18	0.20
I	2000	unif(5, 10)	12	Piecewise-constant	2.41	11.09	0.08	-0.01	0.08
J	2000	10	3	Cox	2.34	10.40	0.06	-0.07	0.10
J	2000	10	3	Piecewise-constant	2.44	11.52	0.06	0.03	0.07
K	2000	10	6	Cox	2.30	10.00	0.06	-0.11	0.13
K	2000	10	6	Piecewise-constant	2.49	12.10	0.07	0.08	0.10
L	2000	10	12	Cox	2.24	9.40	0.07	-0.17	0.19
L	2000	10	12	Piecewise-constant	2.42	11.24	0.08	0.00	0.08

Abbreviations: SE, empirical standard error; RMSE, root mean square error.

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

Table 3.3: Time-specific incident/dynamic AUC.

Scenario	Model	$AUC^{I/D}(1) = 0.71$			$AUC^{I/D}(3) = 0.72$			$AUC^{I/D}(5) = 0.72$		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
A	Cox	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
A	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03
A	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
A	M-spline	0.02	0.02	0.03	0.00	0.02	0.02	-0.01	0.04	0.04
B	Cox	-0.07	0.02	0.08	-0.04	0.01	0.04	-0.03	0.01	0.03
B	Piecewise-constant	-0.06	0.01	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
B	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
B	M-spline	0.01	0.03	0.03	-0.01	0.03	0.03	0.00	0.04	0.04
C	Cox	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
C	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.02	0.05
C	Weibull	-0.01	0.03	0.03	0.00	0.02	0.02	0.00	0.02	0.02
C	M-spline	-0.05	0.04	0.06	0.00	0.03	0.03	0.00	0.04	0.04
D	Cox	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
D	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03
D	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	M-spline	0.02	0.02	0.03	-0.01	0.02	0.03	0.00	0.03	0.04
E	Cox	-0.07	0.02	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
E	Piecewise-constant	-0.06	0.01	0.06	-0.04	0.01	0.04	-0.03	0.01	0.03
E	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
E	M-spline	0.01	0.02	0.03	-0.01	0.03	0.03	0.00	0.04	0.04
F	Cox	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
F	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.01	0.05
F	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
F	M-spline	-0.05	0.04	0.06	0.00	0.03	0.03	0.00	0.04	0.04

Abbreviations: $AUC^{I/D}(x)$, incident/dynamic AUC at year x ; SE, empirical standard error; RMSE, root mean square error.

Table 3.4: Time-specific cumulative/dynamic AUC.

Scenario	Model	$AUC^{C/D}(1) = 0.59$			$AUC^{C/D}(3) = 0.62$			$AUC^{C/D}(5) = 0.64$		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
A	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
A	Piecewise-constant	-0.02	0.01	0.02	-0.01	0.01	0.02	-0.01	0.01	0.02
A	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
A	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
B	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
B	Piecewise-constant	-0.04	0.00	0.04	-0.02	0.01	0.03	-0.02	0.01	0.02
B	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
B	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
C	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
C	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
C	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
C	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
D	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	Piecewise-constant	-0.02	0.01	0.02	-0.01	0.01	0.02	-0.01	0.01	0.02
D	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
E	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
E	Piecewise-constant	-0.04	0.00	0.04	-0.03	0.01	0.03	-0.02	0.01	0.02
E	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
E	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
F	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
F	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
F	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
F	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02

Abbreviations: $AUC^{C/D}(x)$, cumulative/dynamic AUC at year x ; SE, empirical standard error; RMSE, root mean square error.

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

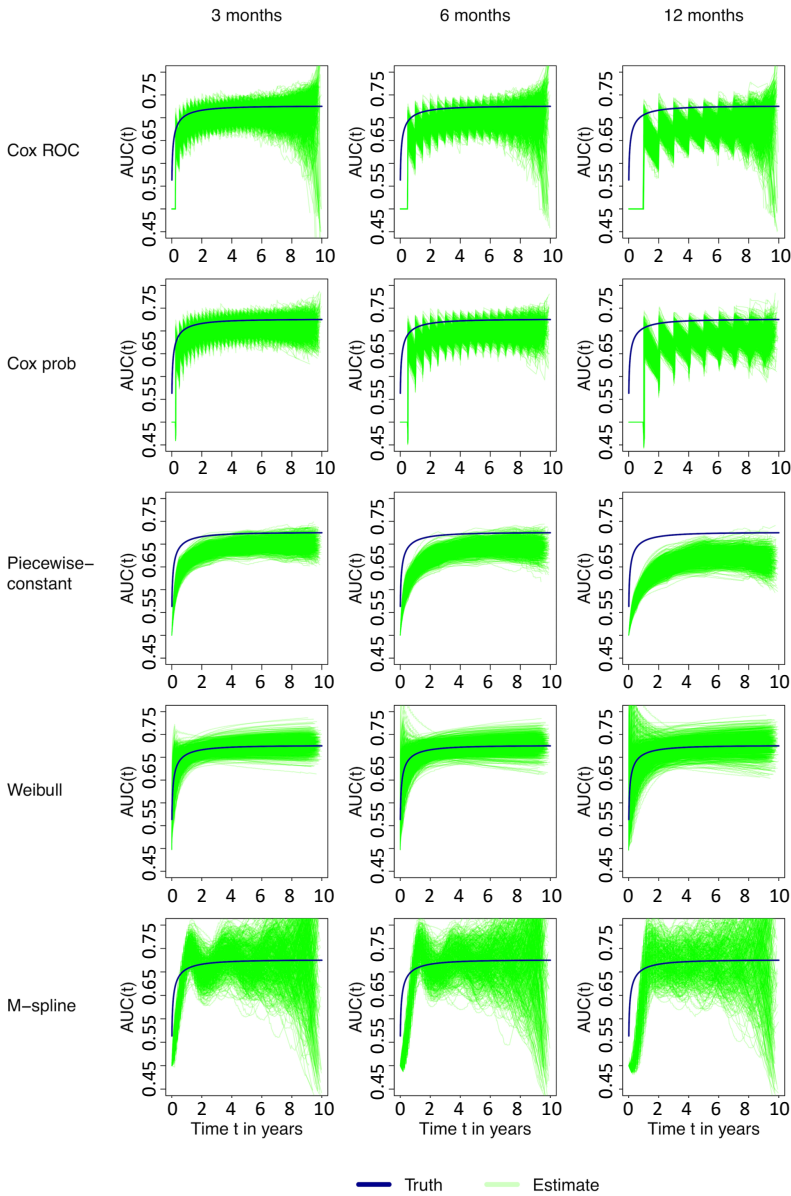


Figure 3.1: Incident/dynamic AUC for scenario A (3 months), B (6 months) and C (12 months). **Abbreviations:** Cox ROC, estimate based on *risksetAUC* function; Cox prob, estimate based on transition probabilities of Cox model.

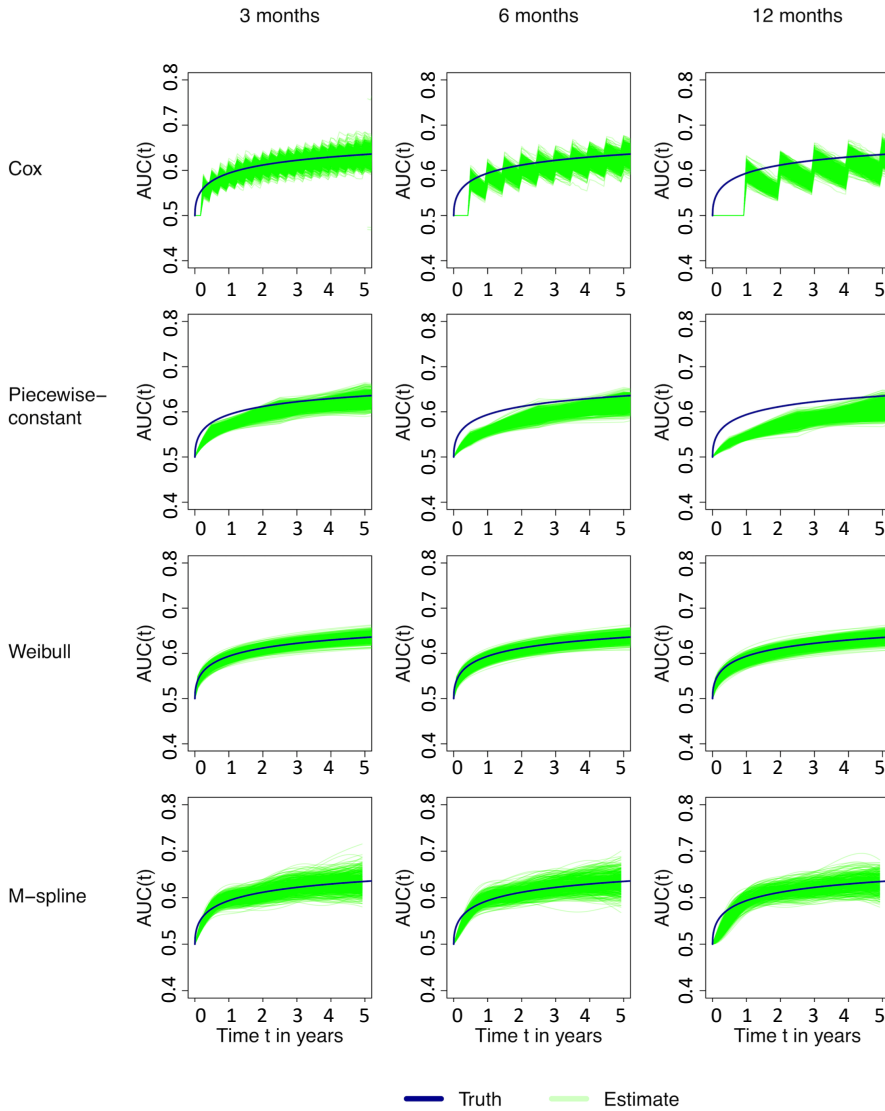


Figure 3.2: Cumulative/dynamic AUC for scenario A (3 months), B (6 months) and C (12 months).

§3.5 Application

The data analyzed in this section was used for the development of a dynamic prediction model for high-grade soft tissue sarcoma patients.[19] The data set contains follow-up information of 2232 patients treated surgically with curative intent. Median follow-up time was 6.42 years. After surgery disease progression can be described by several adverse events: a patient may develop a local recurrence and/or develop distant metastasis (DM) and/or die. The analysis discussed in this section focuses on the effect of DM on death. In total 1034 patients died and 715 patients developed DM (see Figure 3.1).

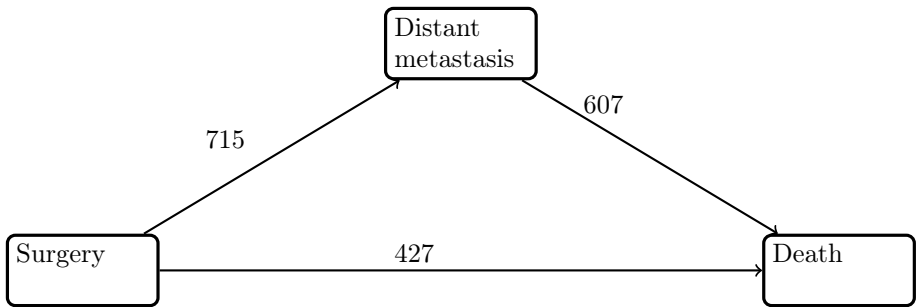


Figure 3.1: Soft tissue sarcoma illness-death model ($n = 2232$).

After surgery a common follow-up visit scheme to screen for DM is to see a patient every 3 months within the first 3 years, then every 6 months until year 5, and from then on once a year.[66] The data did not contain information about exact follow-up times and an approximation of disease screening times was applied. For a patient who was diagnosed with DM during follow-up, the time of DM was interpreted as the first positive screening for DM. Depending on whether DM was diagnosed within the first 3 years, between 3 and 5, or after 5 years the previous screening was assumed to have taken place either 3, 6, or 12 months prior. A patient who was never diagnosed with DM was assumed to have been screened according to the common follow-up scheme described above.

Table 3.1 shows HRs for DM and estimates for the time-specific AUC at different years. The HRs estimated by the Cox and piecewise-constant model are similar, with HRs for the Cox model being slightly larger.

Figure 3.2 displays on the left and the right panel the non-parametric cumulative baseline hazards and a graphical check of their fit to a Weibull distribution, respectively. For this figure the time of DM was assumed to be equal to the time that DM was detected during screening. If the hazards were coming from a Weibull distribution the lines in the right panel of Figure 3.2 would be straight, which is not the case in particular for the transition from surgery to DM. The Weibull model therefore may not be appropriate for this data.

Figure 3.3 shows the AUC over time for the different models. The incident/dynamic AUC of the Weibull model is initially much larger compared to the other models and declines over time. The incident/dynamic AUC of the piecewise-constant model is

the lowest of all three methods. The cumulative/dynamic AUC of the Cox model is generally the largest and the Weibull models the lowest. The M-spline model did not converge for this data set and consequently the incident/dynamic and cumulative/dynamic AUC could not be estimated.

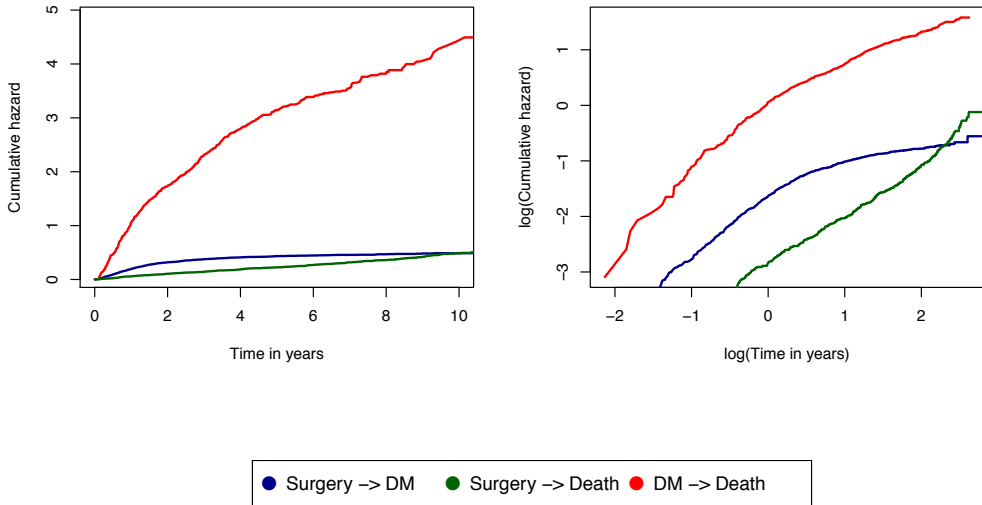


Figure 3.2: Left panel: Cumulative transition hazards. Right panel: plot of $\ln[H(x)]$ vs. $\ln(x)$ to empirically check the fit of the Weibull distribution.

Table 3.1: Effect and predictive accuracy for distant metastasis.

	HR(DM)	$AUC^{I/D}(1)$	$AUC^{I/D}(2)$	$AUC^{I/D}(3)$	$AUC^{I/D}(4)$	$AUC^{I/D}(5)$
Cox ROC	11.71	0.74	0.76	0.76	0.75	0.74
Cox prob	11.71	0.75	0.76	0.76	0.75	0.74
Piecewise-constant	11.28	0.71	0.73	0.73	0.71	0.70
Weibull		0.81	0.78	0.76	0.74	0.72
		$AUC^{C/D}(1)$	$AUC^{C/D}(2)$	$AUC^{C/D}(3)$	$AUC^{C/D}(4)$	$AUC^{C/D}(5)$
Cox ROC						
Cox prob		0.64	0.69	0.70	0.68	0.67
Piecewise-constant		0.62	0.66	0.68	0.66	0.66
Weibull		0.62	0.63	0.64	0.64	0.64

Abbreviations: $AUC^{I/D}(x)$, incident/dynamic AUC at year x ; $AUC^{C/D}(x)$, cumulative/dynamic AUC at year x ; Cox ROC, estimate based on Cox model through `risksetAUC` function; Cox prob, estimate based on Cox model through transition probabilities; HR(DM), hazard ratio of DM.

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

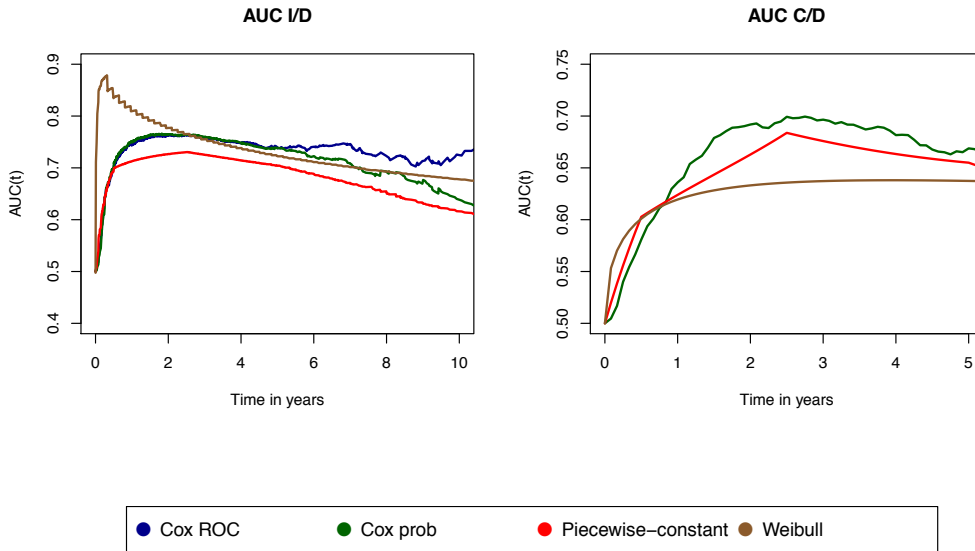


Figure 3.3: Time-specific AUC for distant metastasis. **Abbreviations:** AUC I/D, incident/dynamic AUC; AUC C/D, cumulative/dynamic AUC; Cox ROC, estimate based on Cox model through `risksetAUC` function; Cox prob, estimate based on Cox model through transition probabilities.

§3.6 Discussion

The illness-death model is frequently applied to clinical data to describe disease progression. A patient enters the model disease free, he can then experience disease and die. In clinical practice however, often the time of disease cannot be observed exactly. The information is interval-censored or unobserved because of death or censoring. This can lead to bias in the estimation of disease incidence and regression coefficients.[86, 97]

This article studied the predictive accuracy of a binary time-dependent disease marker in the context of the illness-death model for interval-censored data. A simulation study with several data scenarios was conducted to study four different models: the Cox model with disease as time-dependent marker, the piecewise-constant model implemented in the `msm` package, the Weibull model, and the M-spline model implemented in the `SmoothHazard` package. Both incident/dynamic and cumulative/dynamic AUC estimates were derived from their transition probabilities and studied. The methods were applied to a data set of soft tissue sarcoma patients who were scanned for distant metastasis at scheduled follow-up visits.

The simulation study showed that the HRs from the piecewise-constant model were less biased than those of the Cox model. The number of patients per data set (1000 vs 2000) did not have a large effect on the estimates of the HR, AUC estimates in incident/dynamic and cumulative/dynamic definition except for the M-spline model. The M-spline model converged more reliably with large data sets. The

spacing of follow-up visits at which the disease state was observed did have a large effect on estimates of the incident/dynamic AUC. The Weibull model showed the best performance, however this model had an unfair advantage since the simulated data had Weibull distribution. In practice a Weibull distribution may not be a good fit to the data. The M-spline model showed a good performance when estimating the incident/dynamic and cumulative/dynamic AUC however was not always able to converge and provide AUC estimates. The piecewise-constant model under performed. Even though, incident/dynamic AUC estimates had less bias than the Cox model's for scenarios with large spacing between follow-up visits, cumulative/dynamic estimates had the largest bias of all methods.

Prediction models are becoming more and more important in clinical practice to provide individualized patient care. Dynamic prediction models can incorporate time-dependent disease markers and the predictive accuracy of such a marker may be of interest. In the presence of interval-censored disease time, the results of this study suggest to take the interval-censoring into account not only when estimating parameters of the model, but also when evaluating the predictive accuracy of disease.

Simulations performed studied the effect of an interval-censored binary disease marker. Future research should focus on the predictive accuracy of a time-dependent covariate that can take more than 2 values as well as continuous markers.

Appendix

§3.A Derivation of AUC

§3.A.1 Incident/dynamic AUC

Let i, j be individuals, $X_i(t), X_j(t)$ the binary covariate values at time t , and T_i and T_j the death times. The incident/dynamic AUC is defined as

$$\begin{aligned}
 \text{AUC}^{I/D}(t) &= P(X_i(t) > X_j(t) \mid T_i = t, T_j > t) + 0.5P(X_i(t) = X_j(t) \mid T_i = t, T_j > t) \\
 &= P(X_j(t) = 0 \mid T_j > t)P(X_i(t) = 1 \mid T_i = t) \\
 &\quad + 0.5[P(X_j(t) = 0 \mid T_j > t)P(X_i(t) = 0 \mid T_i = t) + \\
 &\quad \quad P(X_j(t) = 1 \mid T_j > t)P(X_i(t) = 1 \mid T_i = t)] \\
 &= (1 - P(X_j(t) = 1 \mid T_j > t))P(X_i(t) = 1 \mid T_i = t) \\
 &\quad + 0.5[(1 - P(X_j(t) = 1 \mid T_j > t))(1 - P(X_i(t) = 1 \mid T_i = t)) + \\
 &\quad \quad P(X_j(t) = 1 \mid T_j > t)P(X_i(t) = 1 \mid T_i = t)] \\
 &= (1 - \pi_1(t))p(t) + 0.5[(1 - \pi_1(t))(1 - p(t)) + \pi_1(t)p(t)] \\
 &= p(t) - \pi_1(t)p(t) + 0.5[1 - p(t) - \pi_1(t) + \pi_1(t)p(t) + \pi_1(t)p(t)] \\
 &= p(t) - \pi_1(t)p(t) + 0.5 - 0.5p(t) - 0.5\pi_1(t) + \pi_1(t)p(t) \\
 &= 0.5 + 0.5(p(t) - \pi_1(t)),
 \end{aligned}$$

where

$$\pi_1(t) = P(X_i(t) = 1 \mid T_i > t)$$

$$p(t) = P(X_i(t-) = 1 \mid T_i = t).$$

§3.A.2 Cumulative/dynamic AUC

Let i, j be individuals, $X_i(s), X_j(s)$ their binary covariate values at time s , and T_i and T_j their death times. The cumulative/dynamic AUC is then

$$\begin{aligned}
\text{AUC}^{C/D}(s, t) &= P(X_i(s) > X_j(s) \mid T_i > s, T_i \leq t, T_j > t) + \\
&\quad 0.5P(X_i(s) = X_j(s) \mid T_i > s, T_i \leq t, T_j > t) \\
&= P(X_j(s) = 0 \mid T_j > t)P(X_i(s) = 1 \mid T_i > s, T_i \leq t) \\
&\quad + 0.5[P(X_j(s) = 0 \mid T_j > t)P(X_i(s) = 0 \mid T_i > s, T_i \leq t) + \\
&\quad P(X_j(s) = 1 \mid T_j > t)P(X_i(s) = 1 \mid T_i > s, T_i \leq t)] \\
&= (1 - P(X_j(s) = 1 \mid T_j > t))P(X_i(s) = 1 \mid T_i > s, T_i \leq t) \\
&\quad + 0.5[(1 - P(X_j(s) = 1 \mid T_j > t))(1 - P(X_i(s) = 1 \mid T_i > s, T_i \leq t)) + \\
&\quad P(X_j(s) = 1 \mid T_j > t)P(X_i(s) = 1 \mid T_i > s, T_i \leq t)] \\
&= (1 - \pi_1(s, t))p(s, t) + 0.5[(1 - \pi_1(s, t))(1 - p(s, t)) + \pi_1(s, t)p(s, t)] \\
&= p(s, t) - \pi_1(s, t)p(s, t) + 0.5[1 - p(s, t) - \pi_1(s, t) + \pi_1(s, t)p(s, t) + \pi_1(s, t)p(s, t)] \\
&= p(s, t) - \pi_1(s, t)p(s, t) + 0.5 - 0.5p(s, t) - 0.5\pi_1(s, t) + \pi_1(s, t)p(s, t) \\
&= 0.5 + 0.5(p(s, t) - \pi_1(s, t)),
\end{aligned}$$

where

$$\pi_1(s, t) = P(X_j(s) = 1 \mid T_j > t)$$

$$p(s, t) = P(X_i(s) = 1 \mid T_i > s, T_i \leq t).$$

§3.B Results for all scenarios

Table 3.B.1: Effect of disease.

Scenario	N	Censoring	Follow-up	Model	Mean(coef)	exp(mean(coef))	SE(coef)	Bias(coef)	RMSE(coef)
Truth					2.42		11.20		
A	1000	unif(5, 10)		3 Cox ROC	2.35		10.44	0.09	0.11
A	1000	unif(5, 10)		3 Cox prob	2.35		10.44	0.09	0.11
A	1000	unif(5, 10)		3 Piecewise-constant	2.43		11.32	0.09	0.09
B	1000	unif(5, 10)		6 Cox ROC	2.29		9.91	0.10	0.16
B	1000	unif(5, 10)		6 Cox prob	2.29		9.91	0.10	0.16
B	1000	unif(5, 10)		6 Piecewise-constant	2.48		11.90	0.10	0.12
C	1000	unif(5, 10)		12 Cox ROC	2.24		9.37	0.11	0.21
C	1000	unif(5, 10)		12 Cox prob	2.24		9.37	0.11	0.21
C	1000	unif(5, 10)		12 Piecewise-constant	2.41		11.11	0.12	0.12
D	1000		10	3 Cox ROC	2.35		10.45	0.09	0.11
D	1000		10	3 Cox prob	2.35		10.45	0.09	0.11
D	1000		10	3 Piecewise-constant	2.45		11.59	0.08	0.09
E	1000		10	6 Cox ROC	2.31		10.05	0.09	0.14
E	1000		10	6 Cox prob	2.31		10.05	0.09	0.14
E	1000		10	6 Piecewise-constant	2.50		12.19	0.10	0.13
F	1000		10	12 Cox ROC	2.25		9.47	0.10	0.20
F	1000		10	12 Cox prob	2.25		9.47	0.10	0.20
F	1000		10	12 Piecewise-constant	2.43		11.34	0.11	0.11
G	2000	unif(5, 10)		3 Cox ROC	2.34		10.43	0.07	0.10
G	2000	unif(5, 10)		3 Cox prob	2.34		10.43	0.07	0.10
G	2000	unif(5, 10)		3 Piecewise-constant	2.42		11.30	0.06	0.06
H	2000	unif(5, 10)		6 Cox ROC	2.29		9.91	0.07	0.14
H	2000	unif(5, 10)		6 Cox prob	2.29		9.91	0.07	0.14
H	2000	unif(5, 10)		6 Piecewise-constant	2.47		11.88	0.07	0.09
I	2000	unif(5, 10)		12 Cox ROC	2.24		9.38	0.08	0.20
I	2000	unif(5, 10)		12 Cox prob	2.24		9.38	0.08	0.20
I	2000	unif(5, 10)		12 Piecewise-constant	2.41		11.09	0.08	0.08
J	2000		10	3 Cox ROC	2.34		10.40	0.06	0.10
J	2000		10	3 Cox prob	2.34		10.40	0.06	0.10
J	2000		10	3 Piecewise-constant	2.44		11.52	0.06	0.07
K	2000		10	6 Cox ROC	2.30		10.00	0.06	0.13
K	2000		10	6 Cox prob	2.30		10.00	0.06	0.13
K	2000		10	6 Piecewise-constant	2.49		12.10	0.07	0.10
L	2000		10	12 Cox ROC	2.24		9.40	0.07	0.19
L	2000		10	12 Cox prob	2.24		9.40	0.07	0.19
L	2000		10	12 Piecewise-constant	2.42		11.24	0.08	0.08

Abbreviations: SE, empirical standard error; RMSE, root mean square error.

Table 3.B.2: Time-specific incident/dynamic AUC.

Scenario	Model	AUC ^{I/D} (1) = 0.71			AUC ^{I/D} (3) = 0.72			AUC ^{I/D} (5) = 0.72		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
A	Cox ROC	-0.04	0.02	0.04	-0.02	0.01	0.03	-0.02	0.01	0.02
A	Cox prob	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
A	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03

Table 3.B.2: (continued)

Scenario	Model	$AUC^{I/D}(1) = 0.71$			$AUC^{I/D}(3) = 0.72$			$AUC^{I/D}(5) = 0.72$		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
A	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
A	M-spline	0.02	0.02	0.03	0.00	0.02	0.02	-0.01	0.04	0.04
B	Cox ROC	-0.07	0.02	0.08	-0.04	0.01	0.04	-0.03	0.01	0.04
B	Cox prob	-0.07	0.02	0.08	-0.04	0.01	0.04	-0.03	0.01	0.03
B	Piecewise-constant	-0.06	0.01	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
B	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
B	M-spline	0.01	0.03	0.03	-0.01	0.03	0.03	0.00	0.04	0.04
C	Cox ROC	-0.21	0.00	0.21	-0.07	0.02	0.08	-0.06	0.02	0.06
C	Cox prob	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
C	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.02	0.05
C	Weibull	-0.01	0.03	0.03	0.00	0.02	0.02	0.00	0.02	0.02
C	M-spline	-0.05	0.04	0.06	0.00	0.03	0.03	0.00	0.04	0.04
D	Cox ROC	-0.04	0.02	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
D	Cox prob	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
D	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03
D	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	M-spline	0.02	0.02	0.03	-0.01	0.02	0.03	0.00	0.03	0.04
E	Cox ROC	-0.07	0.02	0.08	-0.04	0.01	0.04	-0.03	0.02	0.03
E	Cox prob	-0.07	0.02	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
E	Piecewise-constant	-0.06	0.01	0.06	-0.04	0.01	0.04	-0.03	0.01	0.03
E	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
E	M-spline	0.01	0.02	0.03	-0.01	0.03	0.03	0.00	0.04	0.04
F	Cox ROC	-0.21	0.00	0.21	-0.07	0.02	0.08	-0.06	0.02	0.06
F	Cox prob	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
F	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.01	0.05
F	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
F	M-spline	-0.05	0.04	0.06	0.00	0.03	0.03	0.00	0.04	0.04
G	Cox ROC	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
G	Cox prob	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
G	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03
G	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
G	M-spline	0.02	0.02	0.03	0.00	0.02	0.02	-0.01	0.02	0.03
H	Cox ROC	-0.07	0.01	0.08	-0.04	0.01	0.04	-0.03	0.01	0.03
H	Cox prob	-0.07	0.01	0.08	-0.04	0.01	0.04	-0.03	0.01	0.03
H	Piecewise-constant	-0.06	0.01	0.06	-0.04	0.01	0.04	-0.03	0.01	0.03
H	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
H	M-spline	0.01	0.02	0.02	-0.01	0.02	0.02	-0.01	0.03	0.03
I	Cox ROC	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
I	Cox prob	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
I	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.01	0.05
I	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
I	M-spline	-0.04	0.03	0.05	0.00	0.02	0.02	0.00	0.03	0.03
J	Cox ROC	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
J	Cox prob	-0.04	0.01	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
J	Piecewise-constant	-0.05	0.01	0.05	-0.03	0.01	0.03	-0.02	0.01	0.03
J	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
J	M-spline	0.02	0.01	0.03	-0.01	0.02	0.02	-0.01	0.02	0.03
K	Cox ROC	-0.07	0.01	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
K	Cox prob	-0.07	0.01	0.07	-0.04	0.01	0.04	-0.03	0.01	0.03
K	Piecewise-constant	-0.06	0.01	0.06	-0.04	0.01	0.04	-0.03	0.01	0.03
K	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
K	M-spline	0.01	0.02	0.02	-0.01	0.02	0.02	0.00	0.02	0.03
L	Cox ROC	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
L	Cox prob	-0.21	0.00	0.21	-0.07	0.01	0.07	-0.06	0.01	0.06
L	Piecewise-constant	-0.09	0.01	0.09	-0.06	0.01	0.06	-0.05	0.01	0.05
L	Weibull	-0.01	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01
L	M-spline	-0.04	0.03	0.05	0.00	0.02	0.02	0.00	0.03	0.03

Abbreviations: $AUC^{I/D}(x)$, incident/dynamic AUC at year x ; SE, empirical standard error; RMSE, root mean square error.

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

Table 3.B.3: (continued)

Scenario	Model	$AUC^{C/D}(1) = 0.59$			$AUC^{C/D}(3) = 0.62$			$AUC^{C/D}(5) = 0.64$		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE

Table 3.B.3: Time-specific cumulative/dynamic AUC.

Scenario	Model	$AUC^{C/D}(1) = 0.59$			$AUC^{C/D}(3) = 0.62$			$AUC^{C/D}(5) = 0.64$		
		Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
A	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
A	Piecewise-constant	-0.02	0.01	0.02	-0.01	0.01	0.02	-0.01	0.01	0.02
A	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
A	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
B	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
B	Piecewise-constant	-0.04	0.00	0.04	-0.02	0.01	0.03	-0.02	0.01	0.02
B	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
B	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
C	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
C	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
C	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
C	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
D	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	Piecewise-constant	-0.02	0.01	0.02	-0.01	0.01	0.02	-0.01	0.01	0.02
D	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
D	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
E	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
E	Piecewise-constant	-0.04	0.00	0.04	-0.03	0.01	0.03	-0.02	0.01	0.02
E	Weibull	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
E	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
F	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
F	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
F	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
F	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.02
G	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
G	Piecewise-constant	-0.02	0.00	0.02	-0.01	0.01	0.01	-0.01	0.01	0.01
G	Weibull	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01
G	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
H	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
H	Piecewise-constant	-0.04	0.00	0.04	-0.02	0.01	0.02	-0.02	0.01	0.02
H	Weibull	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.01
H	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
I	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
I	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
I	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
I	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
J	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
J	Piecewise-constant	-0.02	0.00	0.02	-0.01	0.01	0.01	-0.01	0.01	0.01
J	Weibull	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
J	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
K	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
K	Piecewise-constant	-0.04	0.00	0.04	-0.02	0.01	0.03	-0.02	0.01	0.02
K	Weibull	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01
K	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
L	Cox	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
L	Piecewise-constant	-0.05	0.00	0.05	-0.04	0.01	0.04	-0.03	0.01	0.03
L	Weibull	-0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01
L	M-spline	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01

Abbreviations: $AUC^{C/D}(x)$, cumulative/dynamic AUC at year x ; SE, empirical standard error; RMSE, root mean square error.

Table 3.B.4: Invalid estimations for M-spline model.

	Scenario											
	A	B	C	D	E	F	G	H	I	J	K	L
invalid AUC ^{I/D}	379	442	555	17	34	88	201	292	378	1	7	19
invalid AUC ^{C/D}	381	444	560	17	34	89	201	296	381	1	7	19

Number of invalid estimations of AUC^{I/D} and AUC^{C/D} from 1 year based on 1000 data sets.

§3.C Incident/dynamic AUC for scenarios D–L

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

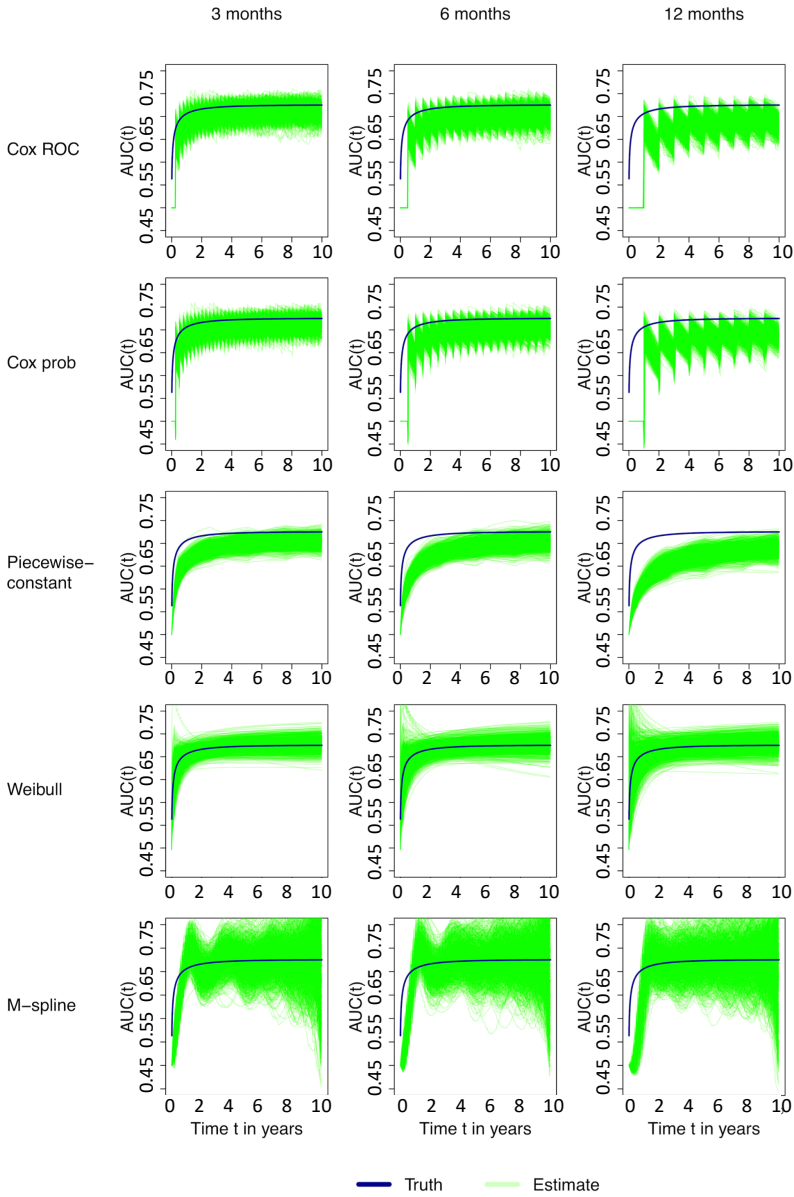


Figure 3.C.1: Incident/dynamic AUC for scenario D (3 months), E (6 months) and F (12 months). **Abbreviations:** Cox ROC, estimate based on *risksetAUC* function; Cox prob, estimate based on transition probabilities of Cox model.

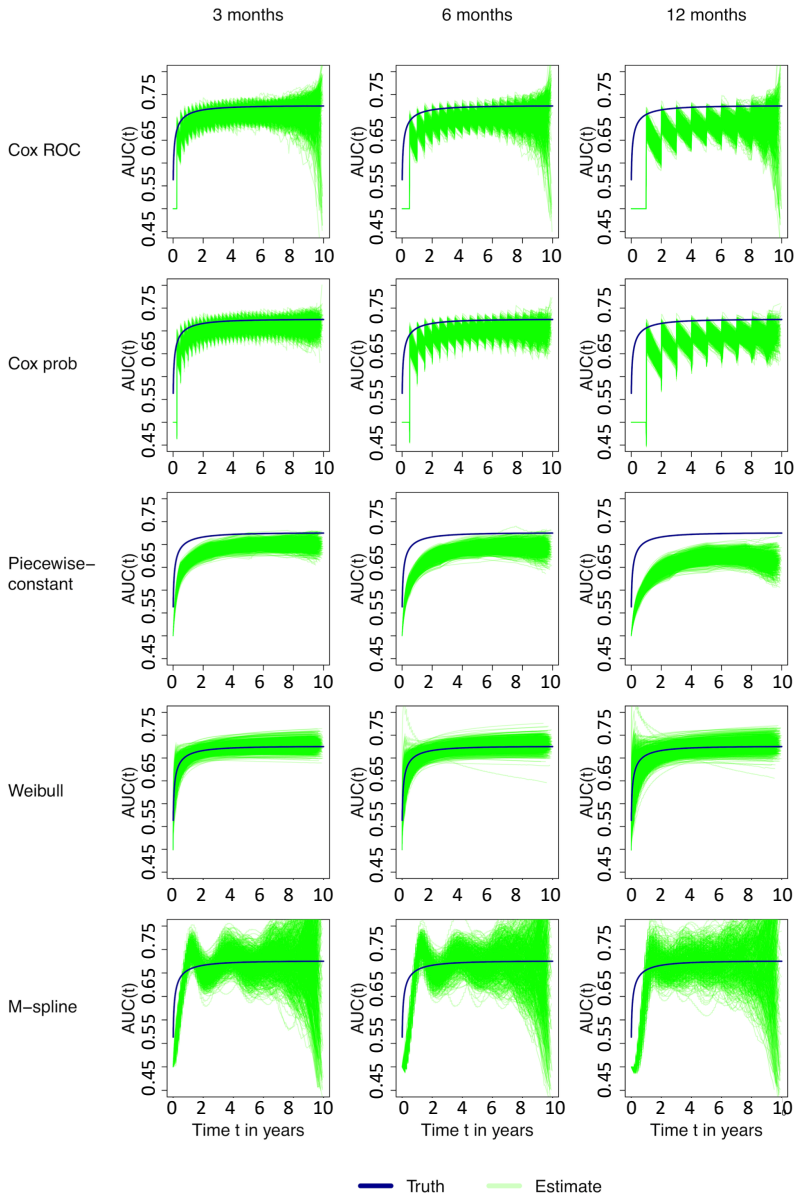


Figure 3.C.2: Incident/dynamic AUC for scenario G (3 months), H (6 months) and I (12 months). **Abbreviations:** Cox ROC, estimate based on *risksetAUC* function; Cox prob, estimate based on transition probabilities of Cox model.

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

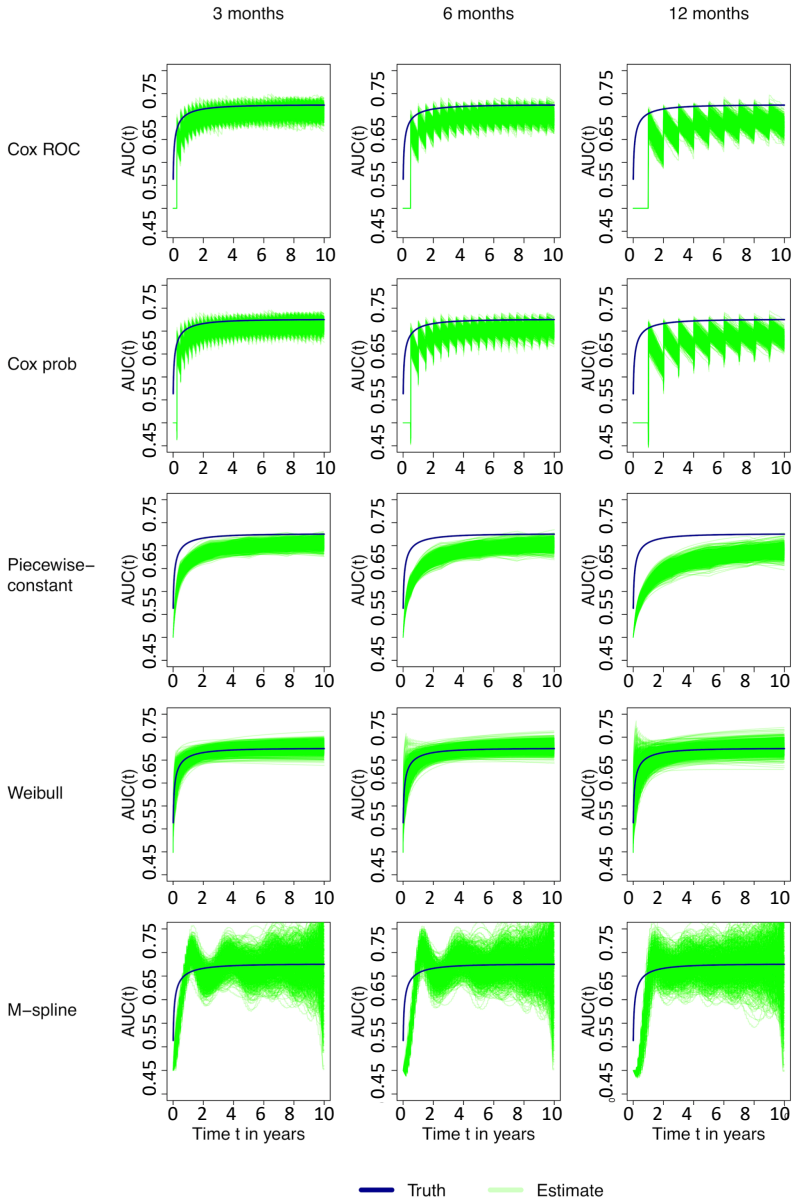


Figure 3.C.3: Incident/dynamic AUC for scenario J (3 months), K (6 months) and L (12 months). **Abbreviations:** Cox ROC, estimate based on *risksetAUC* function; Cox prob, estimate based on transition probabilities of Cox model.

§3.D Cumulative/dynamic AUC for scenarios D–L

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

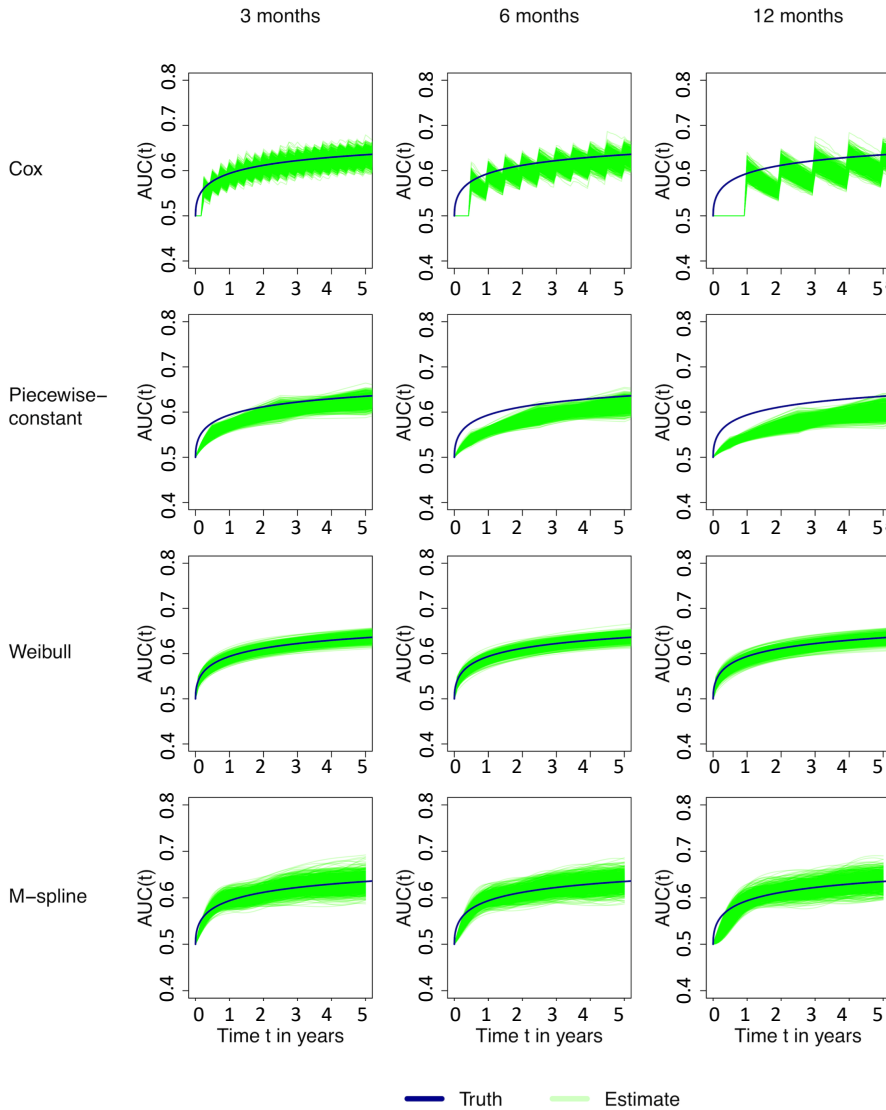


Figure 3.D.1: Cumulative/dynamic AUC for scenario D (3 months), E (6 months) and F (12 months).

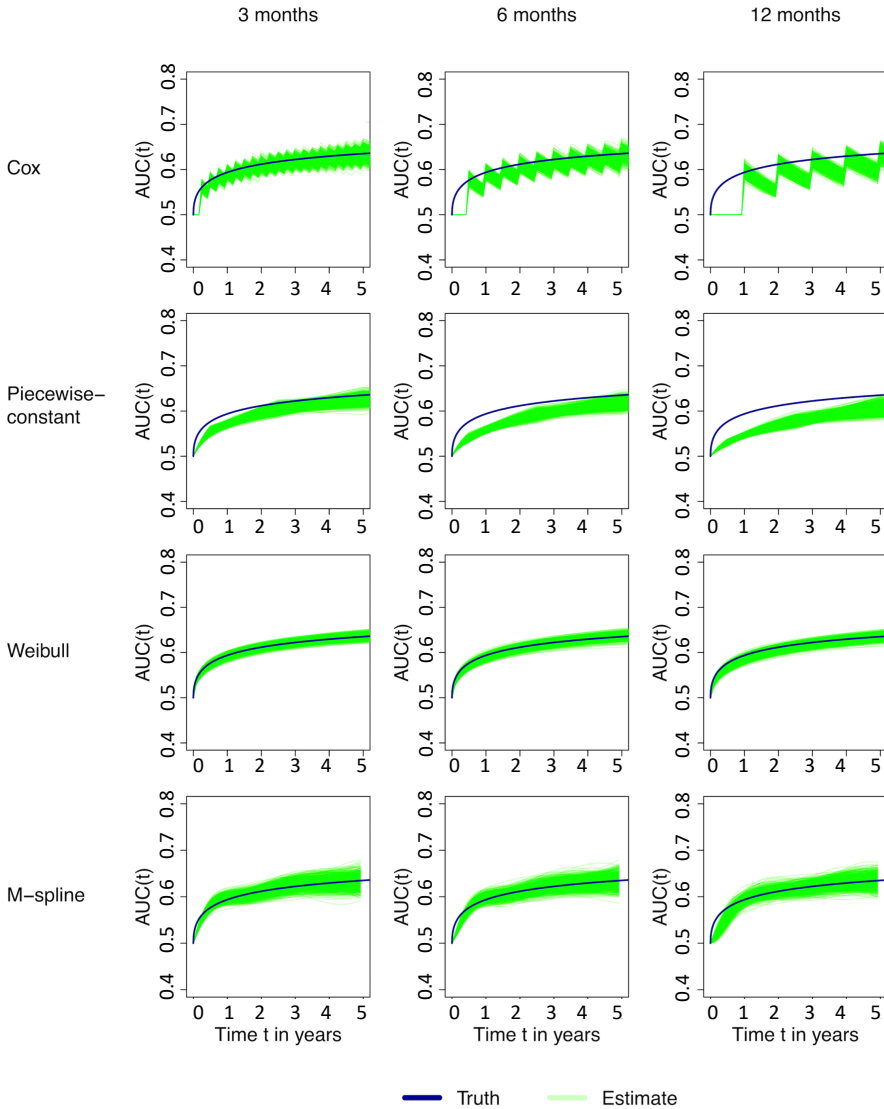


Figure 3.D.2: Cumulative/dynamic AUC for scenario G (3 months), H (6 months) and I (12 months).

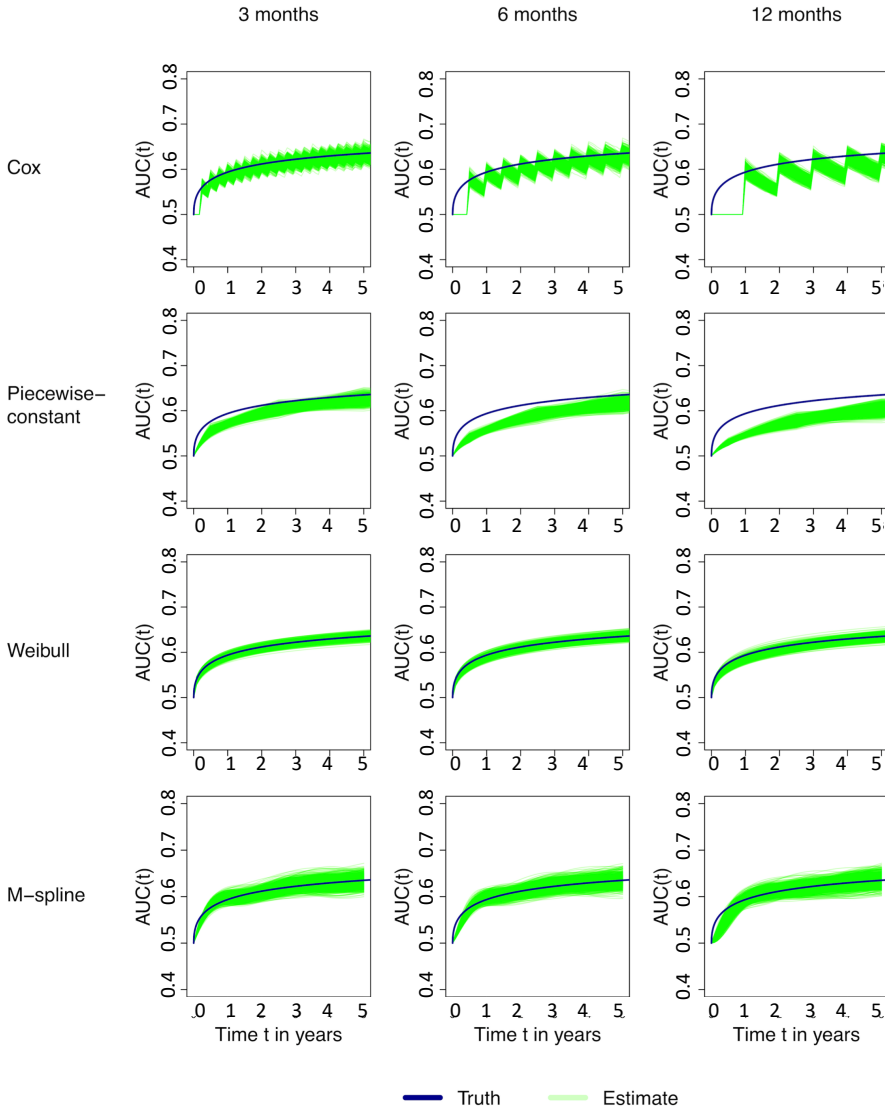


Figure 3.D.3: Cumulative/dynamic AUC for scenario J (3 months), K (6 months) and L (12 months).

3. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker

Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

This chapter has been published in *BMJ Open* 7(2) (2017) doi: 10.1136/bmjopen-2016-012930 as A.J. Rueten-Budde, et al., "Individualised risk assessment for local recurrence and distant metastases in a retrospective transatlantic cohort of 687 patients with high-grade soft tissue sarcomas of the extremities: a multistate model" [21].

Abstract

Objectives: This study investigates the effect of surgical margins and radiotherapy, in the presence of individual baseline characteristics, on survival in a large population of high-grade soft tissue sarcoma of the extremities using a multistate model.

Design: A retrospective multicentre cohort study.

Setting: 4 tertiary referral centres for orthopaedic oncology.

Participants: 687 patients with primary, nondisseminated, high-grade sarcoma only, receiving surgical treatment with curative intent between 2000 and 2010 were included.

Main outcome measures: The risk to progress from 'alive without disease' (ANED) after surgery to 'local recurrence' (LR) or 'distant metastasis (DM)/death'. The effect of surgical margins and (neo)adjuvant radiotherapy on LR and overall survival was evaluated taking patients' and tumour characteristics into account.

Results: The multistate model underlined that wide surgical margins and the use of neoadjuvant radiotherapy decreased the risk of LR but have little effect on survival. The main prognostic risk factors for transition ANED to LR are tumour size (HR 1.06; 95% CI 1.01 to 1.11 (size in cm)) and (neo)adjuvant radiotherapy. The HRs

4. Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

for patients treated with adjuvant or no radiotherapy compared with neoadjuvant radiotherapy are equal to 4.36 (95% CI 1.34 to 14.24) and 14.20 (95% CI 4.14 to 48.75), respectively. Surgical resection margins had a protective effect for the occurrence of LR with HRs equal to 0.61 (95% CI 0.33 to 1.12), and 0.16 (95% CI 0.07 to 0.41) for margins between 0 and 2 mm and wider than 2 mm, respectively. For transition ANED to distant metastases/Death, age (HR 1.64 (95% CI 0.95 to 2.85) and 1.90 (95% CI 1.09 to 3.29) for 25- 50 years and >50 years, respectively) and tumour size (1.06 (95% CI 1.04 to 1.08)) were prognostic factors.

Conclusions: This paper underlined the alternating effect of surgical margins and the use of neoadjuvant radiotherapy on oncological outcomes between patients with different baseline characteristics. The multistate model incorporates this essential information of a specific patient's history, tumour characteristics and adjuvant treatment modalities and allows a more comprehensive prediction of future events.

§4.1 Introduction

Soft tissue sarcomas (STS) are a rare, heterogeneous group of tumours accounting for ~1% of all adult cancers.[137] Approximately 60% of all STS occur in the extremities.[41] High-grade STS are a select subgroup (representing 38% of all STS in one series [83]) of highly aggressive and infiltrative subtypes with an overall poor prognosis.[117, 169] At present, limb salvage surgery with (neo) adjuvant radiotherapy is the standard of care for most patients, while the role of chemotherapy is more controversial.[56] However, locally recurrent disease (LR), distant metastases (DM) and poor survival remain of great concern. Although the risk factors for the occurrence of these adverse events have been the subject of many studies, a solid prognostic profile for individual patients is still lacking.

Considering an individual patient's treatment, two types of prognostic factors can be identified: those that are set at the moment of diagnosis and those that are treatment-related. Prognostic factors such as histology, grade, depth and size [83, 117, 169, 53, 73, 146, 140, 98, 139, 29] are generally recognised and set at the moment of diagnosis. At present, surgical resection margin and the administration of (neo) adjuvant radiotherapy/chemotherapy are the only treatment factors that can be influenced. The intended resection margin is part of an intricate balance between the best oncological outcome and maintenance of quality of life, including limb function. Whether limb function should be sacrificed to achieve a negative or wide margin should be based on its effect on the overall prognosis of that specific patient.

Although the increased risk of LR following an intralesional margin resection is generally recognised,[73, 139, 91] the presence of possible associations between margin status and overall survival (OS) or between LR and OS is still under discussion. Results have been published confirming the absence [140, 106, 164] and presence [29, 23][74, 110, 102, 111] of a prognostic role for margins as well as LR on OS.

Unfortunately, current literature on prognostic factors for STS faces several difficulties: small sample sizes, heterogeneity of study populations and differences in statistical methods applied.[53, 102] Results from prior studies may, therefore, be misleading when applied to an individual patient with a high-grade STS. In an era where clinicians are moving towards individualised patient treatment, it would be preferable to consider the results of planned resection margins for each patient individually. The great importance of individualised cancer treatment is generally accepted because awareness has been created that certain patients have a higher risk of disease recurrence or death than others, and others are more susceptible to possible adverse effects of treatment.

This study aims to investigate the effect of margins and radiotherapy, considering individual patient characteristics, on LR and survival in a large population with only high-grade STS of the extremities using a multistate model. Better stratification of risks will lead to better treatment decisions and improved clinical results for patients with high-grade STS.

§4.2 Patients and methods

A retrospective multicentre analysis of patients surgically treated between 2000 and 2010 for primary, nondisseminated, high-grade (as defined by FNCLCC larger than grade 2) sarcoma, including angiosarcoma, malignant peripheral nerve sheath tumour, synovial sarcoma, spindle cell sarcoma, myxofibrosarcoma and (pleomorphic) STS not-otherwise-specified was performed. All cases were discussed preoperatively in multidisciplinary teams and pretreatment staging was performed with lung CT scans. Postoperative surveillance strategies were comparative between all centres with yearly MRI for local control and chest X-ray/CT scan every 3-4 months according to ESMO guidelines.[56]

Patients were identified from the local sarcoma databases of the four participating institutions, all tertiary referral centres for orthopaedic oncology. Exclusion criteria were metastatic disease at the time of diagnosis, presentation with recurrent disease, treatment without curative intent (ie, no primary intent of (limb-sparing) surgery with intended sufficient margins), adjuvant treatment other than radiotherapy or chemotherapy and an unknown margin status. Initially, 709 patients received treatment in 1 of 4 participating centres and met the inclusion criteria. Five patients met the exclusion criteria and were excluded. Seventeen patients were excluded because there was insufficient information on all covariates.

Medical records including surgical notes and pathology reports were reviewed and the following information was recorded: age (<25; 25-50; >50 years[148]), gender, presentation status (no treatment/biopsy only vs incomplete excision elsewhere prior to referral), tumour size (cm), depth (superficial vs deep to investing fascia), location (upper vs lower extremity), surgical margin, (neo) adjuvant therapy (neoadjuvant, adjuvant, no radiotherapy; chemotherapy vs no chemotherapy) and follow-up data.

Experienced musculoskeletal pathologists in each centre defined the closest surgical margin. Owing to the lack of an international consensus on the definition of margin descriptions, the resection margins were categorised quantitatively: tumour at the inked surface of the resection specimen (0 mm); tumour within 2 mm of ink; tumour at more than 2 mm of ink. The 2 mm cut-off point was based on previous research that identified this as the most optimal differentiating distance.[164, 87]

The decision concerning the use of (neo) adjuvant treatment was not uniform during the study period due to variation in management over time and by centre, although the majority of patients (75%) received radiotherapy. The most common radiotherapy regimens were 50 Gy preoperatively (22.4%) or 50-66 Gy postoperatively (52.3%).

LR was defined as the first radiological or pathological manifestation of tumour within or contiguous to the previously treated tumour bed, 2 or more months after primary treatment. DM was defined by clinical or radiological evident systemic spread of tumour outside the primary tumour bed, including nodal metastasis. Dates of death were extracted from the medical records and local or national death registries.

§4.2.1 Statistical analysis

Multivariate Cox regression model

The effect of prognostic factors on OS was estimated with a Cox regression model with LR included as a time-dependent covariate. The following risk factors were included in the model: age at diagnosis, presentation, tumour location, size, depth, histopathology subtype, surgical margin, limb sparing and radiotherapy. HRs based on the multivariate Cox regression model and their corresponding 95% CIs were estimated.

Multistate model

Disease progression was investigated with a multistate model.[119] A multistate model is a model for time-to-event data, in which all individuals start in one or possibly more starting states (eg, surgery) and eventually may move in one (or more) state(s), for example, progressive distant disease, LR or death. In this approach, transitions are assessed during the course of the disease and prognostic factors for each transition are studied. Figure 4.1 shows the multistate model applied in this study to describe the disease progression. We propose three possible states in which a patient may be at any time. After surgery, a patient may be alive with no evidence of disease (ANED), alive with LR or may have developed DM and subsequent death (Death). In this analysis, the two states death and DM were pooled into one state (DM/Death) since DM will, with very few exceptions, inevitably lead to death; among the 288 patients who developed metastatic disease, 88% had died. Patients with concurrent LR and DM (diagnosed within 3 months of each other; n=30) were registered as entering the state of DM/Death. The direction of arrows in Figure 4.1 indicates the transitions between states. The time scale used is months since definitive surgery.

To estimate the effect of age at diagnosis, presentation, tumour location, size (in cm), depth, histopathology subtype, surgical margin achieved, limb sparing, and radiotherapy on each transition, a Cox proportional hazards (PH) model was used. For transition 3 (LR to DM/Death), the effects of tumour depth, histopathology subtype, surgery type and radiotherapy could not be estimated due to the relatively small number of patients in this transition. Therefore, these covariates were omitted from the model for this specific transition. The PH assumption in the Cox model was tested for each transition.

Individual risk assessment

Multistate models[119] can be used with two different purposes. The first aim is to obtain more biological insight into the disease/recovery process of a patient. It is of interest to determine how certain prognostic factors influence different phases of the evolution of the disease. The second purpose is prediction, as these models help clinicians to obtain more accurate predictions on survival and to adjust predictions by incorporating the occurrence of intermediate events. Predictions are made by estimating the conditional probabilities of future events, given the treatment and patient characteristics.

4. Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

Patient-specific state occupation probabilities presented in stacked charts provide insight into the effect of margins on the occurrence of events after surgery, given the characteristics of a patient. The stacked charts present a visual aid for surgeons to investigate the effect of margin on the probability of being in different states (LR or DM/Death) at different time points after surgery. The multistate model provides information on the ever-changing nature of a specific patient's history and allows a more comprehensive understanding of the data.

The beginning and end of follow-up corresponded to the date of definitive surgery and the last date of follow-up or death, respectively. The median follow-up was assessed by employing the reverse Kaplan-Meier method.[133] The effect of risk factors was estimated by adjusted HRs along with their 95% CIs. *p* Values at or below 0.05 were considered significant. In the analysis, the variable 'centre' was included to account for the presence of heterogeneity between the four treatment centres. All analyses concerning the multistate model were performed using the R-package *mstate* (R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Vienna, Austria 2011. <http://www.r-project.org/>).[48, 49]

§4.3 Results

Table 4.1 summarises patients' demographics and treatments at baseline for the included 687 patients.

The estimated 5-year OS was 52.7% (95% CI 48.8% to 56.6%) with a median follow-up of 71 (95% CI 67 to 75) months. In total, 106 patients (15%) developed LR; however, only 59 patients (9%) developed isolated LR, while the other 47 patients (6%) developed LR synchronous or following DM. In total, 288 (42%) developed DM. Seventy-two patients (10%) died without known DM or LR.

A traditional multivariate Cox regression model with LR as a time-dependent covariate showed a significant effect of age (HR 2.22; 95% CI 1.25 to 3.92 for >50 years compared with <25 years), tumour size (HR 1.06 for every cm; 95% CI 1.04 to 1.08) and actual LR (HR 3.42; 95% CI 2.55 to 4.60) on OS (Table 4.2). Note that tumour size is given in centimetre, implying that a 'k' cm change in size multiplies the hazard by HR^k . For example, an HR equal to 1.34 (95% CI 1.22 to 1.47) and 1.79 (95% CI 1.48 to 2.16) are associated with a tumour of size 5 and 10 cm, respectively. Estimated HRs for histopathology with respect to the reference group angiosarcoma are shown in Table 4.2. Radiotherapy violated the PH assumption and was incorporated in the analysis by fitting a stratified Cox model in which a separate baseline hazard is used for patients with and without (neo) adjuvant radiotherapy.

In the multistate model depicted in Figure 4.1, the number of patients moving from one state to the other is illustrated. The majority moved from the state ANED to DM/Death directly (n=340; 49%). In 42% of the patients (n=288), no further disease was detected; therefore, they remained in their postoperative state ANED. A small group (n=59; 9%) developed LR first, after which 36 of these 59 patients (61%) moved to the final state DM/ Death. To estimate the adjusted HRs for each transition, a multivariate Cox proportional hazard regression model was employed (Table 4.3).

Table 4.1: Patient demographics and treatment characteristics

Characteristics	
Age, mean (SD), years	57.9 (19.8)
Age, no. (%)	
<25	49 (7.1)
25-50	170 (24.7)
>50	468 (68.1)
Gender, no. (%)	
Male	389 (56.6)
Female	298 (43.4)
Tumour presentation, no. (%)	
Primary	555 (80.8)
'Whoops'*	132 (19.2)
Tumour location, no. (%)	
Upper extremity	162 (23.6)
Lower extremity	525 (76.4)
Tumour size, mean (SD), cm	10.0 (6.2)
Depth, no. (%)	
Deep	531 (77.3)
Superficial	115 (16.7)
Deep and superficial	41 (6)
Histopathology, no. (%)	
Angiosarcoma	19 (2.8)
MPNST	81 (11.8)
Myxofibrosarcoma	217 (31.6)
Synovial sarcoma	134 (19.5)
Spindle cell sarcoma	165 (24.0)
Sarcoma NOS	17 (2.5)
MFH/UPS	54 (7.9)
Surgical margin, no. (%)	
0 mm	114 (16.6)
≤2 mm	325 (47.3)
>2 mm	248 (36.1)
Type of surgery, no. (%)	
Limb-sparing	611 (88.9)
Amputation	76 (11.1)
Radiotherapy, no. (%)	
Neoadjuvant	154 (22.4)
Adjuvant	359 (52.3)
No radiotherapy	174 (25.3)
(Neo)Adjuvant chemotherapy, no. (%)	
Yes	82 (11.9)
No	605 (88.1)

Notation: *Incomplete excision elsewhere prior to referral; MPNST, malignant peripheral nerve sheath tumour; NOS, not otherwise specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma.

4. Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

Table 4.2: Cox regression analysis for overall survival

Variable	p Value	HR	95% CI
Age			
<25		1	
25-50	0.115	1.59	0.89–2.82
>50	0.006	2.22	1.25–3.92
Tumour presentation ('whoops'* vs primary)	0.828	1.04	0.75–1.43
Tumour location (lower vs upper)	0.336	1.14	0.87–1.50
Tumour size, cm	0.000	1.06	1.04–1.08
Depth			
Deep		1	
Superficial	0.561	0.90	0.64–1.28
Deep and superficial	0.877	1.04	0.63–1.71
Histopathology			
Angiosarcoma		1	
MPNST	0.005	3.29	1.43–7.54
Myxofibrosarcoma	0.060	2.15	0.97–4.78
Synovial sarcoma	0.027	2.59	1.12–6.02
Spindle cell sarcoma	0.030	2.51	1.09–5.77
Sarcoma NOS	0.057	2.66	0.97–7.27
MFH/UPS	0.025	2.68	1.13–6.37
Surgical margin (mm)			
0		1	
≤2	0.433	0.89	0.66–1.20
>2	0.319	0.83	0.58–1.20
Type of surgery (limb-sparing vs amputation)	0.478	0.86	0.56–1.31
Local recurrence (yes vs no)**	0.000	3.42	2.55–4.60

Notation: *Incomplete excision elsewhere prior to referral; MPNST, malignant peripheral nerve sheath tumour; NOS, not otherwise specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; ** time-dependent variable.

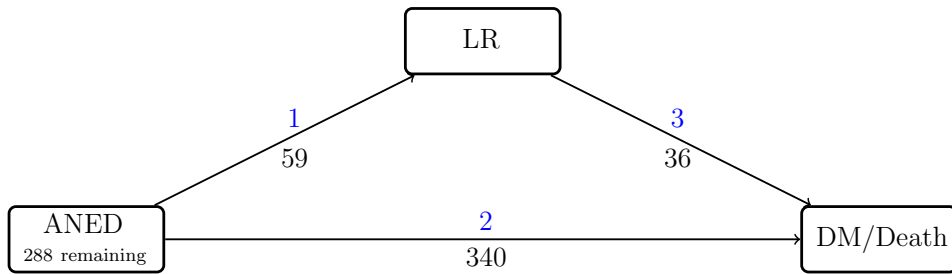


Figure 4.1: Disease progression of high-grade soft tissue sarcomas represented in a multistate model. Blue, transition number; black, number of patients moving from one state to another. ANED, alive, no evidence of disease; LR, local recurrence; DM, distant metastasis.

The main prognostic risk factors for transition 1 (ANED to LR) are tumour size (HR 1.06; 95% CI 1.01 to 1.11 with size in cm) and (neo) adjuvant radiotherapy. The HRs for patients treated with adjuvant or no radiotherapy compared with neoadjuvant radiotherapy are equal to 4.36 (95% CI 1.34 to 14.24) and 14.20 (95% CI 4.14 to 48.75), respectively (Table 4.3). Surgical resection margins had a protective effect on the occurrence of LR with HRs equal to 0.61 (95% CI 0.33 to 1.12) and 0.16 (95% CI 0.07 to 0.41) for margins between 0 and 2 mm and wider than 2 mm, respectively. No statistically significant effect of margins was detected when patients move directly to the state DM/Death from ANED (transition 2). The effect of age on the transition between ANED and DM/Death (transition 2) is equal to 1.64 (95% CI 0.95 to 2.85) and 1.90 (95% CI 1.09 to 3.29) for patients aged 25-50 years and >50 years, respectively, compared with patients <25 years of age. The HR for tumour size (in cm) is equal to 1.06 (95% CI 1.04 to 1.08). There was no significant effect of prognostic factors on the transition hazards between LR and DM/Death (transition 3). There was no significant difference between the centres for each outcome in the classical Cox model and the multistate model.

The estimated multistate model was used to predict outcome probabilities for each specific patient. Estimates of these probabilities are based on the results obtained from the Cox model on the transition hazards between the states. Different resection margins and patient characteristics are considered. The patientspecific state occupation probabilities at different time points after surgery are visualised in stacked charts (Figure 4.2). For any individual patient, three separate charts show the effect of resection margins, in the presence of patient, tumour and (neo) adjuvant treatment characteristics. The distance between two curves represents the probability of being in a specific state (ANED, or LR or DM/Death) at a specific time point. Figure 4.2 illustrates the three margin scenarios for three different patients. After surgery, the probability of occupying the state ‘LR’ (green area) decreases as margins increase in the two patients receiving adjuvant radiotherapy, while the probability of occupying the state ‘ANED’ (light blue area) increases as margins increase. The probability of

4. Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

Table 4.3: HRs and 95% CIs for all prognostic factors and all transitions in the multistate model

Variable	Trans 1: ANED →LR			Trans 2: ANED →DM/Death			Trans 3: LR →DM/Death		
	p Value	HR	95% CI	p Value	HR	95% CI	p Value	HR	95% CI
Age									
<25		1			1			1	
25-50	0.649	0.76	0.23–2.50	0.077	1.64	0.95–2.85	0.413	0.50	0.10–2.60
>50	0.955	1.03	0.32–3.31	0.023	1.90	1.09–3.29	0.302	0.47	0.11–1.97
Tumour presentation (‘whoops’* vs primary)	0.344	1.43	0.68–3.03	0.586	0.91	0.66–1.26	0.539	1.39	0.48–4.03
Tumour location (lower vs upper)	0.116	0.61	0.33–1.13	0.919	1.01	0.78–1.32	0.474	1.43	0.54–3.83
Tumour size, cm	0.018	1.06	1.01–1.11	0.000	1.06	1.04–1.08	0.114	1.05	0.99–1.12
Depth									
Deep		1			1			1	
Superficial	0.093	0.51	0.23–1.12	0.653	0.92	0.66–1.30			
Deep and superficial	0.226	0.26	0.03–2.33	0.253	1.31	0.82–2.09			
Histopathology									
Angiosarcoma		1			1				
MPNST	0.034	0.23	0.06–0.90	0.845	1.08	0.51–2.26			
Myxofibrosarcoma	0.085	0.34	0.10–1.16	0.777	0.90	0.44–1.84			
Synovial sarcoma	0.023	0.21	0.05–0.80	0.972	0.99	0.47–2.07			
Spindle cell sarcoma	0.078	0.32	0.09–1.14	0.910	0.96	0.46–2.01			
Sarcoma NOS	0.918	0.90	0.13–6.14	0.702	0.82	0.31–2.22			
MFH/UPS	0.032	0.19	0.04–0.87	0.560	1.26	0.58–2.76			
Surgical margin (mm)									
0		1			1			1	
≤2	0.113	0.61	0.33–1.12	0.211	0.82	0.61–1.12	0.746	1.15	0.50–2.62
>2	0.000	0.16	0.07–0.41	0.193	0.80	0.56–1.12	0.949	1.04	0.32–3.36
Type of surgery (limb-sparing vs amputation)	0.486	1.55	0.45–5.32	0.717	0.93	0.61–1.40			
Radiotherapy									
Neoadjuvant		1			1				
Adjuvant	0.015	4.36	1.34–14.24	0.840	0.96	0.63–1.46			
No radiotherapy	0.000	14.20	4.14–48.75	0.340	1.24	0.80–1.91			

Notation: *Incomplete excision elsewhere prior to referral; MPNST, malignant peripheral nerve sheath tumour; NOS, not otherwise specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; ** time-dependent variable.

occupying the state ‘DM/Death without LR’ (red area) decreases slightly for patient A (upper panels) as margins increase, while for patient B (middle panels), the probability remains almost the same for the first two margin scenarios and even increases for a margin wider than 2 mm. The probability of occupying the state ‘DM/Death after LR’ (orange area) decreases as the margin increases in patients A and B. Patient C received neoadjuvant radiotherapy and for this patient, the probability of occupying the state ‘LR’ (green area) is very low and it is not affected by the margin. A wider margin also appears to have little effect on the probability of occupying the state ‘DM/Death without LR’ (red area).

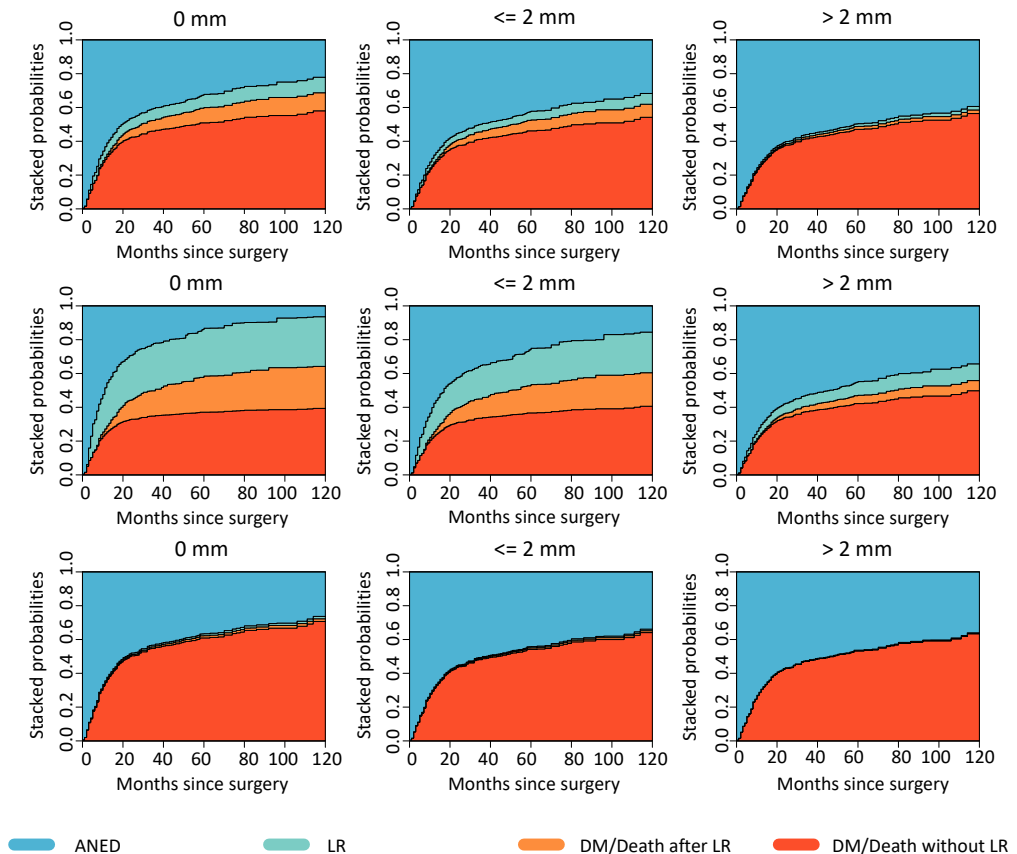


Figure 4.2: Stacked state occupation probabilities for patients for different margins after surgery, based on the model in figure 1. Upper panels: patient A: a woman aged 74 years with a large (>10 cm), high-grade myxofibrosarcoma of the upper leg, resection with adjuvant radiotherapy. Middle panels: patient B: a man aged 60 years with a 7 cm angiosarcoma of the arm, resection with adjuvant radiotherapy. Lower panels: patient C: a woman aged 70 years with a large (>10) synoviosarcoma of the upper leg, resection after neoadjuvant radiotherapy. From left to right: Left panels: a 0 mm margin. Middle panels: margins smaller than or equal to 2 mm. Right panels: margins wider than 2 mm.

§4.4 Discussion

High-grade STS are associated with frequent LRs and poor survival. Since several prognostic factors are set at baseline (ie, tumour size, grade), the resection margin and indication or timing of radiotherapy might be the only prognostic factors that can be affected by the multidisciplinary team. The results of this study stress the importance of individual prediction of survival, considering the different prognostic effects of radiotherapy and surgical margins between patients.

This study brings a new element into the discussion of the effect of margins by using a multistate model. The estimated state occupation probabilities based on the multistate model show the different effect of margins on outcomes between patients with different baseline characteristics and adjuvant treatment modalities. This implies that, in the discussion of the effect of margins, margins cannot be considered as a single entity, but only in combination with patient-specific baseline characteristics and additional radiotherapy. Although previous studies on the effect of margins take patient characteristics into account in their multivariate analysis, it has not earlier been emphasised and visualised how much these characteristics influence the effect of margins. To the best of our knowledge, the stacked charts presented here are the first visualisation of the complex relationship between prognostic factors and probabilities of disease progression for individual patients. An additional asset of the multistate model is that future disease progression can be estimated based on the baseline characteristics of a patient at diagnosis, as well as on his known disease progression after surgery. This enables real-time updates of future outcomes when additional information becomes available over time.

The results from this study can be applied in clinical practice by taking the probabilities of future state occupation for a specific patient into account when weighing invasive surgery against maintaining quality of life, especially in cases with limited expected survival. However, the authors acknowledge that the presented data are too intricate to directly apply in daily practice. Therefore, a user-friendly web-based tool based on the multistate model presented in this study will be developed.

This study presents new knowledge on the effect of neoadjuvant radiotherapy in patients with high-grade STS. In clinical practice, the difference in the effect of preoperative and postoperative use of radiotherapy on LR, DM and survival of patients with high-grade STS of the extremities remains the subject of discussion. Surgery is delayed ~3 months in patients receiving preoperative radiotherapy, compared with patients receiving no or postoperative radiotherapy. Therefore, it is important to assess the effect of our surgical planning and the use of radiotherapy on the course of the disease. The current results show that patients receiving neoadjuvant radiotherapy were less likely to develop LR when compared with patients with no or adjuvant radiotherapy, even though the 95% CI was large. This is consistent with previously published results,[10] although others did not find a true difference in the risk of LR.[112, 108] One recent large retrospective database study showed that neoadjuvant radiotherapy was associated with improved survival.[127] This is in contrast to several other studies that showed no significant effect of timing of radiotherapy on overall survival.[170, 94] Since all these trials face the limitations of retrospective studies, a firm conclusion is

still not possible. Possibly, a larger randomised trial will be able to provide a decisive answer on which sequence is superior.

Undeniably, the question of the definition of a marginal or wide margin remains. Multiple different descriptions are used in the literature.[81] In contrast to its continued use, the Enneking classification[55] is not considered detailed enough in respect of (large) STS with close involvement of essential structures such as vessels, nerves and bone.[111] In addition, the use of (neo) adjuvant radiotherapy has decreased the necessity for radical or even wide margins.[111] The dichotomous classification proposed by Trovik et al[146] may be too simplistic regarding adequate or inadequate margins. While the poor prospect associated with macroscopically intralesional resections is evident, the implications of microscopically positive or marginal resections should not be regarded as identical.[111, 87] The quantitative measurement as applied in this study did not take into account the type of tissue of which the margin consisted (eg, fascia, fat), which might also influence the required minimum width of a margin.[111, 90] As Hoang et al[81] recently proposed, a universally updated surgical margin reporting system would improve communication and understanding regarding surgical treatment of STS. To create a broad basis for such a global system, international collaboration is needed.

The main strengths of this study are its large cohort of high-grade extremity STS only and the use of a multistate model to investigate the evolution of the disease and to estimate the probabilities of clinical future events, given a set of individual patient characteristics. The estimates of these probabilities are based on the results obtained from the Cox model on the transition hazards between the states. The study population is limited to high-grade extremity tumours of the most common sarcoma types, and thus, the results are not attenuated by a diversity of low-grade, low-impact STS. Finally, this study introduces the possibility of a practical aid for clinical practice that would allow for individually tailored treatments, in contrast to many previous studies that provide general prognostic factors for treatment decisions based on groups of patients. Several limitations exist in this study. First, the inherent effects of a retrospective study design, such as selection bias, are present. Second, owing to the multicentre aspect of the study, a revision of all histological data was not possible. However, all centres reported pathology results in the same manner. Margin width as stated in the pathology reports was used for the analyses instead of descriptive results. Additionally, all analyses were corrected for centre effect and there was no significant difference between centres. Despite the limitations, the current analysis is the largest investigation into the effect of margins on LR and OS for patients with high-grade extremity STS.

This study stresses the importance of patient-specific characteristics when evaluating the effect of surgical margins and (neo) adjuvant radiotherapy. On the basis of the estimated state occupation probabilities, the effect of margin differs significantly in individual cases depending on baseline characteristics and the administration of (neo) adjuvant radiotherapy. To use prognostic factors for LR and DM/Death in daily practice and thereby enable personalised care, a user-friendly webbased tool (application) based on the model presented in this study will be validated and published.

4. Individualised risk assessment for local recurrence and distant metastases in for patients with high-grade soft tissue sarcomas of the extremities: a multistate model

A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

This chapter has been published in *European Journal of Cancer* 83 (2017) 313–323 as A.J. Rueten-Budde, et al., "A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)" [20].

Abstract

Background: To support shared decision-making, we developed the first prediction model for patients with primary soft-tissue sarcomas of the extremities (ESTS) which takes into account treatment modalities, including applied radiotherapy (RT) and achieved surgical margins. The PERsonalised SARcoma Care (PERSARC) model, predicts overall survival (OS) and the probability of local recurrence (LR) at 3, 5 and 10 years.

Aim: Development and validation, by internal validation, of the PERSARC prediction model.

Methods: The cohort used to develop the model consists of 766 ESTS patients who underwent surgery, between 2000 and 2014, at five specialised international sarcoma centres. To assess the effect of prognostic factors on OS and on the cumulative incidence of LR (CILR), a multivariate Cox proportional hazard regression and the Fine and Gray model were estimated. Predictive performance was investigated by using internal cross validation (CV) and calibration. The discriminative ability of the model was determined with the C-index.

Results: Multivariate Cox regression revealed that age and tumour size had a significant effect on OS. More importantly, patients who received RT showed better outcomes, in terms of OS and CILR, than those treated with surgery alone. Internal validation of the model showed good calibration and discrimination, with a C-index of 0.677 and 0.696 for OS and CILR, respectively.

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

Conclusions: The PERSARC model is the first to incorporate known clinical risk factors with the use of different treatments and surgical outcome measures. The developed model is internally validated to provide a reliable prediction of post-operative OS and CILR for patients with primary high-grade ESTS.

§5.1 Introduction

Multimodality treatment of high-grade soft-tissue sarcomas of the extremities (ESTS) has improved over the years; however, local recurrence (LR), distant metastasis (DM) and poor survival remain of great concern [164]. Although the effect of several patient-related prognostic factors on overall survival (OS) and LR is well described, the lack of a validated prediction model that includes treatment modalities complicates decision-making aimed at balancing oncologic cure and minimising the risk of disability after treatment.

Factors such as vascular invasion[54], peripheral tumour growth[54], tumour size [54, 107, 74, 29], infiltrative growth[54], necrosis[54], site[107], adjuvant chemo- and/or radiotherapy (RT) [45], histological grade [107, 74, 29] (for fibro- and liposarcomas [145]) and histological subtype [107, 74] have been shown to have a significant impact on survival. While some studies indicate that the prognostic value of tumour depth[54], state at presentation [45], tumour site [102] and age [102] remains unclear, others found some of these factors to be good predictors of outcome[107, 74, 29]. The effect of limb sparing surgery and neoadjuvant chemotherapy and/or RT remains debatable [45]. Surgical margins have an impact on LR [164, 74], but no clear association with OS has been established [164, 74].

In 2003, a prognostic model based on 175 patients with ESTS became available [77] and expanded twice [38, 128]. The first update included patients who were diagnosed at a time (1967) when magnetic resonance imaging (MRI) was not part of the standard care. Prognostic factors included in those studies were tumour size, vascular invasion, necrosis, grade, peripheral growth, depth and location, whereas age, gender, recurrence and metastasis, margins and histology were not included in the model. Callegaro et al. (2016) developed two nomograms for soft-tissue sarcomas of the ESTS and trunk using age, tumour size, histological grade and subtype, using exclusively patients with macroscopically complete surgical resections [37]. In addition, several models only provide prognosis for OS and DM, whereas others underline the relevance of LR. Willeumier et al. (2017) underlined the importance of individual prognostication of LR and OS based on different combinations of surgical margins and possible (neo) and/or adjuvant therapy, while also taking different patient and tumour characteristics into account [21].

To support shared decision-making between patients and physicians, this study aims to develop a prognostic Personalised Sarcoma Care (PERSARC) model to predict the cumulative incidence of LR (CILR) and OS for a patient with high-grade ESTS with specific clinical characteristics and possible treatment modalities at baseline. The prediction model is internally validated by calibration and discrimination.

§5.2 Methods

This multicentre study was approved by each of our hospitals' human subjects review boards.

§5.2.1 Study population

The study population included a consecutive series of 838 patients with primary high-grade ESTS who underwent surgical treatment at one of the five international collaborating hospitals between January 2001 and December 2014. Due to missing values for 72 patients, 766 individuals were included in development of the PERSARC model. Eligible diagnoses included high-grade (Fédération Nationale des Centres de Lutte Contre le Cancer [FNCLCC] grade III) angiosarcoma, malignant peripheral nerve sheath tumour, synovial sarcoma, spindle cell sarcoma, myxofibrosarcoma and (pleomorphic) soft-tissue sarcomas not-otherwise-specified. Excluded patients include those that were treated without curative intent, had LR or DM within 2 months after primary treatment (ruled out by pre-treatment and follow-up (FU) staging with lung computed tomography (CT) scan), had a tumour in their abdomen, thorax, head or neck or received (neo) adjuvant treatment other than RT or chemotherapy.

All collaborating sarcoma centres implemented the guidelines of the European Society for Medical Oncology for soft-tissue sarcoma FU [56]. Patients visited the outpatient clinic for their scheduled clinical and radiographic FU: every 3-4 months in the first 2-3 years, then every 6 months and after 5 years yearly. It was common that FU was ended after 10 years evidence of no disease.

§5.2.2 Study design

This was a retrospective observational study, in which clinical information was gathered retrospectively (medical records) and by using existing prospective sarcoma databases (including documentation of clinic visits, operation reports, histology and radiographic reports). This information included demographics (centre, patient gender and age at presentation, event and FU), tumour characteristics (presentation, localisation, depth, diameter, histology and grade), treatment characteristics (goal, time of operation [weeks], resection margin and categorical, type and dose of [neo] adjuvant therapy), development of LR and/or DM and last known status. All patients had a minimal FU of 2 years or experienced an event (LR, DM or death) before. The primary outcome measure was survival, if the patient was alive at their last documented visit information on the tumour status was gathered. Secondary outcome measure was LR. Long-term FU was obtained through reviewing documentation of all clinical and radiographic FU.

A sarcoma was considered primary if it was previously untreated, a biopsy or whoops excision had been performed before presentation at one of the five contributing specialised sarcoma centres, with no evidence of metastatic disease. LR was defined as the presence of viable tumour at the site of the original tumour bed confirmed by clinical findings, pathological tissue diagnosis or radiological reports more than 2 months after primary surgery. Distant recurrence was defined by clinical or radiological evidence of systemic spread of tumour outside the primary tumour bed.

Tumour size was defined as maximum diameter at pathologic analysis. In patients that received neo-adjuvant RT and/or chemotherapy, tumour size was defined as maximum diameter measured by CT or MRI before treatment. Surgical margin

was defined as follows: intralesional for tumour cells present at the margin of the resection specimen (<0.1 mm), marginal for tumour cells found within 0.1-2 mm of the margin and free for tumour cells found at least 2 mm away from the margin [citewilleumier, rueten2017, kainhofer2016]. Tumour grade was classified as high-grade based on established criteria of the FNCLCC.

§5.2.3 Statistical analysis

Multivariate Cox regression model

To assess the effect of prognostic factors on OS a multivariate Cox proportional hazards regression model was used. Predictor variables incorporated in the model were age, tumour size, depth, histology subtype, surgical margin and RT. Initially, tumour site and tumour presentation were considered; however, previous studies [37] and an initial multivariate analysis (Wald test p-value: tumour site $p = 0.818$, tumour presentation $p = 0.696$) showed no strong predictive value.

Fine and Gray model

To estimate the effect of risk factors on the CILR, a competing risks model, which accounts for the competing event death was used (Appendix 5.A, Figure 5.A.1) [119]. After surgery, a patient may be alive with no evidence of disease. He may then develop LR or die. The cumulative incidence function is defined as the probability of the event occurring before a certain time point. Fine and Gray's method models the effect of covariates on the cumulative incidence in the presence of competing events. Subdistribution hazard ratios (sHRs) estimate the effect of risk factors on the probability of event occurrence directly. The same covariates used in the Cox model were considered.

Prediction and validation

Predictions for OS and LR at 3, 5 and 10 years after surgery together with 95% confidence intervals (95% CIs), which indicate the uncertainty surrounding the estimates are provided. To justify their use in clinical practice, predictive performance of the prediction models was assessed internally by using leave-one-out cross validation (CV). CV is a technique to simulate model performance on new data.

Following van Houwelingen (2000), a prognostic model is defined as a rule to compute probabilities, given the relevant covariates and their validity can be argued on the basis of model calibration.

Calibration refers to how well predicted probabilities agree with observed probabilities. A common practice is to group patients from 'good' to 'bad' prognosis. A model is well calibrated if true and predicted group probabilities do not differ.

The prediction model can be used to categorise patients based on their prognosis. A patient's risk factor information can be summarised into a prognostic index (PI), which is a weighted mean of prognostic variables, where weights are derived from the prognostic model. Patients with a higher value of PI have a higher predicted risk.

Hence, the PI can be used to divide data into four equal sized groups as follows: ‘good prognosis’, ‘fairly good prognosis’, ‘fairly poor prognosis’ and ‘poor prognosis’.

Calibration plots visualise model calibration on a given set of data [150, 165]. Data are divided into prognostic groups. At specific time points the groups’ observed outcome (OS or CILR) is plotted against their predicted outcome. If the points are scattered around the diagonal ($x = y$), the model is valid without adjustment. To investigate calibration for data subgroups, one-sampled T-tests are used, where predicted outcomes were considered the ‘fixed’ value and observed outcomes as the evaluated variable [47].

Discrimination refers to the ability of the model to assign higher predicted risk to patients who experience the event earlier compared with those experiencing the event later or not at all. To visualise this aspect, non-parametric curves are plotted showing the observed outcome (OS or CILR) for different prognostic groups [124]. The spread of the curves indicates how well a model can discriminate. The C-index quantifies discrimination as the proportion of patient pairs that experience events in the order of risk predicted [149]. It can be adjusted for competing risks [165] and can be interpreted as follows: a C-index of 1 means that the model has perfect discrimination and a C-index of 0.5 means that the model predicts just as well as flipping a coin [9].

All statistical analysis was conducted using R software [122]. A p-value of 0.05 was defined as statistically significant.

§5.3 Results

Table 5.1 summarises patients’ characteristics at baseline for the included 766 patients from the five international sarcoma centres. The median FU was 71.8 months (95% CI: 67.6-75.9), assessed with the reverse Kaplan-Meier method. In total, 369 patients died and 116 developed an LR. The majority of patients with an LR died ($n = 83$; 72%). OS was estimated to be equal to 63%, 53% and 39% at 3, 5 and 10 years, respectively. CILR was estimated to be equal to 13.3% (95% CI: 10.9-15.8), 15.1% (95% CI: 12.4-17.7) and 17.2% (95% CI: 13.9-20.5) at 3, 5 and 10 years, respectively. The centre effect on disease progression was investigated but no significant effect was found.

Age, tumour size and additional RT show an independent significant effect on OS (Table 5.2). Patients with larger tumours have a significantly increased risk of dying with HR equal to 1.068 (95% CI: 1.052-1.085) for a unit increase of 1 cm. Older age is associated with a higher risk of death with HR equal to 1.195 (95% CI: 1.116-1.268) for a 10-year increase in age. Note that age and size are included as linear terms in the model, implying that a ‘k*10’ year change in age and a ‘k’ cm change in size multiply the hazard by HRk. Surgical margin has a marginally significant effect on OS, with HR equal to 0.786 (95% CI: 0.599-1.033) and 0.711 (95% CI: 0.524-0.964) for margin equal to 0.1-2 mm and >2 mm, respectively (reference category 0 mm). RT treatment is associated with a decreased risk of dying compared with surgery alone with HRs equal to 0.548 (95% CI: 0.399-0.753) and 0.638 (95% CI: 0.486-0.837) for neoadjuvant and adjuvant RT, respectively.

Table 5.1: Patient characteristics

Characteristics	N(%)
Total	766
Age, mean (SD), years	58.28 (19.39)
Age (%)	
30-60 years	281 (36)
< 30 years	82 (11)
> 60 years	403 (53)
Sex (%)	
Male	435 (57)
Female	331 (43)
Depth (%)	
Deep	579 (76)
Superficial	134 (17)
Deep and superficial	53 (7)
Tumour size, mean (SD), cm	10.06 (6.21)
Tumour location, no. (%)	
Upper extremity	182 (24)
Lower extremity	584 (76)
Tumour presentation (%)	
Primary	622 (81)
'Whoops'*	144 (18)
Histopathology (%)	
Myxofibrosarcoma	238 (31)
MPNST	91 (12)
Synovial sarcoma	142 (18)
Spindle cell sarcoma	167 (22)
MFH/UPS	77 (10)
Other	51 (7)
Surgical margin (%)	
0 mm	140 (18)
≤2 mm	343 (45)
>2 mm	283 (37)
Limb-sparing (%)	
No	81 (11)
Yes	685 (89)
Radiotherapy, no. (%)	
Neoadjuvant	184 (24)
Adjuvant	400 (52)
No radiotherapy	182 (24)

Notation: N, number of patients; *Incomplete excision elsewhere prior to referral; MPNST, malignant peripheral nerve sheath tumour; NOS, not otherwise specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; Depth: relative to the investing fascia.

Table 5.2: Multivariate Cox model for overall survival: hazard ratio (HR) along with 95% confidence interval ($n = 766$).

	HR	95% CI	p-Value
Age	1.195	1.116–1.268	<0.001
Tumour size	1.068	1.052–1.085	<0.001
Depth			0.377
Deep	1		
Superficial	0.813	0.591–1.117	
Deep and superficial	1.110	0.736–1.674	
Histopathology			0.492
Myxofibrosarcoma	1		
MPNST	1.422	0.989–2.044	
Synovial sarcoma	1.261	0.869–1.831	
Spindle cell sarcoma	1.211	0.884–1.661	
MFH/UPS	1.293	0.890–1.876	
Surgical margin			0.080
0 mm	1		
≤2 mm	0.786	0.599–1.033	
>2 mm	0.711	0.524–0.964	
Radiotherapy			
No RT	1		
Neoadjuvant	0.548	0.399–0.753	
Adjuvant	0.638	0.486–0.837	

The HR of age corresponds to a unit increase of 10 years, and the HR of size corresponds to a unit increase of 1 cm. **Notation:** CI, confidence interval; HR, hazard ratio; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; MPNST, malignant peripheral nerve sheath tumour; RT, radiotherapy. Depth: relative to the investing fascia. Depth: relative to the investing fascia

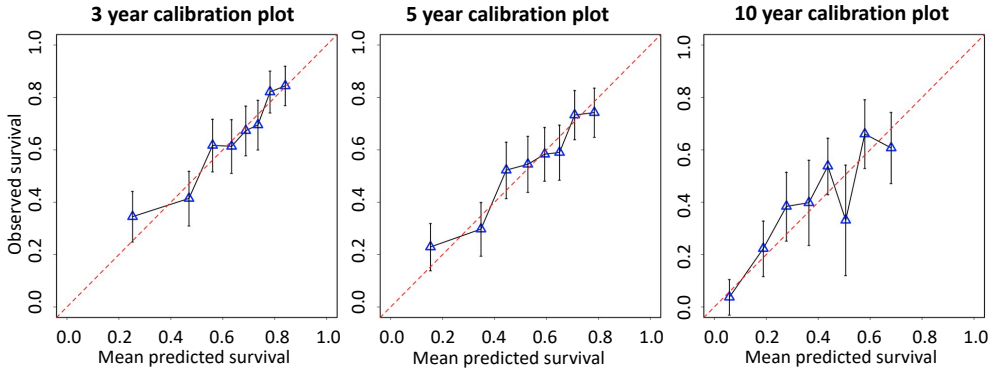


Figure 5.1: Calibration plots for overall survival. Observed survival obtained using Kaplan-Meier estimator is plotted against predicted survival for patients in eight equal sized risk groups identified by their predicted survival at (A) 3 years, (B) 5 years and (C) 10 years, as assessed by cross validation.

Figure 5.1 shows calibration plots for OS at 3, 5 and 10 years. The 3-, 5- and 10-year calibration plots show points (representing risk groups) scattered close to the diagonal, which is contained in the 95% CIs of the observed group survival.

A detailed comparison of observed and predicted survival at 3, 5 and 10 years for data subgroups is given in Table 5.3. Observed and predicted outcome do not differ significantly; however, for smaller and medium sized tumours (<5 cm, 5-10 cm) survival is underestimated at 3, 5 and 10 years, respectively.

Figure 5.2 shows good discrimination of the model visualised by the spread of the Kaplan-Meier estimates (solid lines). Model-based estimates (dotted lines) show the mean predicted survival per group close to the observed survival, indicating good calibration.

The C-index for OS was estimated to be 0.677 (95% CI 0.643-0.701).

In the Fine and Gray model, tumour size, surgical margin and RT show a significant effect on CILR (Table 5.4). Bigger tumours are associated with a higher probability of LR with sHR equal to 1.031 (95% CI: 1.001-1.063) for a unit increase of 1 cm. Patients with larger margins have a significantly lower CILR with sHR equal to 0.635 (95% CI: 0.406-0.992) and 0.282 (95% CI: 0.159-0.500) for 0.1-2 mm and >2 mm, respectively. RT treatment is associated with a lower CILR compared with surgery alone with sHRs equal to 0.312 (95% CI: 0.146-0.668) and 0.700 (95% CI: 0.417-1.175) for neoadjuvant and adjuvant RT, respectively.

Calibration plots for LR are shown in Figure 5.3. Points are scattered around the lower diagonal that lies within the 95% CIs of the observed cumulative incidence, indicating a good calibration. However, the small distance between lower risk groups and the fact that groups observed outcome not always monotonically increases indicate the relative difficulty to discriminate among patients with lower risk profiles.

Figure 5.4 shows crude cumulative incidence curves (solid lines) and model-based estimates (dotted lines) computed as the mean predicted cumulative incidence for LR. The high-risk groups can clearly be distinguished from the rest. However, the

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

Table 5.3: Comparing observed and predicted overall survival, assessed by cross validation, for subgroups of data at 3, 5 and 10 years.

	n (%)	3 years			5 years			10 years		
		Pred	Obs (se)	Diff (% 95 CI)	Pred	Obs (se)	Diff (% 95 CI)	Pred	Obs (se)	Diff (% 95 CI)
Age										
30-60	281 (36.7)	68.9	70.7 (2.8)	-1.8 (-7.3 to 3.7)	60.2	60.5 (3.1)	-0.3 (-6.4 to 5.8)	46.4	45.4 (4.7)	1.0 (-8.2 to 10.2)
<30	82 (10.7)	77.8	74.6 (4.9)	3.2 (-6.4 to 12.8)	70.7	68.9 (5.3)	1.8 (-8.6 to 12.2)	58.7	58.3 (6.9)	0.4 (-3.1 to 3.9)
>60	403 (52.6)	54.0	54.9 (2.6)	-0.9 (-6.0 to 4.2)	43.6	44.4 (2.7)	-0.8 (-6.1 to 4.5)	29.1	28.9 (4.1)	0.2 (-7.8 to 8.2)
Size										
<5cm	123 (16.1)	77.2	87.0 (3.1)	-9.8 (-15.9 to -3.7)	69.8	78.2 (4.1)	-8.4 (-16.4 to -0.4)	57.2	57.8 (8.1)	-0.6 (-16.5 to 15.3)
5cm-10cm	295 (38.5)	69.4	68.7 (2.8)	0.7 (-4.8 to 6.2)	60.3	59.7 (3.0)	0.6 (-5.3 to 6.5)	45.8	53.1 (3.6)	-7.3 (-14.4 to -0.2)
>=10cm	348 (45.4)	50.4	49.3 (2.8)	1.1 (-4.4 to 6.6)	40.0	38.6 (2.8)	1.4 (-4.1 to 6.9)	26.0	20.7 (3.8)	5.3 (-2.1 to 12.7)
Depth*										
Deep	579 (75.6)	60.9	62.8 (2.1)	-1.9 (-6.0 to 2.2)	51.4	52.7 (2.2)	-1.3 (-5.6 to 3.0)	37.3	37.5 (3.1)	-0.2 (-6.3 to 5.9)
Superficial	134 (17.5)	69.3	67.6 (4.2)	1.7 (-6.5 to 9.9)	60.7	60.9 (4.5)	-0.2 (-9.0 to 8.6)	46.9	56.1 (4.9)	-9.2 (-18.8 to 0.4)
Deep and superficial	53 (6.9)	55.0	51.2 (7.0)	3.8 (-9.9 to 17.5)	45.7	37.9 (7.4)	7.8 (-6.7 to 22.3)	32.7	19.0 (10.2)	13.7 (-6.3 to 33.7)
Histology										
Myxofibrosarcoma	238 (31.1)	62.7	62.4 (3.2)	0.3 (-6.0 to 6.6)	53.3	53.4 (3.4)	-0.1 (-6.8 to 6.6)	39.1	36.3 (5.0)	2.8 (-7.0 to 12.6)
MFNST	91 (11.9)	60.0	57.7 (5.2)	2.3 (-7.9 to 12.5)	50.2	50.1 (5.4)	0.1 (-10.5 to 10.7)	35.9	33.4 (7.7)	2.5 (-12.6 to 17.6)
Synovial sarcoma	142 (18.5)	73.8	75.1 (3.8)	-1.3 (-8.7 to 6.1)	65.8	66.6 (4.2)	-0.8 (-9.0 to 7.4)	52.7	50.9 (5.9)	1.8 (-9.8 to 13.4)
Spindle cell sarcoma	167 (21.8)	55.6	59.9 (3.9)	-4.3 (-11.9 to 3.3)	45.4	47.4 (4.3)	-2.0 (-10.4 to 6.4)	30.9	41.5 (5.0)	-10.6 (-20.4 to -0.8)
MFH/UPS	77 (10.1)	53.7	54.8 (5.9)	-1.1 (-12.7 to 10.5)	43.60	44.8 (6.1)	-1.2 (-13.2 to 10.8)	29.6	29.2 (9.0)	0.4 (-17.2 to 18.0)
Margin										
0 mm	140 (18.3)	52.3	51.5 (4.3)	0.8 (-7.6 to 9.2)	42.40	46.8 (4.4)	-4.4 (-13.0 to 4.2)	28.8	37.1 (4.8)	-8.3 (-17.7 to 1.1)
0.1-2 mm	343 (44.8)	63.1	64.6 (2.6)	-1.5 (-6.6 to 3.6)	53.50	52.4 (2.8)	1.1 (-4.4 to 6.6)	39.1	37.9 (3.7)	1.2 (-6.1 to 8.5)
> 2 mm	283 (36.9)	65.5	66.4 (2.9)	-0.9 (-6.6 to 4.8)	56.50	57.5 (3.2)	-1.0 (-7.3 to 5.3)	42.9	39.5 (6.3)	3.4 (-8.9 to 15.7)
RT										
No RT	182 (23.8)	50.9	53.6 (3.8)	-2.7 (-10.1 to 4.7)	40.80	44.5 (3.9)	-3.7 (-11.3 to 3.9)	27.2	29.5 (6.1)	-2.3 (-14.3 to 9.7)
Neoadjuvant	184 (24)	69.3	69.2 (3.5)	0.1 (-6.8 to 7.0)	60.70	60.3 (3.8)	0.4 (-7.0 to 7.8)	47.0	40.3 (6.0)	6.7 (-5.1 to 18.5)
Adjuvant	400 (52.2)	63.7	64.1 (2.5)	-0.4 (-5.3 to 4.5)	54.30	53.7 (2.7)	0.6 (-4.7 to 5.9)	40.0	43.3 (3.1)	-3.3 (-9.4 to 2.8)

Notation: Pred, predicted; Obs, observed; se, standard error; Diff, difference; CI, confidence interval; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; MPNST, malignant peripheral nerve sheath tumour; RT, radiotherapy. * Depth: relative to the investing fascia.

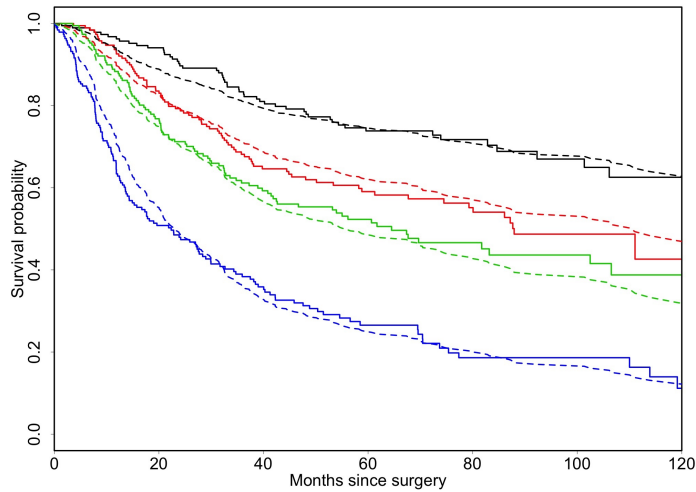


Figure 5.2: Survival curves for four prognostic index groups. Kaplan-Meier survival curves (solid lines) plotted with the model-based survival curves (dotted lines) for four different prognostic index groups. The numbers of patients at risk was 423, 265 and 33 at 3, 5 and 10 years, respectively. Black: patients with good; red: fairly good; green: fairly poor and blue: poor prognosis.

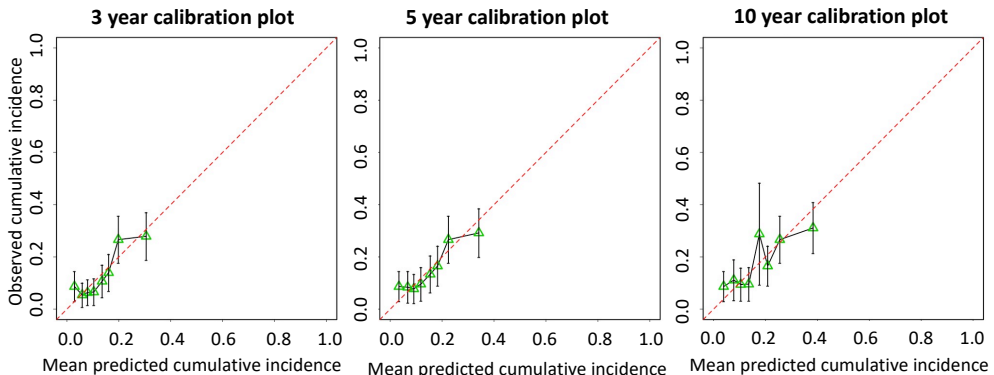


Figure 5.3: Calibration plots for local recurrence. Observed local recurrence (LR) is plotted against predicted LR for patients in eight equal sized risk groups identified by their predicted probability for LR, as assessed by cross validation.

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

Table 5.4: Fine and Gray model for local recurrence. Subdistribution hazard ratio (sHR) along with 95% confidence interval ($n = 766$).

	sHR	95% CI	p-Value
Age	1.051	0.942-1.184	0.337
Size	1.031	1.001-1.063	0.042
Depth*			0.559
Deep	1.000		
Superficial	0.907	0.536-1.535	
Deep & superfiscial	0.563	0.198-1.604	
Histology			0.864
Myxofibrosarcoma	1.000		
MPNST	1.079	0.580-2.009	
Synovial sarcoma	0.779	0.379-1.602	
Spindle cell sarcoma	0.979	0.570-1.681	
MFH/UPS	1.096	0.557-2.156	
Margin			<0.001
0 mm	1.000		
0.1-2 mm	0.635	0.406-0.992	
>2 mm	0.282	0.159-0.500	
RT			0.010
No RT	1.000		
Neoadjuvant	0.312	0.146-0.668	
Adjuvant	0.700	0.417-1.175	

The sHR of age corresponds to a unit increase of 10 years and the sHR of size corresponds to a unit increase of 1 cm. **Notation:** MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; MPNST, malignant peripheral nerve sheath tumour; RT, radiotherapy.* Depth: relative to the investing fascia.

curves of the lower risk groups are located very close to each other, which indicates some difficulty to discriminate between patients with low risk resulting from the small number of LRs observed in those groups.

Figure 5.5 shows the effect of RT on OS and CILR for two patients with the same risk factors (70 years old, 9 cm tumour size, deep depth, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma, resection margin of 0.1-2 mm) with and without neo- adjuvant RT. The patient without RT (red lines) has worse OS and higher CILR.

Detailed comparisons of observed and predicted probabilities for LR for data subgroups are shown in Table 5.5. No significant differences between observed and predicted outcomes were evident. The C-index for LR was 0.696 (95% CI 0.629-0.743).

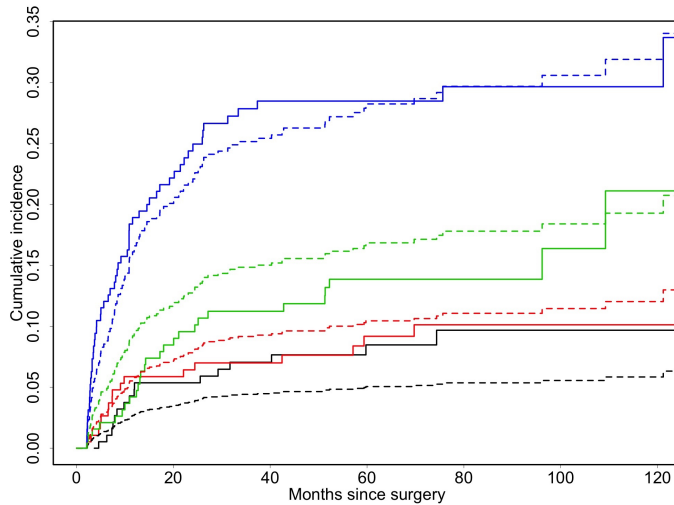


Figure 5.4: Cumulative incidence of local recurrence for four prognostic index groups. Crude cumulative incidence curves (solid lines) plotted with the model-based cumulative incidence curves (dotted lines) for four different prognostic index groups. The numbers of patients at risk were 388, 237 and 29 at 3, 5 and 10 years, respectively. Black: patients with good; red: fairly good; green: fairly poor and blue: poor prognosis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

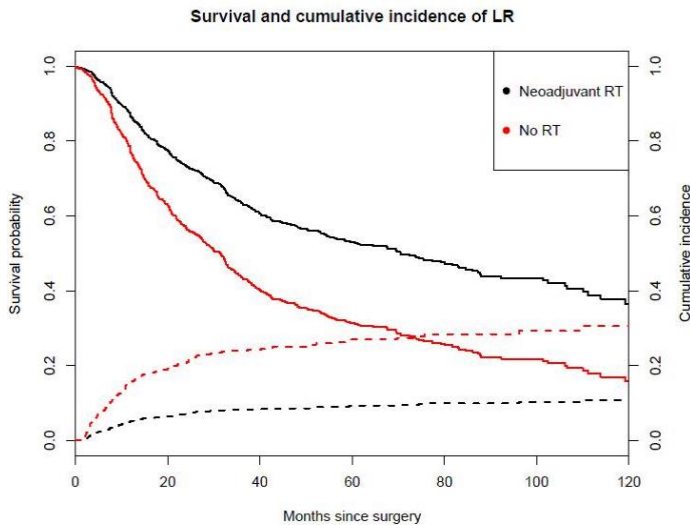


Figure 5.5: Survival and CILR for patient of 70 years, tumour size 9 cm, deep depth, MFH/UPS and resection margin 0.1-2 mm. In red: curves for patient treated with neoadjuvant RT. In black: patient without RT. Solid lines: survival curves. Dotted lines: cumulative incidence for LR. LR, local recurrence; RT, radiotherapy. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

Table 5.5: Comparing observed cumulative incidence and predicted probabilities of local recurrence, assessed by cross validation, for subgroups of data at 3, 5 and 10 years.

	n (%)	3 years			5 years			10 years		
		Pred	Obs (se)	Diff (95% CI)	Pred	Obs (se)	Diff (95% CI)	Pred	Obs (se)	Diff (95% CI)
Age										
30-60	281 (36.7)	11.4	10.2 (1.8)	1.2 (-2.3 to 4.7)	13.0	11.5 (2.0)	1.5 (-2.4 to 5.4)	14.8	12.9 (2.1)	1.9 (-2.2 to 6.0)
<30	82 (10.7)	9.6	12.6 (3.7)	-3.0 (-10.3 to 4.3)	10.8	15.7 (4.2)	-4.9 (-13.1 to 3.3)	12.4	15.7 (4.2)	-3.3 (-11.5 to 4.9)
>60	403 (52.6)	15.5	15.6 (1.8)	-0.1 (-3.6 to 3.4)	17.5	17.4 (2.0)	0.1 (-3.8 to 4.0)	19.9	20.8 (2.8)	-0.9 (-6.4 to 4.6)
Size										
<5cm	123 (16.1)	9.9	8.4 (2.6)	1.5 (-3.6 to 6.6)	11.3	11.9 (3.2)	-0.6 (-6.9 to 5.7)	12.9	18.2 (5.6)	-5.3 (-16.3 to 5.7)
5cm-10cm	295 (38.5)	11.5	10.1 (1.8)	1.4 (-2.1 to 4.9)	13.1	11.9 (2.0)	1.2 (-2.7 to 5.1)	15.0	14.1 (2.5)	0.9 (-4.0 to 5.8)
>=10cm	348 (45.4)	16.2	17.8 (2.1)	-1.6 (-5.7 to 2.5)	18.3	18.8 (2.1)	-0.5 (-4.6 to 3.6)	20.7	19.4 (2.2)	1.3 (-3.0 to 5.6)
Depth*										
Deep	579 (75.6)	13.9	13.9 (1.5)	0.0 (-2.9 to 2.9)	15.7	15.6 (1.6)	0.1 (-3.0 to 3.2)	17.8	17.9 (1.9)	-0.1 (-3.8 to 3.6)
Superficial	134 (17.5)	13.4	12.3 (2.9)	1.1 (-4.6 to 6.8)	15.2	14.8 (3.3)	0.4 (-6.1 to 6.9)	17.3	16.7 (3.7)	0.6 (-6.7 to 7.9)
Deep and superficial	53 (6.9)	8.1	9.6 (4.1)	-1.5 (-9.5 to 6.5)	9.2	9.6 (4.1)	-0.4 (-8.4 to 7.6)	10.5	9.6 (4.1)	0.9 (-7.1 to 8.9)
Histology										
Myxofibrosarcoma	238 (31.1)	12.1	11.9 (2.1)	0.2 (-3.9 to 4.3)	13.7	12.9 (2.2)	0.8 (-3.5 to 5.1)	15.7	15.7 (3.0)	0.0 (-5.9 to 5.9)
MFNST	91 (11.9)	15.6	17.7 (4.0)	-2.1 (-9.9 to 5.7)	17.6	17.7 (4.0)	-0.1 (-7.9 to 7.7)	20.0	19.6 (4.3)	0.4 (-8.0 to 8.8)
Synovial sarcoma	142 (18.5)	7.2	4.5 (1.8)	2.7 (-0.8 to 6.2)	8.2	9.1 (2.6)	-0.9 (-6.0 to 4.2)	9.4	10.4 (2.9)	-1.0 (-6.7 to 4.7)
Spindle cell sarcoma	167 (21.8)	16.8	16.7 (2.9)	0.1 (-5.6 to 5.8)	19.0	19.0 (3.3)	0.0 (-6.5 to 6.5)	21.6	26.3 (7.3)	-4.7 (-19.0 to 9.6)
MFH/UPS	77 (10.1)	18.1	19.3 (4.6)	-1.2 (-10.2 to 7.8)	20.3	20.9 (4.8)	-0.6 (-10.0 to 8.8)	23.0	20.9 (4.8)	2.1 (-7.3 to 11.5)
Margin										
0 mm	140 (18.3)	23.9	26.2 (3.8)	-2.3 (-9.7 to 5.1)	26.9	26.2 (3.8)	0.7 (-6.7 to 8.1)	30.3	26.2 (3.8)	4.1 (-3.3 to 11.5)
0.1-2 mm	343 (44.8)	14.5	13.4 (1.9)	1.1 (-2.6 to 4.8)	16.4	15.9 (2.0)	0.5 (-3.4 to 4.4)	18.7	19.3 (2.6)	-0.6 (-5.7 to 4.5)
> 2 mm	283 (36.9)	6.8	6.7 (1.5)	0.1 (-2.8 to 3.0)	7.8	8.3 (1.8)	-0.5 (-4.0 to 3.0)	9.0	9.1 (1.9)	-0.1 (-3.8 to 3.6)
RT										
No RT	182 (23.8)	16.4	15.3 (2.7)	1.1 (-4.2 to 6.4)	18.5	18.6 (3.1)	-0.1 (-6.2 to 6.0)	20.9	19.9 (3.3)	1.0 (-5.5 to 7.5)
Neoadjuvant	184 (24)	6.0	7.3 (2.0)	-1.3 (-5.2 to 2.6)	6.9	7.3 (2.0)	-0.4 (-4.3 to 3.5)	7.9	7.3 (2.0)	0.6 (-3.3 to 4.5)
Adjuvant	400 (52.2)	15.4	15.1 (1.8)	0.3 (-3.2 to 3.8)	17.4	17.0 (1.9)	0.4 (-3.3 to 4.1)	19.9	21.0 (2.8)	-1.1 (-6.6 to 4.4)

Notation: CI, confidence interval; MFH/UPS, malignant fibrous histiocytoma; undifferentiated pleomorphic sarcoma; MPNST, malignant peripheral nerve sheath tumour; RT, radiotherapy; *Depth: relative to the investing fascia.

§5.4 Discussion

In cancer care, patient characteristics are generally set at presentation, whereas the combination and timing of treatment(s) is a clinical decision based on each patient's specific circumstances. Previously, we developed a multistate model to investigate how these variables affect patient outcomes [21]. In this study, we developed the PERSARC model which uniquely presents clinicians with the possibility to accurately predict outcome of OS and CILR and compare different treatment modalities, for patients with high-grade ESTS that undergo surgical resection with curative intent. It clearly shows the possible added value of (neo) adjuvant RT at an individual patient level (Figure 5.5). Surgical margins, adjuvant therapies and individual baseline characteristics are all incorporated in this model. To assess the predictive value of this model, internal validation was performed.

This prognostic model illustrates that as the tumour size increases, the prognosis worsens for LR and OS with sHR equal to 1.031 (95% CI: 1.001-1.063) and HR equal to 1.068 (95% CI: 1.052-1.085), respectively. These findings are similar to results reported by other groups. As expected, age was an adverse prognostic risk factor for OS[107], which can only be partially explained by comorbidities. Margins are clearly associated with LR and seem to have a marginally significant effect on OS (Tables 5.2 and 5.4). The effect of recurrence on OS might be attributed to biological aggressiveness of the tumour rather than margins itself (Tables 5.2 and 5.4) [164, 75].

Patients who received RT seem to have better outcomes than those who did not (Tables 5.2 and 5.4) [112]. These patients may have been selected out of the total group of ESTS patients based on clinical characteristics, presenting scenarios or capability to undergo neoadjuvant RT [111]. All patients included in this study were treated at one of the five high-volume sarcoma centres following discussion of their case at a multidisciplinary tumour board. Although selection bias may be present, it only reflects every day care decisions. There are two prospective randomised trials on this topic; in both studies, adjuvant RT has shown a decrease in LR but had no significant impact on survival. However, both studies also included patients with low-grade tumours. Furthermore, due to low number of events (death) per arm, they could only detect a minimal benefit of 20% (as mentioned in the trial that had the most patients per arm) [33, 28]. Previous studies along with the results from this investigation suggest that neoadjuvant RT should be considered at multidisciplinary tumour board discussions for all patients undergoing surgery for primary high-grade ESTS [112, 10, 108, 127, 170]. Patients treated with neoadjuvant radiation are at significantly increased risk of wound healing complications, whether they receive conventional treatment or intensity-modulated RT [112]. Therefore, certain patients such as the elderly, those with significant medical comorbidities or those with prosthetic implants adjacent to the location of the sarcoma, may be considered inappropriate candidates for neoadjuvant radiation.

The outcomes presented above must be interpreted with caution because this model is based on clinical routine data and is therefore, susceptible to selection bias. In addition, margin categories are based on millimetres, and histology was not re-evaluated centrally. Therefore, margin assessment and evaluation of specific margins

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

‘close’ to anatomic structures; e.g. periosteum, perineurium or fascia may be subjective to variability [111]. For patients treated in centres where other margin criteria (e.g. Enneking) are in place, this model may be less applicable. Further research should focus on evaluating the different classification methods and agreeing on one standardised margin description for patients with ESTS [55, 90, 81].

While some patients may accept the increased risk of an LR and potential need for subsequent treatment by opting for less aggressive therapy including minimal margins, others may want to minimise the risk of another surgery, for example because of age and comorbidities or because of the potential effect on survival. These trade-offs are delicate and have to be based on clinical experience and substantial evidence. The prediction model developed in this study provides some indication about the possible evolution of the disease and helps in shared decision-making. The Personalised Sarcoma Care model is freely available in the Appstore and Google apps.

Appendix

§5.A Competing risks model

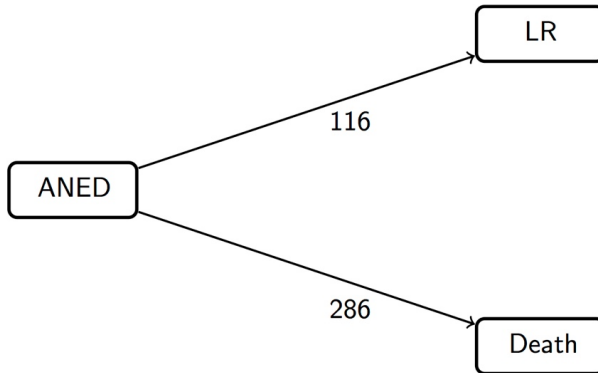


Figure 5.A.1: Competing risk model. A patient enters the state of being alive with no evidence of disease (ANED) after surgery and may move to the state of local recurrence (LR) or death.

5. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC)

Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma

This chapter has been published in *Surgical Oncology* 27 (2018) 695–701 as A.J. Rueten-Budde, et al., "Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma" [19].

Abstract

Purpose: There is increasing interest in personalized prediction of disease progression for soft tissue sarcoma patients. Currently, available prediction models are limited to predictions from time of surgery or diagnosis. This study updates predictions of overall survival at different times during follow-up by using the concept of dynamic prediction.

Patients and methods: Information from 2232 patients with high-grade extremity soft tissue sarcoma, who underwent surgery at 14 specialized sarcoma centers, was used to develop a dynamic prediction model. The model provides updated 5-year survival probabilities from different prediction time points during follow-up. Baseline covariates as well as time-dependent covariates, such as status of local recurrence and distant metastases, were included in the model. In addition, the effect of covariates over time was investigated and modelled accordingly in the prediction model.

Results: Surgical margin and tumor histology show a significant time-varying effect on overall survival. The effect of margin is strongest shortly after surgery and diminishes slightly over time. Development of local recurrence and distant metastases during follow-up have a strong effect on overall survival and updated predictions must account for their occurrence.

Conclusion: The presence of time-varying effects, as well as the effect of local recurrence and distant metastases on survival, suggest the importance of updating predictions during follow-up. This newly developed dynamic prediction model which updates survival probabilities over time can be used to make better individualized treatment decisions based on a dynamic assessment of a patient's prognosis.

§6.1 Introduction

High-grade soft tissue sarcomas (STS) are highly aggressive tumors with poor prognosis [117, 169]. Soft tissue sarcomas of the extremities account for approximately 60% of all STS diagnoses [41]. The effect of prognostic factors measured at the time of surgery (e. g. age, surgical margin, radiotherapy, tumor size, depth, and histology subtype) on overall survival has been previously investigated [117, 169, 41, 20, 37, 21] and is used in the form of prediction tools such as nomograms and online applications to make patient-specific predictions of disease progression [20, 37]. The continuous prediction of OS during treatment and follow-up has proven its clinical benefit in shared decision making and choosing the optimal individualized treatment strategy in several carcinoma cohorts [1, 6, 7].

A weakness of current sarcoma models is that they are designed for use at baseline, such as at the time of diagnosis or surgery, and cannot be used to make adequate predictions at later time points during follow-up. After surgery approximately 10% of high grade STS patients develop local recurrence (LR) with or without synchronous distant metastases (DM). Both will have a significant impact on future disease progression and the difference in prognosis should be incorporated in future treatment protocols. Even the fact that a patient survived a length of time after surgery might give an indication about his future prognosis. In addition, the effect of prognostic factors may change over time (time-varying effect), which has not yet been studied. For example, surgical margin and radiotherapy might have a strong impact on survival in the immediate time after surgery; however, their effect may change during follow-up. The use of time-dependent covariates, such as LR and DM status, and time-varying effects to update survival probabilities during follow-up is known as dynamic prediction [151]. To the best of our knowledge, no previous prediction model has been published taking the time-varying effect of risk factors into account for patients with STS. This study fills a gap in current research by investigation the effect of risk factors for death in high-grade extremity STS patients over time.

The aim of this study was to develop a dynamic prediction model for high-grade (FNCLCC grade II and III [145]) extremity STS patients that updates overall survival probabilities during follow-up. The effect of prognostic factors over time was studied and modelled accordingly in the dynamic model. The model predicts a patient's probability of surviving an additional five years from different prediction time points (t_p) after resection of their sarcoma. Specific patient examples are used to illustrate how predicted probabilities change at different prediction time points during follow-up. To implement these findings in clinical practice, this model will be made available in the updated PERSARC app and online [20].

§6.2 Methods

§6.2.1 Study design

In this study a dynamic prediction model, using a retrospectively collected cohort of patients with STS of the extremities, was developed and internally validated. Clinical data were collected between January 1st, 2000 and December 31st, 2014, at 14 different international specialized sarcoma centers, thereby creating the largest multinational dataset of high-grade surgically treated extremity STS patients in the world. Included centers are Leiden University Medical Center (Leiden, the Netherlands), Royal Orthopaedic Hospital (Birmingham and Stanmore, UK), Netherlands Cancer Institute (Amsterdam, the Netherlands), Mount Sinai Hospital (Toronto, Canada), the Norwegian Radium Hospital (Oslo, Norway), Aarhus University Hospital (Aarhus, Denmark), Skåne University Hospital (Lund, Sweden), and Medical University Graz (Graz, Austria). The outcome measure used was overall survival, which was defined as time from surgery to death from any cause or last recorded follow-up. The prediction model estimates the dynamic probability of surviving an additional five years from a prediction time point tp called dynamic overall survival (DOS). From time of surgery predictions of 5-year DOS can be estimated based on updated patient information.

§6.2.2 Patients and variables

Ethical approval for this study was waived by the institutional review board, because clinical data was collected from medical records. Patients were selected from each hospital's own sarcoma registry based on histological diagnosis. Eligible diagnoses included high-grade (FNCLCC grade II and III [145]) angiosarcoma, malignant peripheral nerve sheath tumor (MPNST), synovial sarcoma, spindle cell sarcoma, myxofibrosarcoma, liposarcoma, leiomyosarcoma, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma (MFH/UPS), (pleomorphic) soft tissue sarcomas not-otherwise-specified (NOS), malignant rhabdoid tumor, alveolar soft part sarcoma, epithelioid sarcoma, clear cell sarcoma, rhabdomyosarcoma (adult form) and conventional fibrosarcoma. Patients were excluded if they were initially treated without curative intent, presented with LR or DM, had Kaposi's or rhabdomyosarcoma (pediatric form), had a tumor in their abdomen, thorax, head or neck, or received isolated limb perfusion as (neo-)adjuvant treatment. For follow-up all collaborating sarcoma centers adhered to the guidelines of the European Society for Medical Oncology [56].

In the following, baseline and time-dependent variables that were included into the dynamic model are defined. Predictors measured at baseline were: age (years), tumor size by the largest diameter measured at pathological examination (centimeters), tumor depth in relation to investing fascia (deep/superficial), and histological subtype according to WHO classification [61]. Radiotherapy (yes/no) was further specified as being either neoadjuvant or adjuvant treatment. Chemotherapy was not included in the model because it was seldom given to patients for primary tumors. Surgical

margins were categorized according to the categorical R-system: 'R0' for a negative margin and 'R1-2' for a positive margin with tumor cells in the inked surface of the resection margin [76]. The potential effect modifier grade was not included, since all included patients had high-grade tumors. Local recurrence was defined as the presence of pathologically and/or radiologically confirmed tumor at the site where it was originally detected, more than two months after primary surgery. Distant metastases were defined as radiological evidence of systemic spread of tumor distant from the primary tumor site.

Initially 2427 patients were considered, however, those who underwent surgery before January 1st, 2000 ($n = 187$) and those with missing outcome information ($n = 8$) were excluded leaving a total of 2232 patients for analysis.

§6.2.3 Statistical analysis

To estimate a prediction model for 5-year DOS a proportional landmark supermodel was used [151, 149]. A landmark model is able to make predictions from a particular landmark time t_{LM} , by using all (updated) information of patients in follow-up at that time. A landmark supermodel combines several landmark models corresponding to distinct landmark time points to make predictions at different prediction times t_p during follow-up.

To fit such a model, landmark time points t_{LM} were chosen every three months between zero and five years after surgery. At each of these time points a Cox proportional hazards model was estimated on the subset of patients still at risk: patients alive and in follow-up at time t_{LM} . The status of LR and DM is determined at landmark time point t_{LM} for each patient and considered fixed. These Cox models were then combined into a landmark supermodel.

The main covariates as well as the linear and quadratic effect of time in form of the term t_{LM} and t_{LM}^2 were included into the model. Some histology subtypes were not sufficiently represented in the data ($n \leq 35$) and it was not possible to estimate a separate effect for them on survival. For this reason, they were grouped together under the label "Other".

A backward selection procedure was used to select further time-varying covariates. The time-varying effect of a covariate is modelled by the interaction term between the covariate and time. Initially all interactions of covariates with t_{LM} and t_{LM}^2 were included in the model, after which interactions with t_{LM}^2 without significant effect were removed. In the next step, interactions for these prognostic factors with t_{LM} were considered and removed from the model if they had no significant effect. A p-value of ≤ 0.05 was considered significant. The validity of the prediction model was assessed in terms of model calibration, which refers to how well predicted probabilities agree with observed probabilities. The model was internally calibrated using the heuristic shrinkage factor [154]. Shrinkage of a linear prognostic index towards the mean can improve the predictions of a prognostic model [149]. The estimated shrinkage factor on new data is an estimate of necessary calibration needed to improve the model fit on new data. Without an external data set the shrinkage factor can be determined using a heuristic formula and may take values between zero and one, where values

close to one represent a good calibration.

Model discrimination refers to the ability of the model to predict higher risks for patients with an early event compared to those with later or no event and was assessed using the dynamic cross-validated C-index [149]. A C-index equal to one means that the model has perfect discrimination and a C-index of 0.5 means that the model predicts just as well as flipping a coin [9].

Most statistical methods are not able to include observations with missing values, which leads to the removal of patients with missing information. To make optimal use of the collected data multiple imputation was applied. The R-package Amelia II was used to impute five complete data sets with plausible values [82]. Across these data sets observed values stay the same, however missing values were inserted with a distribution that reflects the uncertainty surrounding the missing data. Statistical methods were applied to each individual complete data set and the results were then combined following Rubin's rule [126]. The analysis was adjusted for country effect by including country as a fixed covariate into the model. The items on both the checklist of STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) and the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) we considered during model development [158, 42]. All statistical analyses were performed in the R-software environment [122].

§6.3 Results

The number of patients used for this analysis was 2232, with a median follow-up of 6.42 years (95% confidence interval: 6.17-6.72), assessed with the reverse Kaplan-Meier method [133]. Table 6.1 provides a summary of the patient characteristics.

An overview of the number of patients used at each landmark time point is given in Figure 6.1 together with information about their disease status at that time. In total 1034 patients died, 143 patients developed LR, 556 DM, and 159 developed both.

Table 6.2 shows hazard ratios (HR) together with 95% confidence intervals (95%CI) for the risk factors included in the Cox proportional hazard model. Hazard ratios for covariates with time-constant and time-varying effects are shown in the upper and lower part of the table respectively.

Table 6.1: Patient demographics.

Characteristics	Overall
Total	2232
Age, mean (SD), years	60.86 (18.74)
Gender (%)	
Male	1203 (53.9)
Female	1029 (46.1)
Tumour size in cm mean (SD)	8.95 (5.85)
Tumor depth* (%)	
Deep	1269 (56.9)
Superficial	551 (24.7)
Unknown	412 (18.5)
Histology (%)	
Myxofibrosarcoma	432 (19.4)
MPNST	167 (7.5)
Synovial sarcoma	277 (12.4)
Sarcoma – NOS	108 (4.8)
Spindle cell sarcoma	492 (22.0)
MFH/UPS	604 (27.1)
Other	152 (6.8)
Margin (%)	
R1-2	274 (12.3)
R0	1890 (84.7)
Unknown	68 (3.0)
Radiotherapy (%)	
No radiotherapy	916 (41.0)
Neoadjuvant	265 (11.9)
Adjuvant	1004 (45.0)
Unknown	47 (2.1)
Chemotherapy (%)	
No chemotherapy	1876 (84.1)
Neoadjuvant	98 (4.4)
Adjuvant	228 (10.2)
Unknown	30 (1.3)

Notation: N, number of patients; sd, standard deviation; cm, centimeters; MPNST, malignant peripheral nerve sheath tumor; sarcoma – NOS, (pleomorphic) soft tissue sarcomas not-otherwise-specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; Histology "Other", angiosarcoma, leiomyosarcoma, liposarcoma, malignant rhabdoid tumor, alveolar soft part sarcoma, epithelioid sarcoma, clear cell sarcoma, rhabdomyosarcoma (adult form) and conventional fibrosarcoma. *Depth: relative to the investing fascia.

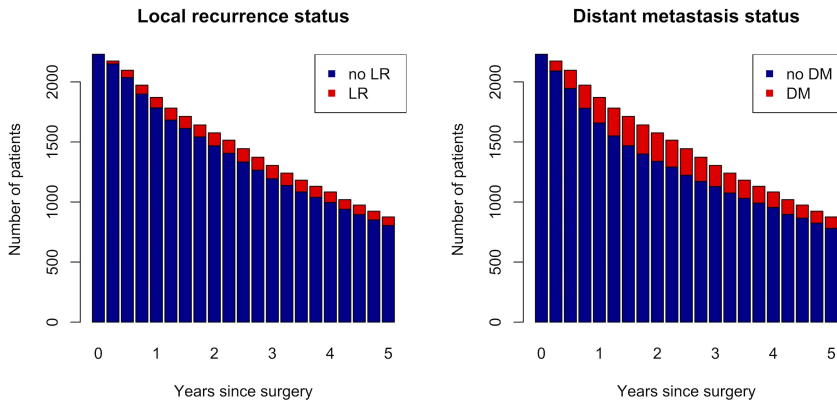


Figure 6.1: Number of patients at risk at each landmark time point t_{LM} . A) Red, patients with local recurrence; blue, patients without local recurrence. B) Red, patients with distant metastases; blue, patients without distant metastases.

Table 6.2: Dynamic prediction model for overall survival: hazard ratio (HR) along with 95% confidence interval ($n = 2232$).

	HR	95% CI	P-value
Covariates with time-constant effects			
Age (ref: 60 years, per 10 years)			
Age	1.444	1.381 - 1.510	<0.001
Age ²	1.065	1.048 - 1.082	<0.001
Tumor size (ref: 0 cm, per 1 cm)			
Size	1.120	1.072 - 1.169	<0.001
Size ²	0.997	0.996 - 0.999	0.002
Tumor depth (superficial vs. deep)	0.784	0.654 - 0.940	0.020
Radiotherapy (RT)			
No RT	1		
Neoadjuvant	0.773	0.572 - 1.044	0.095
Adjuvant	0.903	0.763 - 1.068	0.238
Local recurrence (yes vs. no)	1.998	1.622 - 2.461	<0.001
Distant metastasis (yes vs. no)	7.572	6.501 - 8.818	<0.001
Covariates with time-varying effects			
Prediction time (ref: time of surgery, per year)			
t_p	0.431	0.330 - 0.562	<0.001
t_p^2	1.127	1.066 - 1.192	<0.001
Histology			
Constant			
Myxofibrosarcoma	1		

6. Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma

Table 6.2: (continued)

	HR	95% CI	P-value
MPNST	1.807	1.270 - 2.571	0.001
Synovial sarcoma	1.323	0.971 - 1.801	0.076
Sarcoma – NOS	1.181	0.784 - 1.781	0.426
Spindle cell sarcoma	0.819	0.638 - 1.051	0.117
MFH/UPS	1.000	0.789 - 1.269	0.974
Other	1.229	0.828 - 1.825	0.307
Linear time-varying effect			
Myxofibrosarcoma	1		
MPNST	0.916	0.692 - 1.212	0.539
Synovial sarcoma	1.368	1.084 - 1.727	0.008
Sarcoma – NOS	1.067	0.739 - 1.540	0.730
Spindle cell sarcoma	1.184	0.959 - 1.461	0.116
MFH/UPS	1.256	1.024 - 1.540	0.029
Other	1.050	0.742 - 1.486	0.781
Quadratic time-varying effect			
Myxofibrosarcoma	1		
MPNST	0.985	0.930 - 1.044	0.618
Synovial sarcoma	0.913	0.864 - 0.964	0.001
Sarcoma – NOS	0.983	0.913 - 1.058	0.648
Spindle cell sarcoma	0.990	0.947 - 1.035	0.660
MFH/UPS	0.968	0.928 - 1.010	0.137
Other	0.985	0.913 - 1.062	0.689
Margin			
Constant			
R0 vs. R1-2	0.764	0.606 - 0.964	0.024
Linear time-varying effect			
R0 vs. R1-2	1.417	1.127 - 1.783	0.003
Quadratic time-varying effect			
R0 vs. R1-2	0.947	0.902 - 0.993	0.026

Notation: HR, hazard ratio; CI, confidence interval; t_p , prediction time points; MPNST, malignant peripheral nerve sheath tumor; sarcoma – NOS, (pleomorphic) soft tissue sarcomas not-otherwise-specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma. Depth: relative to the investing fascia.

Age, tumor size, and depth show a significant time-constant effect on 5-year DOS. Age and tumor size are modelled by both a linear and quadratic term (age in steps of 10 years, size in cm), due to significant nonlinearity. The HR corresponding to a particular age and size consists of two components: their linear effect HR_{lin} and their quadratic effect HR_{quad} . For the risk factor age the HR of a 70-year-old patient compared to a 60-year-old patient (reference) is equal to

$$HR_{lin}^{step} \times HR_{quad}^{step^2} = 1.444 \times 1.065 = 1.538$$

where ‘step’ in the computation represents the age difference between the two patients, and one step corresponds to a 10-year increase.

The HR of an 80-year-old patient (20-year increase, corresponding to a step of 2) compared to a 60-year-old one is equal to $1.4442 \times 1.0654 = 2.682$.

Both LR and DM show a significant time-constant effect with HR equal to 1.998 (95%CI: 1.622-2.461) and 7.572 (95%CI: 6.501-8.818) respectively. The occurrence of LR significantly decreases the 5-year DOS predictions (Figure 6.2). Figure 6.2 shows the probability of dying within five years for patients with different baseline characteristics and states of disease progression, from different prediction time points t_p . In Figure 6.2A the probability of dying within five years is displayed for two 61-year old patients with 9 cm deep myxofibrosarcoma, R0 margin, no radiotherapy treatment and no DM. The blue and red lines represent the probability of dying within five years for patients with the previous characteristics in the absence and presence of LR at prediction time point t_p respectively. If still alive at one year after surgery, the probability of dying within five years is 30% and 52% for the patient without and with LR respectively. Figure 6.2B shows that patients with the same risk factors as individuals in Figure 6.2A who developed DM before the prediction time point t_p have a much higher dynamic prediction of death within five years. Figure 6.2C and D illustrate a very different prediction pattern for a patient with other characteristics.

Surgical margin and histology subtype show a significant time-varying effect. To explain how the time component is incorporated in the model and affects a patient’s risk, the HR at one year after surgery for a patient with an R0 margin compared to a patient with an R1-2 margin is calculated by using the following formula

$$\begin{aligned} \text{HR} &= [\text{constant} \times (\text{linear time-varying effect})^{t_p} \times (\text{quadratic time-varying effect})^{t_p^2}] \\ &= 0.764 \times 1.417 \times 0.947 = 1.025 \end{aligned}$$

where $t_p = 1$ and $t_p^2 = 1$ (Table 6.3).

Table 6.3: Values of HR for 5-year dynamic overall survival for a patient operated with an R0 margin at different prediction time points t_p (reference: R1-2).

t_p	constant	linear time-varying effect	quadratic time-varying effect	HR	95% CI	P-value
0	0.76	1.417 ⁰	0.947 ⁰	0.764	0.606 - 0.964	0.024
1	0.76	1.417 ¹	0.947 ¹	1.025	0.828 - 1.269	0.821
2	0.76	1.417 ²	0.947 ⁴	1.234	0.943 - 1.614	0.128
3	0.76	1.417 ³	0.947 ⁹	1.332	0.965 - 1.838	0.085
4	0.76	1.417 ⁴	0.947 ¹⁶	1.289	0.859 - 1.934	0.232
5	0.76	1.417 ⁵	0.947 ²⁵	1.119	0.628 - 1.992	0.730

t_p prediction time point; HR, hazard ratio; CI, confidence interval.

The HR changes from 0.764 at time of surgery to 1.025 one year after surgery. At a prediction time point of two years after surgery, the HR further increases to 1.234. The change in HRs over time for margin is depicted in Figure 6.3. The figure shows

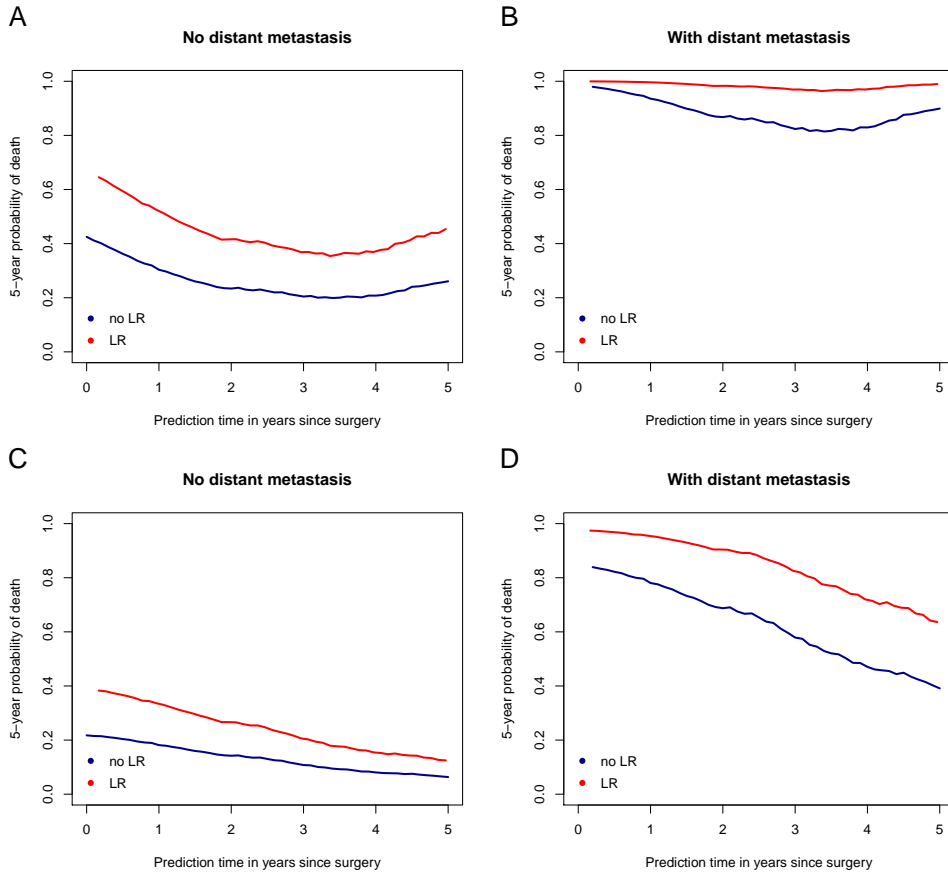


Figure 6.2: 5-year probability of death estimates for patients with different characteristics and at different states of disease progression. A and B: 61 years, tumor of 9 cm, deep myxofibrosarcoma, treated with an R0 margin and no radiotherapy. (A) Without DM at time of prediction (t_p). (B) diagnosed with DM before time of prediction (t_p). C and D: 45 years, 5 cm superficial synovial sarcoma, treated with an R0 margin, and adjuvant radiotherapy. (C) Without DM at time of prediction (t_p). (D) diagnosed with DM before time of prediction (t_p). Blue: without LR; red: with LR.

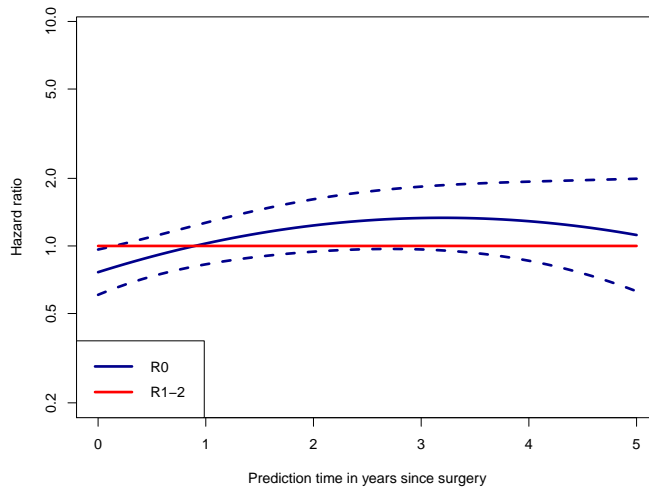


Figure 6.3: Time-varying hazard ratio for surgical margin. Blue: R0 margin; red: R1-2 margin (reference). Dashed line: pointwise confidence interval for HR of R0 margin.

that an R0 margin right after surgery appears to have a protective effect on 5-year DOS. However, the effect decreases with time.

The (time-varying) effect of histology subtype may be calculated analogously to the margin example. The interpretation of its effect however, is more difficult since all HRs are given relative to the chosen reference category myxofibrosarcoma.

Figure 6.4 displays the time-varying effect of histology subtype on two example patients. The left panels (A, C, and E) display the 5-year probability of death for a 61 year old patient with a 9 cm deep tumor, treated with no radiotherapy and R0 margin. Panel A shows the probabilities in case this specific patient had no adverse event at time of prediction. Panel C and E show the probabilities of death in case the patient had LR or DM at time of prediction respectively. Different colored lines correspond to different histology subtypes. Analogously, the left panels (B, D, and F) show probabilities for a 45 year old patient with 5 cm superficial tumor, treated with adjuvant radiotherapy and R0 margin.

Good model calibration was indicated by a heuristic shrinkage factor equal to 0.996. The discriminative ability of the model was measured with dynamic cross-validated C-indices of 0.694, 0.777, 0.813, 0.810, 0.798, and 0.781 at 0-, 1-, 2-, 3-, 4-, and 5-years after surgery respectively. The C-indices are quite high, implying a very good discriminative ability of the model. The reason for this is the strong predictive value that DM has for survival. A patient with DM will have a much worse prognosis compared to a patient without DM.

6. Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma

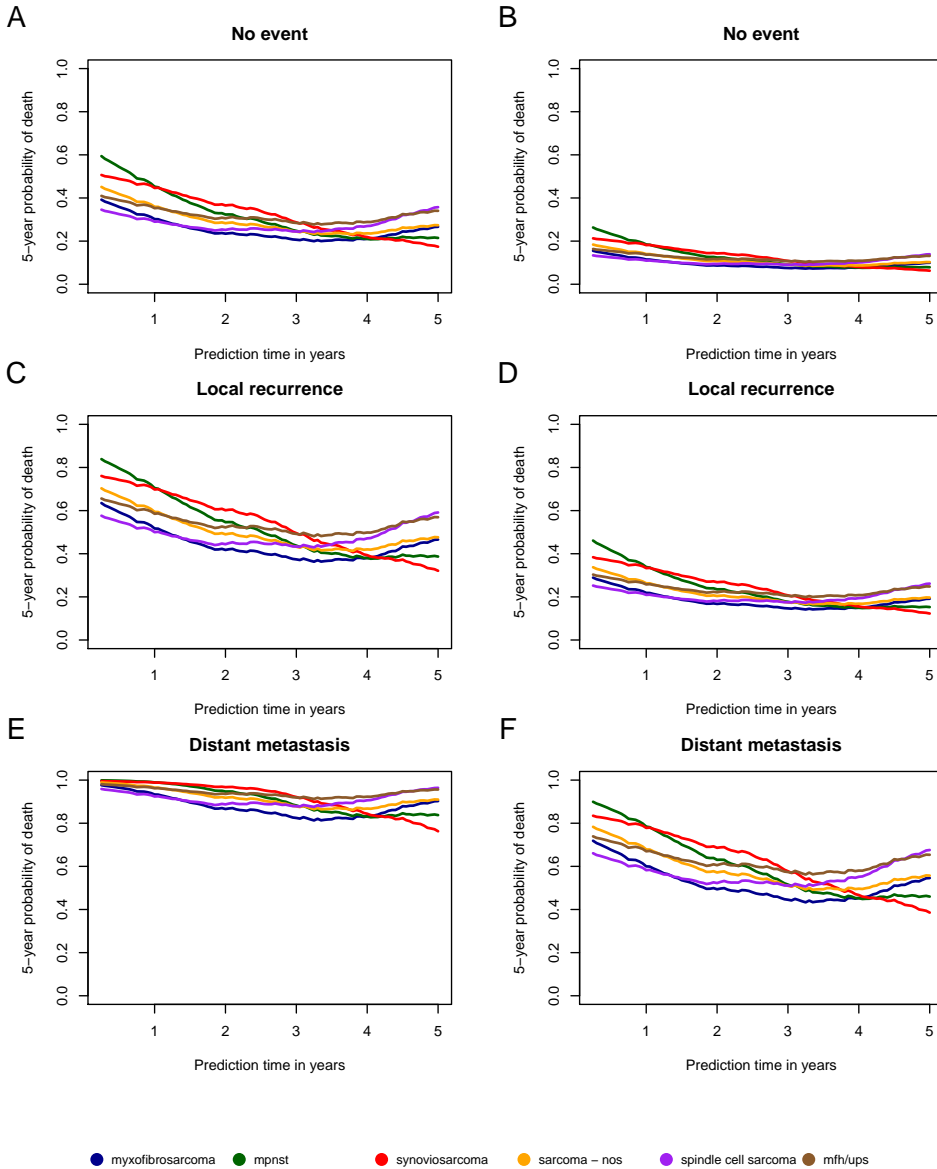


Figure 6.4: 5-year probability of death estimates for patients with different characteristics and at different states of disease progression. A, C, and E: 61 years, 9 cm deep tumor, with R0 margin and no radiotherapy. (A) Without LR or DM at time of prediction (t_p). (C) diagnosed with LR before time of prediction (t_p). (E) diagnosed with DM before time of prediction (t_p). B, D, and F: 45 years, 5 cm superficial tumor, with R0 margin and adjuvant radiotherapy. (B) Without LR or DM at time of prediction (t_p). (D) diagnosed with LR before time of prediction (t_p). (F) diagnosed with DM before time of prediction (t_p).

§6.4 Discussion

The prediction model developed in this study is able to provide estimates for the probability of surviving an additional five years from a prediction time point after surgery (t_p). It can be used from time of surgery up until five years post-surgery to make updated predictions for patients with high-grade STS of the extremities treated surgically with curative intent. This allows for optimization of evidence based shared decision-making and may improve the personalization of sarcoma treatment. Information about a patient's LR and DM status is used in the model, since those factors significantly influence a patient's prognosis. Additionally, it allows for personalization of the treatment options in progressive disease. Internal calibration using the heuristic shrinkage factor showed that the model was well calibrated and dynamic cross-validated C-indices demonstrate its ability to discriminate between high- and low-risk patients.

Additionally, this study investigated the effect of prognostic factors over time and found a significant time-varying effect for surgical margin and histology subtype on overall survival. Initially an R0 margin is associated with a better 5-year DOS compared to an R1-2 margin, however, this effect changes over time. At later time points during follow-up, no significant effect of margin on 5-year DOS could be found. This result should be interpreted with caution since the majority of patients were treated with (neo)adjuvant radiotherapy (see Table 6.1).

The strength of this research is that the data were collected from a very large number of relatively homogeneous sarcoma patients world-wide and patients were not selected (i.e. this is a 'real world' patient population). A limitation of this study is that re-evaluations of tumor histology could not be performed due to practical and logistical constraints. Additionally, when a patient has developed DM and/or is receiving care in the palliative setting, the routine checks for LR are not always performed and therefore underestimation of the total incidence of LR might be possible.

To the best of the authors' knowledge, this is the first dynamic prediction model for patients with high-grade extremity STS, which allows for prediction of 5-year DOS during follow-up. A similar model has been used to make dynamic predictions for breast cancer patients [62]. This model is an essential addition to current models, since it provides updated predictions after surgery (instead of at the time of surgery alone).

The results of this study will be made freely available through the updated Personalized SARcoma Care (PERSARC) mobile application. With the app it will be possible to make personalized dynamic predictions during follow-up, taking specific patient, tumor, and treatment characteristics into account [20].

External Validation and Adaptation of a Dynamic Prediction Model for Patients with High-Grade Extremity Soft Tissue Sarcoma

This chapter is joined work with Michiel van de Sande, Veroniek van Praag, the PERSARC studygroup, and Marta Fiocco.

Abstract

Background: A dynamic prediction model for patients with soft tissue sarcoma of the extremities has been previously developed and published to predict updated overall survival probabilities from time of surgery and throughout follow-up. This study updates and externally validates the dynamic model to allow for further implementation in clinical practice.

Methods: Data from 3826 patients with high-grade extremity soft tissue sarcoma, treated surgically with curative intent were used to update the dynamic Personalised Sarcoma Care (PERSARC) model. More patients were added to the original model development cohort and grade was included in the model. The model was externally validated with data from 1111 patients treated at a single tertiary sarcoma center.

Results: Calibration plots, to compare observed and predicted survival for the external data set show good calibration. Dynamic C-indices suggest that the model can adequately discriminate between high and low risk patients. Values for the dynamic C-indices at 0-, 1-, 2-, 3-, 4-, and 5-years after surgery were equal to 0.697, 0.790, 0.822, 0.818, 0.812, and 0.827 respectively.

Conclusion: Results from the external validation show that the dynamic PERSARC model is reliable and robust in predicting the probability of surviving an additional 5 years from a specific prediction time point during treatment and follow-up. The model combines patient characteristics, treatment-specific and time-dependent variables such as local recurrence and distant metastasis to provide reliable and accurate predictions of overall survival during follow-up and is available through the PERSARC App.

§7.1 Introduction

Extremity soft tissue sarcomas (eSTS) not only represent a wide variety of histological subtypes, sizes and grades but also affect patients of all age groups. This reflects the clear and substantial differences in their clinical course and prognosis [61]. As treatment protocols differ for specific patients between institutes and countries, several prognostic prediction models for overall survival (OS) and local recurrence have been developed [103, 35, 37, 20, 36, 114, 115]. However, these models are designed to estimate prognosis at the time of treatment or diagnosis and do not take new events that occur during treatment and follow-up into account. In addition, they do not account for possible time-varying effects of baseline risk factors.

A dynamic prediction model for patients with eSTS was therefore developed, the dynamic Personalised Sarcoma Care (PERSARC) model, to predict the probability of surviving an additional 5 years from a prediction time point during follow-up [19]. Before the introduction of the dynamic PERSARC model, prediction models for eSTS patients were limited to predictions from baseline, e.g. time of surgery or diagnosis [103, 35, 37, 20, 36, 114, 115]. The dynamic PERSARC model uses updated patient information such as occurrence of local recurrence (LR) and distant metastasis (DM) which become available during follow-up, to update predictions over time. Additionally, it accounts for the time-varying effects of histology subtype and surgical margin on survival. The dynamic model has been internally validated through the use of cross-validation, but so far, no external validation has been performed for any dynamic model in sarcoma prediction. As the original publication on dynamic PERSARC did not account for grade, the model is updated to meet current clinical demands and improve possibilities for implementation.

The aim of this study was to update and improve the existing dynamic prediction model as well as to validate it using a large external data set. The model was adapted in two ways: (1) new patients were added to the model development cohort, and (2) the grade of disease was included in the model.

§7.2 Methods

§7.2.1 Study design

In this study the original dynamic prediction model developed by Rueten-Budde et al. (2018) [19] was updated and externally validated, using a retrospectively collected cohort of patients with eSTS. The model development data was augmented for the update and contained data from Leiden University Medical Center (Leiden, the Netherlands), Royal Orthopaedic Hospital (Birmingham and Stanmore, UK), Netherlands Cancer Institute (Amsterdam, the Netherlands), Mount Sinai Hospital (Toronto, Canada), the Norwegian Radium Hospital (Oslo, Norway), Aarhus University Hospital (Aarhus, Denmark), Skåne University Hospital (Lund, Sweden), Medical University Graz (Graz, Austria), Royal Marsden Hospital (London, UK), Daniel den Hoed (Rotterdam, the Netherlands), Radboud University Medical Center (Nijmegen, the

Netherlands), University Medical Center Groningen (Groningen, the Netherlands), Haukeland University Hospital (Bergen, Norway), Helios Klinikum Berlin-Buch (Berlin, Germany), MedUni Vienna (Vienna, Austria), Vienna General Hospital (Vienna, Austria), and the EORTC trial 62931, a randomized controlled trial which studied the effect of intensive adjuvant chemotherapy on several outcome measures.

External data were provided by Istituto Nazionale dei Tumori (Milan, Italy). For both, the model development and external cohort data were collected from centers between January 1st, 2000 and December 31st, 2014. Data from the EORTC trial 62931, which is part of the development cohort, were collected between February 1995, and December 2003.

The outcome of interest was overall survival, defined as time from surgery to death due to any cause or last recorded follow-up. The dynamic model predicts 5-year dynamic overall survival (DOS) from a particular prediction time point during follow-up. For example, at one-year post-surgery the model predicts the probability of surviving an additional five years (therefore until 6 years post-surgery). To determine the predictive performance of the model, calibration and discrimination were evaluated with the external data set. Ethical approval for this study was waived by the institutional review board CME (G16.022), because clinical data was collected from medical records and were pseudo-anonymized.

§7.2.2 Patients and Variables

Selection and exclusion criteria were identical for the model development cohort and the external cohort [19]. All patients were selected from the sarcoma registry based on histological diagnosis from each hospital. Histologically, tumors were classified according to the WHO's criteria [61] and patients were grouped into eight categories. Included eSTS subtypes included high-grade (FNCLCC grade II and III [145]) angiosarcoma, malignant peripheral nerve sheath tumor (MPNST), synovial sarcoma, spindle cell sarcoma, myxofibrosarcoma, liposarcoma, leiomyosarcoma, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma (MFH/UPS), (pleomorphic) soft tissue sarcomas not-otherwise-specified (NOS), epithelioid sarcoma, clear cell sarcoma, rhabdomyosarcoma (adult form), conventional fibrosarcoma, giant cell sarcoma, malignant granular cell tumor, unclassified soft tissue sarcoma and undifferentiated sarcoma.

Patients were excluded if they were initially treated without curative intent, presented with LR or DM, had Kaposi's or rhabdomyosarcoma (pediatric form), had tumor in their abdomen, thorax, head or neck, or received isolated limb perfusion as (neo-)adjuvant treatment.

Three types of risk factors were included into the dynamic model. Patient specific predictors assessed at baseline were: age (years), tumor size by the largest diameter measured at pathological examination (centimeters), tumor depth in relation to investing fascia (deep/superficial), grade (II/III), and histological subtype according to the WHO classification [61]. Treatment related predictors measured at baseline were: radiotherapy ((neo)adjuvant/no radiotherapy), surgical margin categorized according to the categorical R-system, 'R0' for negative margin and 'R1-2' for a positive margin

with tumor cells in the inked surface of the resection margin [76]. Risk factors measured during follow-up were: local recurrence defined as the presence of pathological and/or radiologically confirmed tumor at the site where it was originally detected, more than two months after primary surgery and distant metastasis defined as radiological evidence of systemic spread of tumor distant from the primary tumor site.

The original dynamic prediction model was based on 2232 patients [19]. For the revised model additional data was collected resulting in 3826 patients for the development of the updated dynamic model. For external validation 1111 patients were considered.

§7.2.3 Statistical analysis

The dynamic prediction model developed in Rueten-Budde et al. (2018) [19] was revised by adding more patients and the variable grade to the model. The prediction model was based on landmark methodology. Technical details about landmark models for dynamic prediction are provided in van Houwelingen and Putter (2012) [149]. Additionally, the association between chemotherapy and survival was investigated.

The predictive ability of the updated model was assessed in terms of calibration and discrimination using an external data set. Model discrimination refers to how well the model is able to discriminate between high and low risk patients; dynamic C-indices [149] were computed to evaluate the performance of the model. A C-index equal to one corresponds to perfect discrimination and a C-index of 0.5 means that the model predicts just as well as flipping a coin [9]. Model calibration on the external data refers to how well predicted and observed survival probabilities have similar values and was assessed with yearly calibration plots.

Calibration plots visualize calibration at a particular prediction time point (e.g. 1 year post-surgery). Patients at risk at a specific time were divided into 8 prognostic groups based on their predicted survival. This means that the dynamic model was used to predict 5-year DOS for patients in the external data set and based on these probabilities, patients were grouped into 8 different risk groups. Five years after the prediction time point (e.g. 6 years post-surgery) the observed survival probabilities of the risk groups were estimated by applying Kaplan-Meier's method. In the calibration plot the observed survival is plotted against the predicted survival, where each point represents a risk group. If the points lay on the diagonal ($x=y$), predicted and observed survival are the same, implying that the predictions for the risk groups were perfect. The arbitrary choice for the number of risk groups was made based on the number of patients at risk over time; initially 1111 patients were at risk, however 5 years after surgery only 529 patients remain in the risk set. To have a reasonable number of patients per risk group even at 5 years post-surgery, 8 risk groups were chosen.

The items on the checklist of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) were considered during model development [42]. Statistical analyses were performed in the R-software environment [122].

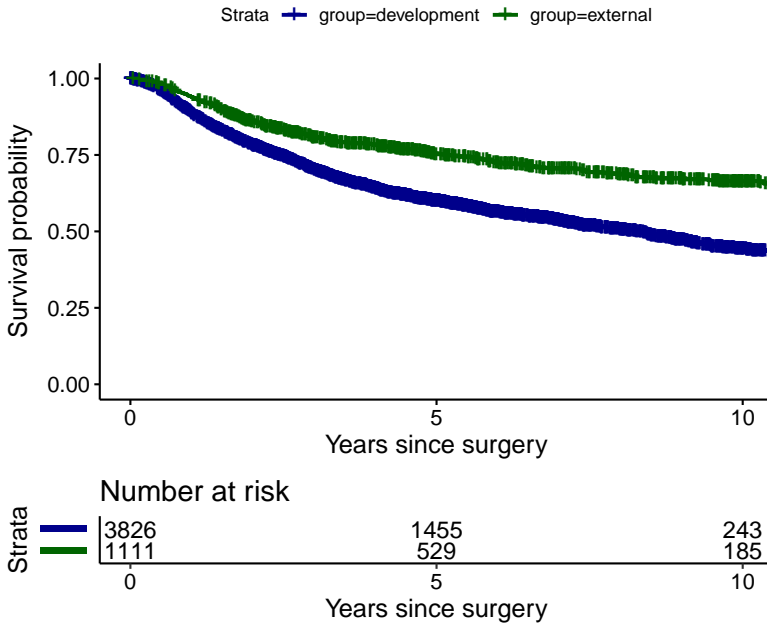


Figure 7.1: Kaplan-Meier curves for development and external cohort.

§7.3 Results

The model was developed on a cohort of 3826 patients with median follow-up equal to 6.00 years (95% confidence interval (CI): 5.86-6.18), assessed with the reverse Kaplan-Meier method [133]. The external validation cohort consisted of 1111 patients with a median follow-up equal to 6.89 years (95% CI: 6.47-7.61). Table 7.1 provides a summary of the patient characteristics for the cohort used to develop the dynamic model and the external cohort.

Figure 7.1 shows Kaplan-Meier survival curves for both development and external cohort.

An overview of the number of patients at risk in the development and external data set is given in Figure 7.2 together with information about the disease status. In the development cohort in total 1602 patients died, 241 patients developed LR, 949 DM, and 385 developed both. In the external cohort 306 patients died, 70 had LR, 279 DM and 77 developed both.

Table 7.2 shows hazard ratios (HR) together with 95% CI for the risk factors included in the revised dynamic model. Age and tumor size are both modelled by a linear and a quadratic term (age in steps of 10 years and size in steps of 1 cm). This means that the HRs consist of two components: the linear (HR_{lin}) and the quadratic effect (HR_{quad}). For example, for the risk factor age the HR of an 80-year old compared to a 60-year old patient (reference) is equal to

Table 7.1: Patient demographics for the two cohorts used to develop and to validate the model.

Characteristics	Development	External
Total	3826	1111
Age mean (sd)	59.40 (18.10)	55.46 (17.03)
Gender (%)		
Female	1680 (43.9)	504 (45.4)
Male	2011 (52.6)	607 (54.6)
Unknown	135 (3.5)	0 (0.0)
Tumor size in cm mean (sd)	9.04 (5.77)	8.33 (5.66)
Margin (%)		
R1-2	515 (13.5)	142 (12.8)
R0	3028 (79.1)	969 (87.2)
Unknown	283 (7.4)	0 (0.0)
Histology (%)		
Myxofibrosarcoma	689 (18.0)	197 (17.7)
MPNST	261 (6.8)	60 (5.4)
Synovial sarcoma	411 (10.7)	122 (11.0)
MFH/UPS and NOS	1204 (31.5)	202 (18.2)
Spindle cell	191 (5.0)	0 (0.0)
LMS	368 (9.6)	150 (13.5)
LPS	388 (10.1)	167 (15.0)
Other	314 (8.2)	213 (19.2)
Tumor depth (%)		
deep	2493 (65.2)	802 (72.2)
superficial	912 (23.8)	309 (27.8)
Unknown	421 (11.0)	0 (0.0)
Grade		
2	639 (16.7)	432 (38.9)
3	3111 (81.3)	679 (61.1)
Unknown	76 (2.0)	0 (0.0)
Radiotherapy (%)		
No radiotherapy	1331 (34.8)	474 (42.7)
Neoadjuvant	517 (13.5)	138 (12.4)
Adjuvant	1878 (49.1)	499 (44.9)
Unknown	100 (2.6)	0 (0.0)
Chemotherapy (%)		
No	3189 (83.4)	739 (66.5)
Yes	470 (12.3)	372 (33.5)
Unknown	167 (4.4)	0 (0.0)

Notation: sd, standard deviation; cm, centimeters; MPNST, malignant peripheral nerve sheath tumor; sarcoma - NOS, (pleomorphic) soft tissue sarcomas not-otherwise-specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; LMS, leiomyosarcoma; LPS, liposarcoma ; Histology ‘Other’, angiosarcoma, epithelioid sarcoma, clear cell sarcoma, rhabdomyosarcoma (adult form), giant cell sarcoma, malignant granular cell tumor, conventional fibrosarcoma, unclassified soft tissue sarcoma and undifferentiated sarcoma. Tumor depth: relative to the investing fascia.

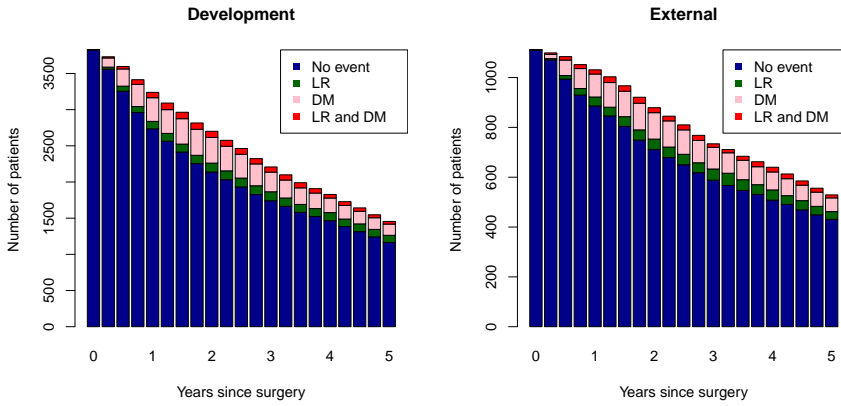


Figure 7.2: Number of patients at risk in development and external data set respectively. Red: patients with local recurrence and distant metastasis; pink: patients with distant metastasis; green: patients with local recurrence; blue: patients without local recurrence or distant metastasis.

$$\text{HR}_{in}^{step} \times \text{HR}_{quad}^{step^2} = 1.366^2 \times 1.052^4 = 2.285$$

where ‘step’ corresponds to the age difference between the two patients in units of 10 years.

Surgical margin and histology subtype are modelled as time-varying variables, which means that the effect on the outcome changes over time. For example, the HR one-year postop for a patient with R0 margin compared to a R1-2 margin is equal to

$$\begin{aligned} \text{HR} &= [\text{constant} \times (\text{linear time-varying effect})^{t_p} \times (\text{quadratic time-varying effect})^{t_p^2}] \\ &= 0.827 \times 1.334 \times 0.954 = 1.052 \end{aligned}$$

where $t_p = 1$ and $t_p^2 = 1$. The HR changes from 0.827 at time of surgery to 1.052 one year later. The model shows that the effect of surgical margin changes from being protective at surgery time to having no effect on survival after one year.

Table 7.2: Dynamic prediction model for overall survival: hazard ratio (HR) along with 95% confidence interval ($n = 3826$).

	HR	95% CI	P-value
Covariates with time-constant effects			
Age (ref: 60 years, per 10 years)			
Age	1.366	1.304 - 1.431	<0.001
Age ²	1.052	1.028 - 1.076	<0.001
Tumor size (ref: 0 cm, per 1 cm)			
Size	1.158	1.116 - 1.202	<0.001
Size ²	0.996	0.995 - 0.998	<0.001
Tumor depth (superficial vs. deep)	0.790	0.673 - 0.927	0.004
Grade (3 vs. 2)	1.425	1.174 - 1.730	<0.001
Radiotherapy (RT)			
No RT	1		
Neoadjuvant	0.719	0.583 - 0.886	0.002
Adjuvant	0.818	0.716 - 0.936	0.003
Local recurrence (yes vs. no)	2.232	1.892 - 2.634	<0.001
Distant metastasis (yes vs. no)	6.446	5.662 - 7.338	<0.001
Covariates with time-varying effects			
Prediction time (ref: time of surgery, per year)			
t_p	0.507	0.415 - 0.621	<0.001
t_p^2	1.095	1.050 - 1.141	<0.001
Histology			
Constant			
Myxofibrosarcoma	1		
MPNST	2.132	1.633 - 2.783	<0.001
Synovial sarcoma	1.458	1.145 - 1.856	0.002
MFH/UPS and NOS	1.207	1.004 - 1.452	0.045
Spindle cell	1.396	1.054 - 1.848	0.020
LMS	1.065	0.819 - 1.386	0.638
LPS	0.915	0.706 - 1.185	0.501
Other	1.419	1.095 - 1.841	0.008
Linear time-varying effect			
Myxofibrosarcoma	1		
MPNST	0.845	0.669 - 1.068	0.159
Synovial sarcoma	1.261	1.037 - 1.534	0.020
MFH/UPS and NOS	1.002	0.851 - 1.179	0.981
Spindle cell	1.058	0.824 - 1.357	0.660
LMS	1.166	0.941 - 1.444	0.160
LPS	1.010	0.812 - 1.256	0.929
Other	0.863	0.663 - 1.124	0.276
Quadratic time-varying effect			
Myxofibrosarcoma	1		

7. External Validation and Adaptation of a Dynamic Prediction Model for Patients with High-Grade Extremity Soft Tissue Sarcoma

Table 7.2: (continued)

	HR	95% CI	P-value
MPNST	1.000	0.947 - 1.056	1.000
Synovial sarcoma	0.939	0.897 - 0.983	0.007
MFH/UPS and NOS	1.009	0.976 - 1.044	0.585
Spindle cell	0.972	0.906 - 1.043	0.434
LMS	0.989	0.946 - 1.034	0.636
LPS	1.011	0.967 - 1.058	0.622
Other	1.019	0.963 - 1.078	0.510
Margin			
Constant			
R0 vs. R1-2	0.827	0.698 - 0.981	0.029
Linear time-varying effect			
R0 vs. R1-2	1.334	1.114 - 1.597	0.002
Quadratic time-varying effect			
R0 vs. R1-2	0.954	0.918 - 0.990	0.014

Notation: HR, hazard ratio; CI, confidence interval; tp, prediction time points; MPNST, malignant peripheral nerve sheet tumor; sarcoma - NOS, (pleomorphic) soft tissue sarcomas not-otherwise-specified; MFH/UPS, malignant fibrous histiocytoma/undifferentiated pleomorphic sarcoma; LMS, leiomyosarcoma; LPS, liposarcoma; Histology ‘Other’, angiosarcoma, epithelioid sarcoma, clear cell sarcoma, rhabdomyosarcoma (adult form), giant cell sarcoma, malignant granular cell tumor, conventional fibrosarcoma, unclassified soft tissue sarcoma and undifferentiated sarcoma. Tumor depth: relative to the investing fascia.

In a preliminary analysis, the association of risk factors to the outcome chemotherapy treatment (yes (neoadjuvant or adjuvant) vs. no) was evaluated. Most baseline risk factor showed a significant association (age, tumor size, depth, histology, radiotherapy, grade). Country of treatment was significantly associated to chemotherapy treatment. This means that, correcting for the other risk factors (age, tumor size, depth, margin, histology, radiotherapy, grade) in the model, countries had different approaches in giving chemotherapy treatment. The association of chemotherapy to survival was investigated by including this risk factor in the dynamic model and no significant effect was found (chemotherapy yes vs. no; HR = 1.131; 95% CI: 0.946-1.352; p value = 0.178). Chemotherapy was therefore not included in the updated dynamic prediction model.

The quality of the model can be assessed from the calibration plots (Figure 7.3A-F). Each point in the plot represents a risk group; the figure shows they are relatively close to the diagonal line implying that predictions are accurate. Figure 7.3 also suggests that the model generally slightly underestimated survival.

The discriminative ability of the model was assessed with dynamic C-indices, with values equal to 0.697, 0.790, 0.822, 0.818, 0.812, and 0.827 at 0-, 1-, 2-, 3-, 4-, and 5-years after surgery respectively. High values for the C-indices are due to the strong predictive value of DM on survival. A patient who experience DM has much worse

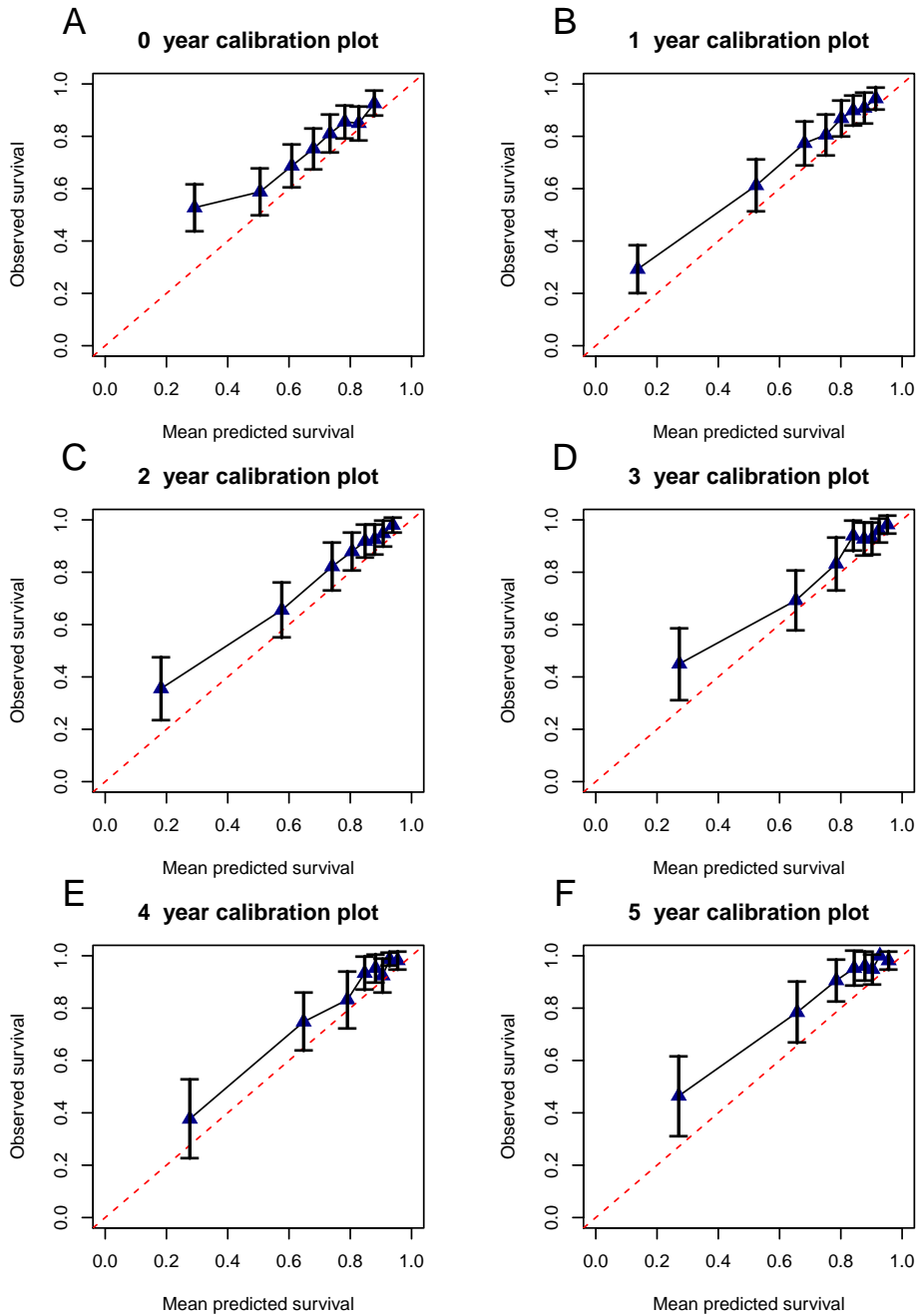


Figure 7.3: Calibration plots for predictions of 5-year DOS from 0-, 1-, 2-, 3-, 4- and 5-years post-surgery.

7. External Validation and Adaptation of a Dynamic Prediction Model for Patients with High-Grade Extremity Soft Tissue Sarcoma

prognosis compared to a patient without DM.

§7.4 Discussion

The previously developed dynamic prediction model has been updated and successfully externally validated. The sample size of the model development cohort was increased from 2232 to 3826 patients and the risk factor grade was added to the updated model [19]. The model can estimate the probability of surviving an additional 5 years from a prediction time point during follow-up. It can be used from time of surgery up until 5 years post-surgery for patients with high-grade eSTS treated with curative intent.

Even though calibration plots showed that predicted survival was close to observed survival the model generally underestimated survival in the external cohort. Kaplan-Meier curves estimated for the development and external cohort indicate that the external cohort had better survival. There are several possible reasons for the underestimation of survival: the effect of risk factors might be different in the development cohort compared to the external cohort, or patients might differ in terms of an unobserved covariate which might affect survival and cannot be taken into account.

The association of chemotherapy with survival is controversial, and its indication greatly depends on other risk factors (indication bias). When added to the dynamic model, chemotherapy showed no significant association with survival.

The updated dynamic prediction models will be implemented in the updated PERSARC application; available for free at the Apple Store and Google Play Store.

Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model

This chapter has been published in *Pediatric Blood & Cancer* 66 (2019) Article e27943 as A.J. Rueten-Budde, et al., "Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model" [14].

Abstract

Background: We investigated the effects of surgical margins, histological response, and radiotherapy on local recurrence (LR), distant metastasis (DM), and survival in Ewing sarcoma.

Procedure: Disease evolution was retrospectively studied in 982 patients with Ewing sarcoma undergoing surgery after chemotherapy using a multistate model with initial state surgery, intermediate states LR, pulmonary metastasis (DMpulg), other DM \pm LR (DMother), and final state death. Effect of risk factors was estimated using Cox proportional hazard models.

Results: The median follow-up was 7.6 years (95% CI, 7.2-8.0). Risk factors for LR are pelvic location, HR 2.04 (1.10-3.80), marginal/intralesional resection, HR 2.28 (1.25-4.16), and radiotherapy, HR 0.52 (0.28-0.95); for DMpulg the risk factors are <90% necrosis, HR 2.13 (1.13-4.00), and previous pulmonary metastasis, HR 4.90 (2.28-8.52); for DMother are 90% to 99% necrosis, HR 1.56 (1.09-2.23), <90% necrosis, HR 2.66 (1.87-3.79), previous bone/other metastasis, HR 3.08 (2.03-4.70); and risk factors for death without LR/DM are pulmonary metastasis, HR 8.08 (4.01-16.29), bone/other metastasis, HR 10.23 (4.90-21.36), and <90% necrosis, HR 6.35 (3.18-12.69). Early LR (0-24 months) negatively influences survival, HR 3.79 (1.34-10.76). Once DMpulg/DMother arise only previous bone/other metastasis remain prognostic for death, HR 1.74 (1.10-2.75).

Conclusion: Disease extent and histological response are risk factors for progression to DM or death. Tumor site and surgical margins are risk factors for LR. If disease progression occurs, previous risk factors lose their relevance. In case of isolated LR, time to recurrence is important for decision-making. Radiotherapy seems protective

8. Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model

for LR especially in pelvic/axial. Low percentages of LR in extremity tumors and associated toxicity question the need for radiotherapy in extremity Ewing sarcoma.

§8.1 Introduction

Ewing sarcoma is an aggressive primary bone tumor, predominantly affecting children and young adults.[61] At the time of diagnosis, 20% to 25% of the patients present with pulmonary (70-80%) and/or osseous (40-50%) metastases. A multimodal approach to treatment drastically improved survival of patients with localized Ewing sarcoma, with a 10-year overall survival (OS) of 55% to 65% nowadays. However, local recurrence, distant metastasis, and poor survival in patients with metastatic disease with a 5-year OS of 20% to 35% still remain of great concern.[95, 68] One of the strongest risk factors is the presence of metastasis at diagnosis[96, 52] and site of metastatic lesions; patients with extrapulmonary metastasis do significantly worse than patients with pulmonary metastasis alone.[95, 43] Other well-known risk factors are the primary tumor site[11, 30, 84] and tumor volume and/or size.[43, 25, 67, 63] Principles of treatment consist of neoadjuvant chemotherapy followed by local treatment of the primary tumor, either by surgery, radiotherapy, or both, and adjuvant chemotherapy. The histological response, assessed after surgery, is a strong additional prognostic factor for OS.[11, 25, 67] The effect of surgical margins on survival is controversial. The risk of local relapse is significantly lower after wide resection compared with marginal or intralesional resections.[113, 26] How the occurrence of a local recurrence may affect OS is not yet clearly established.[84, 22] If surgery with or without radiotherapy is superior compared with radiotherapy only in order to maximize local control alone is also under debate. Existing evidence is based on retrospective, nonrandomized trials.[134, 136] Several studies show advantage of post-operative radiotherapy (PORT) for patients with marginal or intralesional resections in terms of improved local control and event-free survival.[25, 63, 113, 134, 160] Possible associations between PORT and overall survival, and between local recurrence and OS, are not yet clearly established. The main problem in current studies on prognostic factors for Ewing sarcoma is that they are hampered by the choice of outcome variable. In general, overall survival, local recurrence-free survival, and disease-free survival are reported. Multiple analyses for these different endpoints are usually utilized; however, the relationship between those different endpoints cannot be investigated by using separate models. Multistate models can overcome these problems since the evolution of the disease and the occurrence of intermediate events, such as local recurrence and distant metastasis, which occur after surgery, are incorporated in the model, which provides useful insights into their relation with the considered endpoint, usually death.[12, 119, 21]

This study aims to investigate the effect of surgical margins, histological response, and radiotherapy, on local recurrence (LR), distant metastasis (DM), and OS in a large cohort of patients with Ewing sarcoma treated according to the EURO-E.W.I.N.G 99 protocol (EUROpean Ewing tumor Working Initiative of National Groups-Ewing Tumor Studies).

§8.2 Methods

This retrospective study was reviewed and approved by the institutional review board of the Leiden University Medical Center (Leiden, the Netherlands), and a waiver for informed consent was granted. A retrospective analysis of patients from the GPOH registry (Gesellschaft für Pädiatrische Onkologie und Hämatologie) treated in or according to the EURO-E.W.I.N.G 99 (EE99) protocol[8] was performed. All patients were treated between 1999 and 2009, and followed up until the end of 2017. All patients were treated according to the protocol with the aim to administer six cycles of VIDE (vincristine, ifosfamide, doxorubicin, etoposide) induction chemotherapy followed by local treatment of the primary tumor. The choice of local treatment, surgery with or without radiotherapy or definitive radiotherapy, was directed by specific guidelines in the protocol; however, the choice of the local multidisciplinary team prevailed. According to the EE99, protocol surgery was favored whenever feasible; only in case of an inoperable lesion that cannot be completely resected or a tumor in a critical site where complete surgery would cause unacceptable morbidity, definitive radiotherapy is indicated. Preoperative radiotherapy was indicated in case of clinical progression under chemotherapy or anticipated marginal or intralesional resectability. PORT was indicated in intralesional or marginal surgery and advised in cases with a poor histological response (<90% necrosis) regardless of surgical margins. Advised radiotherapy doses were 44.8-54.4 Gy, with a boost to a maximum of 64 Gy using a shrinking field technique. After local treatment, patients received maintenance chemotherapy. Only patients who underwent surgery (with or without radiotherapy) of the primary tumor after induction chemotherapy were eligible for inclusion in this study. A total of 982 patients, 470 study patients and 512 registry patients (who were treated according to the protocol but not randomized), were found to be eligible for inclusion in this study.

§8.2.1 Measures

The following data were extracted from the GPOH registry: age (0-10 years; 11-18 years; >18 years), gender, disease extent (localized, pulmonary metastasis only, other metastasis), tumor volume (<200 mL/≥ 200 mL), tumor location (extremity/axial nonpelvic/pelvic), PORT (yes/no), surgical margin (wide/marginal/intralesional), histological response (<90%/90-99%/100% necrosis), and follow-up data on LR, distant metastasis pulmonary (DMpulm), distant metastasis extrapulmonary with or without pulmonary metastasis (DMother). Histological response and resection margins were assessed on the surgical specimen by experienced local pathologists. Local recurrence was defined as local regional recurrence after initial complete response. Distant metastasis was defined as new metastatic disease or recurrence of metastatic disease after initial complete response.

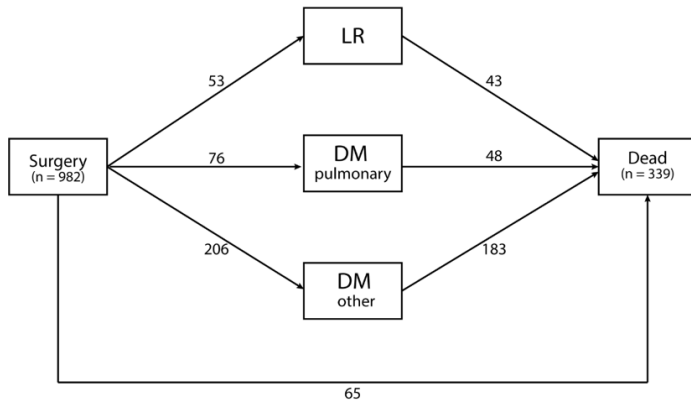


Figure 8.1: Multistate model for Ewing sarcoma

§8.2.2 Statistical analysis

OS was measured from the date of surgery, until the last day of follow-up, or date of death and evaluated using Kaplan-Meier estimates. To model disease progression, the multistate model illustrated in Figure 8.1 was estimated. The following five states are considered: alive after surgery without adverse events (state 1, surgery); alive with LR (state 2, LR); alive with pulmonary DM (state 3, DMpulm); alive with other DM (state 4, DMother); death (state 5). The effect of risk factors on each specific transition was estimated by using a Cox proportional hazard regression model; hazard ratios (HRs) along with their 95% confidence intervals (95% CI) were estimated.

§8.2.3 Missing data

For 776 (79%) of the 982 patients, information on all the covariates of interest was complete. Missing data were observed for the variable histological response (19%) and surgical margins (5%). In order to make full use of the available data, missing values were imputed using multiple imputation. Five complete data sets were generated. The multistate model was estimated on each of the imputed data sets, and the results were then combined using the Rubin rule.[126] Multiple imputation is a well-known technique used to reconstruct data when there is a small percentage of missing data. Another common approach is to drop cases with missing values and only analyze complete cases; however, this reduces the number of patients and therefore the power of the statistical tests and may even lead to biased results in some scenarios.[138]

All analyses were performed using R version 3.5.1.[122] The R-package mstate[49] was used to estimate the multistate model and to compute the occupation probabilities. The R-package Amelia II was used to impute the missing data.[82]

§8.3 Results

Table 8.1 summarizes patient and tumor characteristics and treatment for the 982 included patients, 470 study and 512 registry patients, at the time of surgery. The median follow-up, estimated with reversed Kaplan-Meier analysis, was 7.6 years (95% CI, 7.2-8.0 years). The 5-year OS was 74% (71-77%) for localized disease, 56% (47-55%) for pulmonary metastasis, and 43% (33-53%) for extrapulmonary metastasis. For patients who only had surgery as local treatment, 5-year OS was 75% (71-79%) for localized disease, 52% (39-65%) for pulmonary metastasis, and 41% (28-54%) for extrapulmonary metastasis. For patients who had surgery with radiotherapy, the 5-year OS was 74% (69-79%) for localized disease, 59% (47-71%) for pulmonary metastasis only, and 48% (31-65%) for extrapulmonary metastasis. In the group of patients treated with surgery and radiotherapy, there were more pelvic tumors (21% vs 15% in the surgery group), more marginal and intralesional surgical margins (39% vs 21% in the surgery group), and fewer patients with 100% tumor necrosis (33% vs 52% in the surgery group). The other patient and tumor characteristics were similar between both groups; see also Appendix 8.A.

In total, 53 patients of 982 (5%) developed isolated LR, 8% (14 of 169) of pelvic tumors, 8% (30 of 388) of nonpelvic axial tumors, and 2% (9 of 425) of extremity tumors. The percentage of LR was similar for patients treated with surgery and surgery with radiotherapy, 6% versus 5%, respectively. Seventy-six (7%) of the patients moved from surgery to DMpulum, and 28 (of 128) patients with isolated pulmonary metastasis at diagnosis developed new pulmonary metastasis during follow-up. The percentage of patients who developed new pulmonary metastasis was similar for patients treated with surgery and surgery with radiotherapy, 7% versus 8%, respectively. Two hundred six (21%) of the patients moved from surgery to DMother, and 39% (33 of 84) patients with previous bone/other metastasis and 21% (27 of 128) of patients with pulmonary metastasis only developed new extrapulmonary metastasis during follow-up. The percentage of patients who developed new extrapulmonary metastasis was similar for patients treated with surgery and surgery with radiotherapy, 20% versus 22%, respectively. Table 8.2 provides more details of the patient and tumor characteristics of patients who developed local recurrence, pulmonary metastasis, and other/bone metastasis with or without local recurrence with respect to the local treatment modality used. Sixty-five (7%) of the patients died without the occurrence of LR or DM. Sixty percent (39/65) of these patients had metastatic disease at diagnosis and died of progressive disease. Nine percent (6/65) died of therapy-related complications, and 15% (10/65) due to a secondary malignancy. For the remaining 10 patients, the cause of death was unknown. In total, 339 patients (35%) died.

HRs for each risk factor along with their 95% CI for each transition are reported in Table 8.3. The main prognostic factors for moving from surgery to LR are primary tumors located in the pelvis (HR 2.04; 95% CI, 1.10-3.80) and marginal or intralesional resection margins (HR 2.28; 95% CI, 1.25-4.16). The administration of radiotherapy seems protective for LR for all tumor sites combined (HR 0.52; 95% CI, 0.28-0.95). Radiotherapy was not randomized in this study, but was recommended, in the EE99 protocol, in case of intralesional or marginal resection and in case of poor histological

Table 8.1: Patient demographics and treatment characteristics after surgery for the 982 included patients

Characteristic	n (%)	Study n (%)	Registry n (%)
Total	982	470	512
Gender			
Male	590 (60)	280 (60)	310 (60)
Female	392 (40)	190 (40)	202 (40)
Age			
0-10 years	252 (26)	117 (25)	135 (26)
11-18 years	452 (46)	225 (48)	227 (44)
>18 years	278 (28)	128 (27)	150 (30)
Primary tumor localization			
Pelvic	169 (17)	75 (16)	94 (18)
Non-pelvic	813 (83)	395 (84)	418 (82)
Extremity	425 (43)	224 (48)	201 (40)
Axial	388 (40)	171 (36)	217 (42)
Volume at diagnosis			
<200 ml	577 (59)	311 (66)	266 (52)
≥200 ml	405 (41)	159 (34)	246 (48)
Disease extent at diagnosis			
Localized	770 (78)	417 (89)	353 (69)
Pulmonary metastasis	128 (13)	53 (11)	75 (15)
Extrapulmonary metastasis	84 (9)	0 (0)	84 (16)
Surgical margin			
Wide	717 (73)	352 (75)	365 (71)
Marginal	161 (16)	74 (16)	87 (17)
Intralesional	104 (11)	44 (9)	60 (12)
Histological response			
100%	426 (43)	225 (48)	202 (39)
90-99%	284 (29)	151 (32)	133 (26)
<90%	271 (28)	94 (20)	177 (35)
Post-operative radiotherapy			
No	550 (56)	284 (60)	266 (52)
Yes	432 (44)	186 (40)	246 (48)

8. Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model

Table 8.2: Patient, tumor, and treatment characteristics of patients who developed local recurrence, pulmonary metastasis, and other/bone metastasis with or without local recurrence

Characteristic	Total, n (%)	Surgery (n = 550)	Surgery + radiotherapy (n = 432)
Local recurrence	53/982 (5)	33 (6)	20 (5)
Location primary tumor			
Extremity	9/425 (2)	7	2
Non-pelvic axial	30/388 (8)	16	14
Pelvic	14/169 (8)	10	2
Surgical margin			
Wide	30/717 (4)	23	7
Marginal	12/161 (7)	4	8
Intralesional	11/104 (11)	6	5
Histological response			
100%	16/426 (4)	13	3
90-99%	19/284 (7)	11	8
<90%	18/271 (7)	9	9
Distant metastasis - pulmonary	76/982 (8)	41 (7)	35 (8)
Disease extent			
Localized	46/770 (6)	24	22
pulmonary metastasis	28/128 (22)	15	13
Bone/other metastasis	2/84 (2)	2	0
Histological response			
100%	25/426 (6)	18	7
90-99%	26/284 (9)	15	11
<90%	25/271 (9)	8	17
Distant metastasis - bone/other with or without LR	206 (21)	110 (20)	96 (22)
Disease extent			
Localized	146/770 (19)	76	70
Metastatic pulmonary	27/128 (21)	11	16
Metastatic bone/other	33/84 (39)	23	10
Histological response			
100%	65/426 (15)	37	28
90-99%	60/284 (21)	35	25
<90%	81/271 (30)	38	43

response (<90% necrosis) regardless of surgical margins. Guidelines were not always followed: 143 patients (97/266 [36%] registry and 46/284 [16%] study patients) treated with surgery alone had, based on the protocol guidelines, an indication for PORT and 190 patients (106/246 [43%] registry and 84/186 [45%] study patients) who received PORT had no indication for it based on the protocol guidelines. The main prognostic factor for patients moving from surgery to new pulmonary metastasis is a histological response of less than 90% necrosis (HR 2.13; 95% CI, 1.13- 4.00) and previous pulmonary metastasis (HR 4.90; 95% CI, 2.28-8.52). Risk factors for the transition surgery to new bone/other DM with or without LR are histological response (HR 1.56; 95% CI, 1.09-2.23 for 90-99% necrosis and HR 2.66; 95% CI, 1.87-3.79 for <90% necrosis) and previous bone/other metastasis with or without pulmonary metastasis (HR 3.08; 95% CI, 2.03-4.70). Disease extent (HR 8.08; 95% CI, 4.01-16.29 for pulmonary metastasis and HR 10.23; 95% CI, 4.90- 21.36 for bone/other metastasis) and histological response (HR 6.35; 95% CI, 3.18-12.69 for <90% necrosis) are risk factors for transition surgery to death. The administration of radiotherapy, which is not given randomly, seems to be protective (HR 0.45; 95% CI, 0.26-0.76). The effect of time to recurrence is prognostic for survival with an HR of 3.79 (95% CI, 1.34-10.76) for recurrence in the first 0 to 24 months. Histological response and disease extent are risk factors for DMpulg, but in the presence of new pulmonary disease, no statistically significant effect of histological response and disease extent on survival was observed. Histological response was also a risk factor for transition surgery to DMother, but in the presence of new metastatic disease, only previous bone/other metastasis with or without pulmonary metastasis remains of prognostic value in the presence of new metastatic disease (HR 1.74; 95% CI, 1.10-2.75).

The estimated multistate model was used to estimate outcome probabilities for specific patients Figure 8.1 visualizes the effect of local treatment modality on the patient-specific state occupation probabilities at different time points after surgery. The distance between two curves represents the probability of being in a specific state at a specific time point. After surgery, the probability of occupying the state "local treatment" decreases. The probabilities of occupying the states "local recurrence", "DMpulg", and "DMother" are similar for patients treated with surgery and surgery with radiotherapy regardless of the tumor site, surgical margins and histological response. However, radiotherapy was not randomized, so these results should be interpreted with caution.

8. Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model

Table 8.3: HRs and 95% CIs for all prognostic factors and the different transitions in the multistate model

Predictor	1: Surgery → LR HR 95% CI	2: Surgery → DMpulm HR 95% CI	3: Surgery → DM other HR 95% CI	4: Surgery → death HR 95% CI	5: LR → death HR 95% CI	6: DMpulm → death HR 95% CI	7: DMother → death HR 95% CI
Age							
0-10 years			1				1
11 - 18 years			1.55 1.05 - 2.27				0.85 0.56 - 1.31
>18 years			1.85 1.24 - 2.76				0.97 0.60 - 1.55
Tumor site							
Non-pelvic							
Pelvic	1	1	1				1
Disease extent	2.04 1.10 - 3.80	1.64 0.96 - 2.83	1.32 0.93 - 1.86				0.98 0.66 - 1.44
Localized + <200ml		1	1	1		1	1
Localized + ≥200 ml		1.47 0.82 - 2.62	1.34 0.96 - 1.86	1.87 0.91 - 3.85		1.34 0.62 - 2.88	1.06 0.73 - 1.53
Metastatic pulmonary		4.90 2.82 - 8.52	1.52 0.98 - 2.36	8.08 4.01 - 16.29		0.77 0.38 - 1.57	0.96 0.59 - 1.56
Metastatic other		0.62 0.15 - 2.62	3.08 2.03 - 4.70	10.23 4.90 - 21.36		1.38 0.16 - 11.67	1.74 1.10 - 2.75
Surgical margin							
Wide	1		1				1
Marginal / intraleisional	2.28 1.25 - 4.16		0.79 0.57 - 1.10				1.31 0.92 - 1.87
Histological response							
100%	1	1	1	1	1	1	1
90-99%	1.43 0.73 - 2.79	1.49 0.84 - 2.63	1.56 1.09 - 2.23	1.17 0.50 - 2.74	1.02 0.37 - 2.81	1.79 0.82 - 3.92	0.91 0.60 - 1.38
<90%	1.13 0.48 - 2.66	2.13 1.13 - 4.00	2.66 1.87 - 3.79	6.35 3.18 - 12.69	0.86 0.18 - 4.14	1.32 0.60 - 2.92	1.14 0.77 - 1.69
Radiotherapy							
No	1	1	1	1	1	1	1
Yes	0.52 0.28 - 0.95	0.82 0.51 - 1.31	0.99 0.73 - 1.32	0.45 0.26 - 0.76	1.53 0.65 - 3.64		1.08 0.78 - 1.49
Time to recurrence							
>24 months					1		1
<24 months					3.79 1.34 - 10.76		1.06 0.65 - 1.73

Abbreviations: DMpulm, distant metastasis solitary pulmonary; DMother, distant metastasis extrapulmonary with or without pulmonary metastasis; HR, hazard ratio; 95%CI, 95% confidence interval; LR, local recurrence

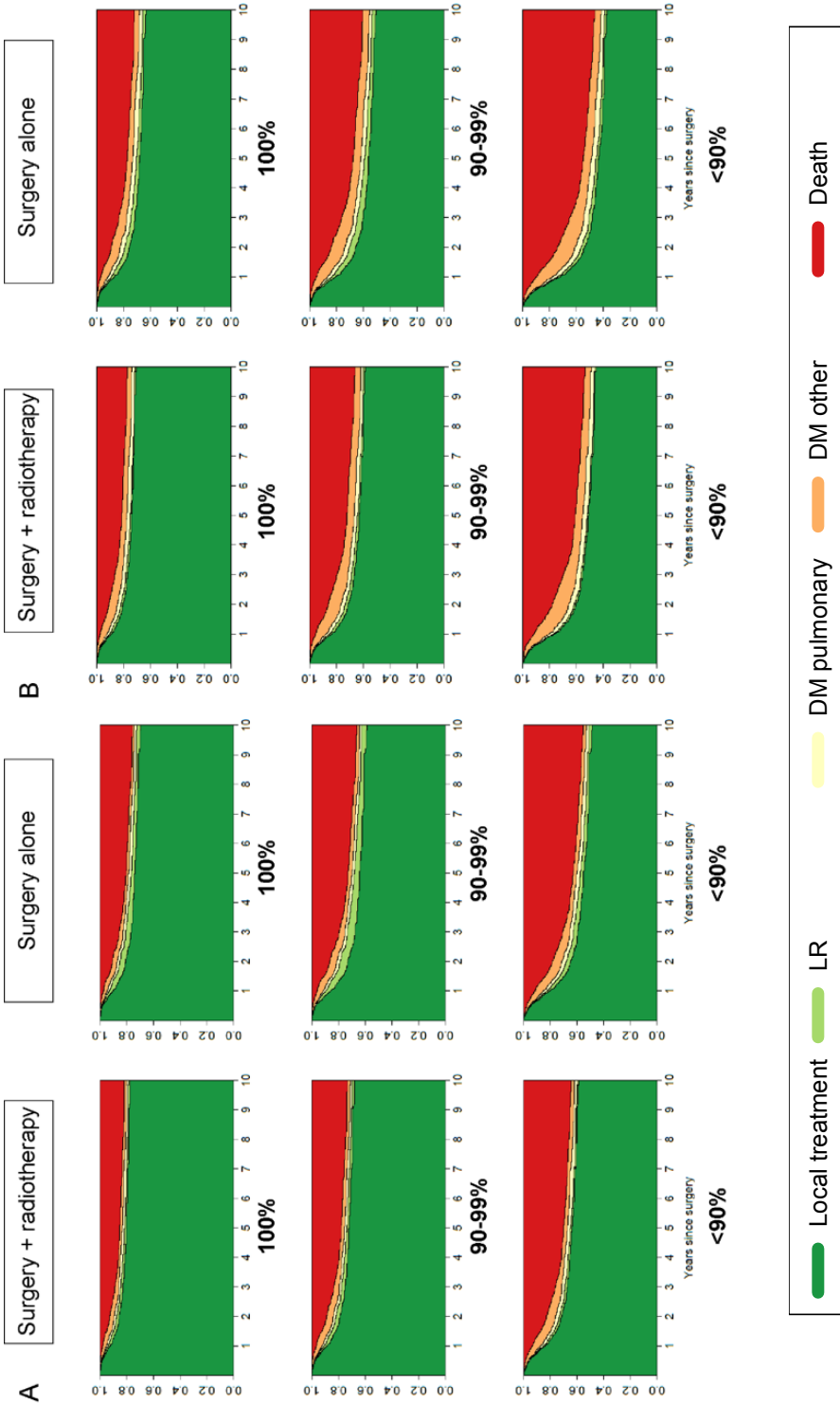


Figure 8.1: State occupation probabilities for patients with different local treatment strategies. Estimations are based on the model presented in Figure 1 and Table 2 and shows state occupation probabilities for (A) patient 1: aged 10-18 with a localized nonpelvic Ewing sarcoma and marginal or intralesional resection margins who is treated with surgery + radiotherapy (left panel) or surgery alone (right panel) for 100% necrosis, 90-99% necrosis, and <90% necrosis. (B) Patient 2: aged 10-18 with a localized pelvic Ewing sarcoma and wide resection margins who is treated with surgery + radiotherapy (left panel) or surgery alone (right panel) for 100% necrosis, 90-99% necrosis, and <90% necrosis

§8.4 Discussion

In Ewing sarcoma, local recurrence, distant metastasis, and poor survival in patients with metastatic disease remain of great concern. Associations between local treatment modality, local recurrence, distant metastasis, and death are not yet clearly established. In this study, we investigated the effect of surgical margins, histological response, and radiotherapy on the intermediate events local recurrence, distant metastasis, and on survival in a large cohort of patients with Ewing sarcoma using a multistate model.

Marginal or intralesional surgical margins are an important risk factor for transition from surgery to LR, and when a patient reaches the LR state, it was observed that the probability of death is higher in case of early LR (0-24 months), so time to recurrence could be considered as most relevant in these situations. Histological response is a strong prognostic factor for transition from surgery to distant metastasis and death. When a patient experiences new distant metastasis (either pulmonary, bone, other, or combined), histological response loses relevance as a risk factor as the occurrence of distant metastasis more dramatically affects survival. Administration of radiotherapy seems to be protective for LR. Other prognostic factors identified in this study were the primary tumor site and disease extent. A pelvic tumor site is an important risk factor for transition from surgery to LR. Previous pulmonary metastasis is a risk factor for transition to new pulmonary disease, but when a patient experiences new pulmonary disease, previous pulmonary metastasis is no longer prognostic factor for survival. Previous pulmonary or bone/other metastasis is a risk factor for transition to new bone/other metastasis with or without simultaneous LR. When reaching the DMother state only previous bone/other metastasis remains of prognostic value for survival.

The prognostic value of disease extent,[95, 96, 52, 43] histological response,[11, 25] primary tumor site,[11, 30, 84] and surgical margins[25, 113, 26, 22] observed in this study is consistent with previous studies. Several large studies show an advantage of PORT for patients with marginal or intralesional resections.[25, 63, 113, 134, 160] In addition to previous studies, this study has extended the knowledge about the effect of prognostic factors for intermediate events and final event death in Ewing sarcoma. We showed that prognostic factors have different effects on different transitions and that the impact on the next state in the evolution of the disease depends on the state a patient occupies. Apart from the patient's history, the time element is also of paramount importance for decision-making. LR within 2 years or the occurrence of distant metastasis with or without subsequent LR significantly affects survival chances, and despite our efforts as physicians almost all patients who experience such an event died of progressive disease. Therefore, the balance between the toxicity of intensive salvage treatments and quality of life in the remaining life span of these patients should be carefully considered. In case of late local recurrence (at least 2 years after treatment) there is no standard approach. The patients' age and preferences, previous treatment and tumor characteristics such as location, should all be considered and discussed in a multidisciplinary setting.

Radiotherapy seems protective for LR in all tumor sites combined, even in case

of good histological response. However, radiotherapy is not given randomly and is strongly correlated to patient and tumor characteristics; therefore, a note of caution in the interpretation of the results is required here. Patients treated with PORT generally have more tumor located in the pelvic, more inadequate surgical margins, and poorer histological response, which could have biased the results (see also Appendix 8.A). The incidence of local recurrence, especially in extremity Ewing sarcoma, is low. Only 2% (9 of 425) of the patients with extremity tumors developed isolated LR versus 8% (14 of 169) of the pelvic tumors and 8% (30 of 388) of the non-pelvic axial tumors. The number needed to treat (NNT) with surgery and radiotherapy to prevent the occurrence of a single LR is 72 for all tumor sites combined. In contrast, the NNT for extremity tumors is 80 and the NNT for pelvic tumors is 10. Which questions the value of radiotherapy in patients with an extremity Ewing sarcoma, where an individual patient with an extremity Ewing sarcoma might benefit, only few really are in need for this potentially toxic treatment, especially in the growing child. Radiotherapy is associated with a significant risk for secondary radiotherapy-induced malignancies, growth disturbance and postoperative complications of surgical reconstructions.[70] In case of Ewing sarcoma in a high-risk location, such as the pelvic or axial skeleton, this study showed that the administration of radiotherapy seems protective for LR, proton beam therapy could, in theory, be the solution in these cases; however, long-term data on radiation-induced late effects of proton beam radiation are not available yet. Prevention of distant metastasis and local recurrence appears to be the key to improve outcome in Ewing sarcoma, but distant metastases are still the main cause of treatment failure, and the results suggest that the use of radiotherapy is not protective for the occurrence of distant metastasis.

We compared the results presented in this article, which were computed using multiple imputation for missing data, to 776 complete cases and found that HRs were of similar magnitude. More details can be found in Appendix 8.B. We used a large cohort of patients with Ewing sarcoma, which strengthens this study. However, several limitations exist. Some subgroups are small; therefore, we cannot ensure that our findings of no effect of certain risk factors are not a result of the low number of events in these subgroups. Secondly, histological response and surgical margins were assessed by the local pathologist. The design of the study, in which a retrospective analysis was performed using a prospectively collected cohort, made revision of surgical margins and histological response not possible. Clear definitions were stated in the protocol, but differences in interpretation and evaluation could still exist. Third, cohorts often contain more variables than can reasonably be used for prediction, and for sufficient power one needs at least 10 events per variable. We therefore choose to select the most predictive and sensible predictors to be included in the analysis. Using a more extensive variable profile would have led to reduced predictability. Lastly, the recommendations for the use of radiotherapy were not consistently followed, and the results from this study are subjected to confounding by indication. Therefore, caution is needed when interpreting these results. Since the cohort used in this study is large and treated according to one protocol, we feel that the cohort adequately represents the population of interest and that the results are generalizable.

§8.5 Conclusion

Disease extent at diagnosis and histological response are the main risk factors for progression to distant metastasis or death after surgery. Tumor site and surgical margins are important risk factors for local recurrence. In case disease progression occurs, previous risk factors lose significance. Only time to recurrence is important for decision-making, since early LR (0-24 months) negatively influences survival. Both local recurrence and distant metastasis significantly affect survival, and despite our efforts as physicians, almost all patients who experience an event died of progressive disease. Therefore, the balance between the toxicity of intensive salvage treatments and quality of life in the remaining life span of these patients should be carefully considered in these cases. Radiotherapy seems protective for LR when all tumor sites are combined. However, a very low percentage of local recurrence in extremity tumors and the associated long-term toxicity with the use of radiotherapy questions the indication of radiotherapy in all extremity cases. Indications for radiotherapy should be explored further, preferably in a prospective randomized setting.

Appendix

§8.A Patient characteristics

Table 8.A.1

Characteristic	Surgery	Surgery + radiotherapy
	n (%)	n (%)
Total	550	432
Gender		
Male	335 (61)	255 (59)
Female	215 (39)	177 (41)
Age		
0-10 years	149 (27)	103 (24)
11-18 years	235 (43)	217 (50)
>18 years	166 (30)	112 (26)
Primary tumor localization		
Pelvic	80 (15)	89 (21)
Non-pelvic	470 (85)	343 (79)
Extremity	272 (50)	153 (35)
Axial	198 (35)	190 (44)
Volume at diagnosis		
<200 ml	336 (61)	241 (56)
≥ 200 ml	214 (39)	191 (44)
Disease extent at diagnosis		
Localized	431 (78)	339 (79)
Pulmonary metastasis	62 (11)	66 (15)
Extrapulmonary metastasis	57 (10)	27 (6)
Surgical margin		
Wide	453 (82)	264 (61)
Marginal	58 (11)	105 (24)
Intralesional	39 (7)	65 (15)
Histological response		
100%	284 (52)	142 (33)
90-99%	165 (30)	119 (28)
<90%	100 (18)	171 (39)
Transition to state		
Local recurrence	33 (6)	20 (5)
DMPulm	41 (8)	36 (8)
Extrapulmonary metastasis	113 (21)	99 (23)
Alive without disease	359 (65)	284 (66)

§8.B Complete case analysis

8. Individual risk evaluation for local recurrence and distant metastasis in Ewing sarcoma: a multistate model

Table 8.B.1: Hazard ratios and 95% confidence intervals for all prognostic factors and the different transitions in the multistate model only containing complete cases (n = 776)

	1: Surgery → LR HR 95% CI	2: Surgery → DM/pulm HR 95% CI	3: Surgery → DM other HR 95% CI	4: Surgery → death HR 95% CI	5: LR → death LR 95% CI	6: DM/pulm → death HR 95% CI	7: DMother → death HR 95% CI
Age							
0-10 years	1		1				1
11-18 years			1.52 0.99 - 2.32				0.84 0.53 - 1.35
>18 years			1.76 1.12 - 2.79				1.15 0.55 - 1.58
Tumor site							
Non-pelvic	1	1	1				1
Pelvic	2.56 1.23 - 5.33	1.50 0.83 - 2.72	1.53 1.06 - 2.20				0.99 0.66 - 1.49
Disease extent							
Localized + <200ml		1	1	1		1	1
Localized + ≥200 ml		1.26 0.68 - 2.33	1.22 0.84 - 1.76	1.82 0.81 - 4.06		1.34 0.60 - 3.00	0.96 0.64 - 1.45
Metastatic pulmonary		3.91 2.14 - 7.11	1.50 0.93 - 2.42	8.94 4.10 - 19.46		0.76 0.35 - 1.63	0.91 0.53 - 1.55
Metastatic other		0.67 0.16 - 2.86	3.22 2.03 - 5.11	10.82 4.70 - 24.94		1.35 0.16 - 11.73	1.63 0.98 - 2.71
Surgical margin							
Wide	1		1				1
Marginal / intralesional	2.40 1.07 - 5.39		0.88 0.58 - 1.32				1.30 0.84 - 2.01
Histological response							
100%	1	1	1	1	1	1	1
90-99%	1.50 0.71 - 3.16	1.62 0.90 - 2.91	1.64 1.12 - 2.42	1.27 0.55 - 2.94	1.00 0.40 - 2.50	1.97 0.89 - 4.36	0.90 0.58 - 1.39
<90%	1.08 0.42 - 2.79	2.25 1.21 - 4.17	2.81 1.91 - 4.12	8.32 4.16 - 16.64	0.57 0.15 - 2.22	1.32 0.57 - 3.06	1.15 0.76 - 1.73
Radiotherapy							
No	1	1	1	1	1		1
Yes	0.51 0.24 - 1.09	0.96 0.58 - 1.61	1.04 0.75 - 1.45	0.54 0.30 - 0.96	1.87 0.62 - 5.61		1.00 0.70 - 1.43
Time to recurrence							
>24 months					1		1
0-24 months					5.76 1.65 - 20.17		1.05 0.61 - 1.79

Abbreviations: DM/pulm, distant metastasis solitary pulmonary; DM/other, distant metastasis extrapulmonary with or without pulmonary metastasis; HR, hazard ratio; 95%CI, 95% confidence interval; LR, local recurrence

CHAPTER 9

Discussion and future perspectives

This thesis aimed at developing clinically relevant survival models, in particular the development and validation of prediction models for use in clinical practice. In the medical field statistical models need to address the research questions asked by clinicians and take into account the particular data structure of the problem at hand. Models used for prediction need to additionally be validated before they can be used in clinical practice. This final chapter, summarizes and discusses previous chapters. They are put in broader perspective and future research directions are suggested.

§9.1 Custom-made models

Statistical research is often motivated by a practical problem. A statistician is assigned to answer an important research question with an often already collected data set. Many times standard techniques can be applied to analyze the data. However, frequently the data does not perfectly meet the criteria for standard techniques or the research question suggests to use more advanced techniques. In some cases, existing methods cannot properly address the problem at hand and new methodology needs to be developed.

Survival methodology for example, was developed to analyze time to event data. This type of data is incomplete because not for all subjects the event of interest can be observed, referred to as right-censoring. This is the most common type of censoring found in survival data, however there are many more types of missing information that occur in practice. Methods have been developed to deal with different kinds of missing information.

Some survival methods such as the Kaplan-Meier estimator, the log-rank test and the Cox model are well established in clinical literature [60]. These methods are often adequate to answer clinically relevant research questions and clinicians can understand the output of these methods. Sometimes however, when the research question or the data are complex other statistical methods need to be used. Even though a variety of statistical models have been developed by statisticians there seems to be a disconnect between available methods and methods used for clinical research [118, 159, 132]. Several reasons for this exist. First, clinical researchers may not be aware of the best method to answer the research question. Second, a lack of understanding of complex methods and the output leads to them being undesirable. Third, a lack of available software makes methods difficult to apply.

Regardless the reason for the discrepancy between available statistical methods and methods applied in clinical research a close collaboration between statistician and clinician can contribute to the solution. Additionally, statisticians can benefit from such a collaboration twofold. On the one hand, statisticians can demonstrate the application of their methods and popularize them while at the same time benefiting clinical research. On the other hand, inspiration for new statistical methods may come from interesting clinical research questions. The interdisciplinary collaboration between statistician and clinician is of tremendous importance and may deliver great contributions to both fields [60, 132].

An example of statistical models that are under used in clinical practice are competing risks models [93, 156, 24]. In a frail patient population in which a non terminal event of interest is studied the competing risk of death may prevent the event of interest to be observed. Ignoring the risk of death will lead to wrong conclusions about the incidence of the event of interest and the effect of risk factors on the event of interest. A correct analysis of competing risks is therefore of great importance. In Chapter 2 we proposed a new competing risks model for two competing events that is able to quantify hospital heterogeneity using correlated frailties. The cause-specific proportional hazards model is used to model the risk of the two competing events and additionally frailties on the hospital level are modelled with gamma distribution. The model was developed for multi-center data which is data that was collected from multiple treatment centers. This is often necessary to obtain sufficient data for rare disease. The model adjusts for competing risks while at the same time quantifying the difference in risk between hospitals using Empirical Bayes methodology. It is a delicate task to compare the performance of hospitals and Empirical Bayes methodology is used to take the larger variances of smaller hospitals into account. In Chapter 5 Fine and Gray's model for competing risks was used to predict the cumulative incidence of local recurrence in soft tissue sarcoma patients. Local recurrence is an adverse event for patients who were surgically treated to remove the primary tumor. It means that tumor growth was found at site of surgery and its occurrence may mean additional surgery for some patients. It is therefore an important event for patients and clinicians and the predicted probability of local recurrence is an important information that can be used in patient care.

Multi-state models naturally extend competing risks models. The evolution of disease can sometimes be described by a series of events such as relapse, recovery and death that a patient may experience. These can be directly modelled with a multi-state model. Even though multi-state models can describe disease progression close to reality they are seldom applied in clinical practice. Apart from being more difficult to apply they are also still not well known in the clinical world. If applied they need to be carefully explained so that parameters are interpreted properly. Multi-state models are however an important tool that can give a deeper insight into the association of risk factors and the different states of disease which should be exploited in clinical research. In Chapter 4 and 8 multi-state models were used to understand the disease process for soft tissue sarcoma patients and Ewing sarcoma patients, respectively. In Chapter 4 a multi-state model to investigate the association of risk factors and adverse disease events for soft tissue sarcoma patients was proposed. Several prognostic factors such

as histology, grade, depth and size [83, 117, 169, 53, 73, 146, 140, 98, 139, 29] were already identified. An increase in risk for local recurrence following an intralesional margin resection was recognised [73, 139, 91], the association between margin status and survival and between local recurrence and survival however was still unclear [140, 106, 164, 29, 23, 74, 110, 102, 111]. After surgery a patient is alive with no evidence of disease and may then have local recurrence, distant metastasis or die. We proposed a three state multi-state model to obtain more insight into how certain prognostic factors affect phases of the evolution of disease with particular interest in surgical margin and radiotherapy. In our analysis distant metastasis and death were modelled as a single state, leading to three states (starting state, local recurrence, distant metastasis/death). The multi-state model showed that wide surgical margins and the use of (neo)adjuvant radiotherapy decreased the risk of local recurrence but had little effect on the risk of distant metastasis/death without local recurrence. This study contributed to a better understanding of the effect of risk factors on the different states of disease progression. In Chapter 8 we developed a multi-state model for Ewing sarcoma patients that were treated surgically to gain insight into the effect of surgical margin, histological response and radiotherapy on disease progression. Other studies had conducted single end point analyses of risk factors for adverse events in Ewing sarcoma. Multi-state models can estimate the effect of risk factors on different disease states simultaneously. Five states of disease progression were considered in the multi-state model. After surgery a patient enters the starting state, from here he may develop local recurrence, distant metastasis in the lungs, distant metastasis at a different site and he may die. The data only contained information of the first adverse event a patient experienced, so transitions from adverse events to other non terminal events were not possible. It was found that disease extent at start of treatment and histological response had a strong association for the transitions to distant metastasis and death. For the transition to local recurrence the location of the tumor and surgical margin were important risk factors.

Competing risks and multi-state models are natural model choices when the evolution of disease can be represented by multiple events. Another reason to apply more complex models lies in the data collection process. Methods for survival analysis were developed to deal with the common issue of right-censoring and other forms of missing information that can be present in survival data. Sometimes the event of interest cannot be observed exactly because it may only be diagnosed at pre-specified follow-up visits by for example a blood test. If the test is positive then it is only known that the event had occurred prior to the test. If regular tests are conducted then it is known that the event occurred between the time of the last negative and the first positive test. This type of censoring is referred to as interval-censoring. Methods for interval-censored data have been developed but they are not frequently used in clinical practice. We studied the effect that ignoring interval-censoring has on the predictive accuracy of a binary time-dependent marker in Chapter 3. The binary time-dependent marker may represent the presence or absence of disease that can be acquired over time. Prediction models for survival outcomes can inform patient and clinician and give an indication of a patients prognosis. Dynamic prediction models are able to incorporate updated information from time-dependent variables into the predictions

as they become available. The predictive accuracy of such time-dependent marker is therefore a parameter of interest. We conducted a simulation study to investigate the predictive accuracy of a binary time-dependent marker for the outcome death in the presence of interval-censoring. The marker may represent the presence or absence of disease that a patient can acquire over time. We studied several data scenarios and compared four different models, one of them ignoring interval-censoring. Different interval lengths between the observation of the time-dependent marker were studied. We used Area Under the Curve (AUC) measures that were adapted to survival outcomes to quantify the predictive accuracy of the time-dependent marker and found that the spacing between observation times did have a large effect on the AUC. The results of this study suggest to take interval-censoring into account when evaluating the predictive accuracy of a time-dependent marker.

The choice of statistical method can be motivated from the research question and the collected data. There are however choices to be made like for example regarding the selection of variables to be included in the model. Throughout this thesis clinical researchers chose the variables that were included in the models. In Chapter 4 and 8 multi-state models were estimated and not all covariate effects were estimated for all transitions. The reason for this was the limited number of patients that made some transitions. The selection of covariates for these transitions was done by clinical experts. Categorization of covariates and dichotomization of continuous covariates have been motivated clinically as well and these have changed throughout the time of this research. In Chapter 4 the variable age has been modelled in three categories while in Chapter 5, 6, and 7 age has been modelled continuously. The choice to include a quadratic effect of age into the models of Chapter 6, and 7 have been based on the significant nonlinearity of this variable. For the dynamic prediction model of Chapter 6, and 7 the covariates have been chosen by clinical experts, the interactions with time have been chosen using a backward selection procedure described in Chapter 6.

§9.2 Prediction models in clinical practice

Statistical models are able to make predictions as well as provide interpretable parameters, which contribute to the understanding of the underlying event process. The main goal of prediction models is to make good predictions which can be verified by validating predictions using a data set that has not been used in the model building process. In Chapter 5, 6 and 7 we developed and validated prediction models that predict the probability of local recurrence and survival for soft tissue sarcoma patients. In the first two chapters the models were validated internally using cross validation and in the last chapter the prediction model was validated externally using an external data set.

The parameters of the prediction models can be interpreted clinically and contribute to the understanding of the underlying disease process. The variables used in the models were chosen from clinical experts. The baseline covariates included in the models were of two different kinds: (1) they included patient- and disease-specific covariates, such as age, histology subtype, and tumor size. (2) they included treatment related covariates, such as surgical margin and radiotherapy treatment. The

first type of covariates are given and cannot be influenced. The treatment related covariates however, can be influenced by the treating clinician. In Chapter 4 and 5 we found that a wider surgical margin and radiotherapy treatment were associated with a decreased probability of local recurrence compared to an intralesional margin and no radiotherapy treatment. In Chapter 6, 7 we found that a wider surgical margin was associated to longer survival for a prediction time point just after surgery, but as the patient survived a period of time after surgery, no association with survival was found. Because of these associations the predictions of survival probability for a wider margin and radiotherapy treatment are higher than for an intralesional margin and no radiotherapy treatment. However these predictions do not make a fair comparison between treatment options because the prediction models developed in this thesis depend on current treatment practice. The data used for model development are not randomized controlled trials. For this reason they cannot be used to base treatment decisions on.

§9.3 Future perspectives

In this thesis clinically relevant survival models have been developed. In Chapter 3, 4 and 8 we used multi-state methodology to investigate risk factors for soft tissue sarcoma and Ewing sarcoma. The hazard ratios estimated with these models illustrate the association between covariates and disease related events and are very informative for clinicians. Because of the large number of parameters and insufficient transitions between states not all possible transitions between disease related events could be considered in the multi-state models. In Chapter 4 the event distant metastasis and death were combined because of the small number of patients transitioning from local recurrence to these states. Additionally, because of the small number of patients making some transitions it was not possible to estimate the effect of all relevant covariates for each transition. The soft tissue sarcoma data set has grown since the research conducted in Chapter 4 and now would allow for a more sophisticated multi-state model, one with more states and transitions and more covariates per transition. At the moment a project on the construction of a more complex multi-state model for soft tissue sarcoma patients is ongoing. In Chapter 8 a multi-state model for Ewing sarcoma was estimated based on data from the GPOH registry (Gesellschaft für Pädiatrische Onkologie und Hämatologie) treated in or according to the EURO-E.W.I.N.G 99 (EE99) protocol [8]. This data comprises only half of the total available data. At the moment researchers at Leiden University are working on receiving the second half of this data to conduct a more sophisticated analysis.

In Chapter 5, 6 and 7 we developed prediction models for patients with soft tissue sarcoma. We started with two models to be used at baseline just after surgery which were implemented in the PERSARC mobile application. They predict the probability of overall survival and local recurrence at 3, 5, and 10 years post surgery. In Chapter 6 a dynamic prediction model for soft tissue sarcoma patients was developed to predict the probability of surviving an additional 5 years from a prediction time point during follow-up. This dynamic model was updated and externally validated in Chapter 7 and is in the process of being implemented in the PERSARC mobile application.

Implementation of prediction tools into clinical practice remains challenging despite their utility. The adoption of prediction models to support shared decision making in clinical practice is a current subject of interest. A group of researchers from the Leiden University Medical Center were granted funding to implement shared decision making in treatment decisions in high-grade soft tissue sarcoma of the extremities in the Netherlands. The goal is to ensure that soft tissue sarcoma patients receive personalised care, in which risks and benefits of treatment options and patient preferences are balanced. Part of the implementation strategy is the introduction of the PERSARC mobile application to clinical practice through educational outreach. This type of research may lead to improved prediction tools and facilitate their introduction to clinical practice.

Finally, we stress the importance of future interdisciplinary collaboration between statisticians and clinicians. Multi-state models and dynamic prediction models are important tools that are underused in the clinical field. Other complex available methods should be introduced and explained carefully by statisticians to the clinical community so that they are able to benefit from their advantages.

Bibliography

- [1] Adjuvant! online program. <https://www.adjuvantonline.com>.
- [2] Data analysis and survival for personalised oncology. <https://sites.google.com/view/daspo>.
- [3] Mayo Clinic soft tissue sarcoma. <https://www.mayoclinic.org/diseases-conditions/soft-tissue-sarcoma/symptoms-causes/syc-20377725>. Accessed: 12-10-2019.
- [4] Personalised sarcoma care, apple store. <https://apps.apple.com/nl/app/personalized-sarcoma-care/id1189577003>.
- [5] Personalised sarcoma care, google play store. <https://play.google.com/store/apps/details?id=com.mobilepioneers.sarcoma&hl=en>.
- [6] Predict tool version 2.0: Breast cancer overall survival. <https://www.predict.nhs.uk>.
- [7] Prostate cancer nomograms. <https://www.mskcc.org/nomograms/prostate>.
- [8] Protocol: Euro ewing 99. european ewing tumour working initiative of national groups, ewing tumour studies 1999/ee99; euro ewing 99, 2006.
- [9] A. K. Akobeng. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*, 96(5):644–647, 2007.
- [10] E. Al-Absi, F. Farrokhyar, R. Sharma, K. Whelan, T. Corbett, M. Patel, and M. Ghert. A systematic review and meta-analysis of oncologic outcomes of pre-versus postoperative radiation in localized resectable soft-tissue sarcoma. *Ann Surg Oncol*, 17(5):1367–1374, 2010.
- [11] J. I. Albergo, C. L. Gaston, M. Laitinen, A. Darbyshire, L. M. Jeys, V. Sumathi, M. Parry, D. Peake, S. R. Carter, R. Tillman, A. T. Abudu, and R. J. Grimer. Ewing’s sarcoma: only patients with 100% of necrosis after chemotherapy should be classified as having a good response. *Bone Joint J*, 98-B(8):1138–1144, 2016.
- [12] P. Andersen and N. Keiding. Multi-state models for event history analysis. *Stat Methods Med Res*, 11(2):91–115, 2002.
- [13] J. R. Anderson, K. C. Cain, and R. D. Gelber. Analysis of survival by tumor response. *J Clin Oncol*, 1(11):710–719, 1983.

- [14] **A. J. Rueten-Budde**, S. Bosma, C. Lancia, A. Ranft, U. Dirksen, Krol, H. Gelderblom, M. van de Sande, P. Dijkstra, and M. Fiocco. Individual risk evaluation for local recurrence and distant metastasis in ewing sarcoma: a multistate model. *Pediatr Blood Cancer*, 66(e27943):doi: 10.1002/pbc.27943, 2019.
- [15] **A. J. Rueten-Budde**, C. Liu, A. Ranft, U. Dirksen, H. Gelderblom, and M. Fiocco. Dynamic prediction of overall survival for patients with ewing sarcoma. *Submitted*.
- [16] **A. J. Rueten-Budde**, H. Putter, and M. Fiocco. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker. *Submitted*.
- [17] **A. J. Rueten-Budde**, H. Putter, and M. Fiocco. Investigating hospital heterogeneity with a competing risks frailty model. *Stat Med*, 38(2):269–288, 2018.
- [18] **A. J. Rueten-Budde**, M. van de Sande, V. van Praag, PERSARC studygroup, and M. Fiocco. External validation and adaptation of a dynamic prediction model for patients with high-grade extremity soft tissue sarcoma. *Submitted*.
- [19] **A. J. Rueten-Budde**, V. van Praag, PERSARC studygroup, M. van de Sande, and M. Fiocco. Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma. *Surg Oncol*, 27(4):695–701, 2018.
- [20] **A. J. Rueten-Budde**, V. M. van Praag, L. M. Jeys, M. K. Laitinen, R. Pollock, W. Aston, J. A. van der Hage, P. S. Dijkstra, P. C. Ferguson, A. M. Griffin, J. J. Willeumier, J. S. Wunder, M. A. van de Sande, and M. Fiocco. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: personalised sarcoma care (persarc). *Eur J Cancer*, 83:313–323, 2017.
- [21] **A. J. Rueten-Budde**, J. J. Willeumier, L. M. Jeys, M. Laitinen, R. Pollock, W. Aston, P. D. S. Dijkstra, P. C. Ferguson, A. M. Griffin, J. S. Wunder, M. Fiocco, and M. A. J. van de Sande. Individualised risk assessment for local recurrence and distant metastases in a retrospective transatlantic cohort of 687 patients with high-grade soft tissue sarcomas of the extremities: a multistate model. *BMJ Open*, 7(2):e012930, 2017.
- [22] E. Arpacı, T. Yetisyigit, M. Seker, D. Uncu, U. Üyetürk, B. Oksuzoglu, U. Demirci, U. Coskun, M. Kucukoner, A. Isikdogan, M. Inanc, N. Alkis, and M. Ozkan. Prognostic factors and clinical outcome of patients with ewing’s sarcoma family of tumors in adults: Multicentric study of the anatolian society of medical oncology. *Med Oncol*, 30(1):469, 2013.
- [23] I. Atean, Y. Pointreau, P. Rosset, P. Garaud, G. De-Pinieux, and G. Calais. Prognostic factors of extremity soft tissue sarcoma in adults. a single institutional analysis. *Cancer/Radiother*, 16(8):661–666, 2012.
- [24] P. C. Austin and J. P. Fine. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. *Stat Med*, 36(8):1203–1209, 2017.

-
- [25] G. Bacci, S. Ferrari, A. Longhi, D. Donati, E. Barbieri, C. Forni, F. Bertoni, M. Manfrini, S. Giacomini, and P. Bacchini. Role of surgery in local treatment of ewing's sarcoma of the extremities in patients undergoing adjuvant and neoadjuvant chemotherapy. *Oncol Rep*, 11(1):111–120, 2004.
- [26] G. Bacci, A. Longhi, A. Briccoli, F. Bertoni, M. Versari, and P. Picci. The role of surgical margins in treatment of ewing's sarcoma family tumors: experience of a single institution with 512 patients treated with adjuvant and neoadjuvant chemotherapy. *Int J Radiat Oncol Biol Phys*, 65(3):766–772, 2006.
- [27] T. Balan and H. Putter. *frailtyEM: an R package for estimating semiparametric shared frailty models*, 2017. R package version 0.7.9.
- [28] J. D. Beane, J. C. Yang, D. White, S. M. Steinberg, S. A. Rosenberg, and U. Rudloff. Efficacy of adjuvant radiation therapy in the treatment of soft tissue sarcoma of the extremity: 20-year follow-up of a randomized prospective trial. *Ann Surg Oncol*, 21(8):2484–2489, 2014.
- [29] D. J. Biau, P. C. Ferguson, P. Chung, A. M. Griffin, C. N. Catton, B. O'Sullivan, and J. S. Wunder. Local recurrence of localized soft tissue sarcoma: a new look at old predictors. *Cancer*, 118(23):5867–5877, 2012.
- [30] B. Biswas, S. Rastogi, S. Khan, N. Shukla, S. Deo, S. Agarwala, B. Mohanti, M. Sharma, S. Vishnubhatla, and S. Bakhshi. Developing a prognostic model for localized ewing sarcoma family of tumors: a single institutional experience of 224 cases treated with uniform chemotherapy protocol. *J Surg Oncol*, 111(6):683–689, 2015.
- [31] A. Boruvka and R. Cook. Sieve estimation in a markov illness-death process under dual censoring. *Biostatistics*, 17(2):350–363, 2016.
- [32] S. Bosma, C. Lancia, **A. J. Rueten-Budde**, A. Ranft, H. Gelderblom, M. Fiocco, M. van de Sande, P. Dijkstra, and U. Dirksen. Easy-to-use clinical tool for survival estimation in ewing sarcoma at diagnosis and after surgery. *Sci Rep*, 9(11000), 2019.
- [33] M. F. Brennan, E. S. Casper, L. B. Harrison, M. H. Shiu, J. Gaynor, and S. I. Hajdu. *Ann Surg*, 214(3):328–338, 1991.
- [34] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99, 1974.
- [35] O. Cahlon, M. F. Brennan, X. Jia, L.-X. Qin, S. Singer, and K. M. Alektiar. A postoperative nomogram for local recurrence risk in extremity soft tissue sarcomas after limb-sparing surgery without adjuvant radiation. *Ann Surg*, 255(2):343–347, 2012.
- [36] D. Callegaro, R. Miceli, S. Bonvalot, P. Ferguson, D. C. Strauss, and A. Levy. Impact of perioperative chemotherapy and radiotherapy in patients with

- primary extremity soft tissue sarcoma: retrospective analysis across major histological subtypes and major reference centres. *Eur J Cancer*, 105:19–27, 2018.
- [37] D. Callegaro, R. Miceli, S. Bonvalot, P. Ferguson, D. C. Strauss, A. Levy, A. Griffin, A. J. Hayes, S. Stacchiotti, C. L. Pechoux, M. J. Smith, M. Fiore, A. P. D. Tos, H. G. Smith, L. Mariani, J. S. Wunder, R. E. Pollock, P. G. Casali, and A. Gronchi. Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: a retrospective analysis. *Lancet Oncol*, 17(5):671–680, 2016.
- [38] A. Carneiro, P.-O. Bendahl, J. Engellau, H. A. Domanski, C. D. Fletcher, P. Rissler, A. Rydholm, and M. Nilbert. A prognostic model for soft tissue sarcoma of the extremities and trunk wall based on size, vascular invasion, necrosis, and growth pattern. *Cancer*, 117(6):1279–1287, 2011.
- [39] G. Casella and R. Berger. *Statistical Inference*. Duxbury Press, 2001.
- [40] D. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.
- [41] J. Coindre, P. Terrier, N. Bui, F. Bonichon, F. Collin, V. Le Doussal, A. Mandard, M. Vilain, J. Jacquemier, H. Duplay, X. Sastre, C. Barlier, M. Henry-Amar, J. Macé-Lesech, and G. Contesso. Prognostic factors in adult patients with locally controlled soft tissue sarcoma. a study of 546 patients from the french federation of cancer centers sarcoma group. *J Clin Oncol*, 14(3):869–877, 1996.
- [42] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *Ann Int Med*, 162(1):55–63, 2015.
- [43] S. Cotterill, S. Ahrens, M. Paulussen, H. Jürgens, P. Voûte, H. Gadner, and A. Craft. Prognostic factors in ewing’s tumor of bone: analysis of 975 patients from the european intergroup cooperative ewing’s sarcoma study group. *J Clin Oncol*, 18(17):3108–3114, 2000.
- [44] D. Cox. Regression models and life-tables. *J Royal Stat Soc*, 34(2):187–220, 1972.
- [45] L. Davis and C. Ryan. Preoperative therapy for extremity soft tissue sarcomas. *Curr Treat Options Oncol*, 16(6):25, 2015.
- [46] G. de Bock, H. Putter, J. Bonnema, J. van der Hage, H. Bartelink, and C. van de Velde. The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model. *Breast Cancer Res Treat*, 117(2):401–408, 2009.

-
- [47] N. A. de Glas, E. Bastiaannet, C. C. Engels, A. J. M. de Craen, H. Putter, C. J. H. van de Velde, A. Hurria, G. J. Liefers, and J. E. A. Portielje. Validity of the online predict tool in older patients with breast cancer: a population-based study. *Br J Cancer*, 114(4):395–400, 2016.
- [48] L. de Wreede, M. Fiocco, and H. Putter. mstate: An r package for the analysis of competing risks and multi-state models. *J Stat Softw*, 38(7):1–30, 2011.
- [49] L. C. de Wreede, M. Fiocco, and H. Putter. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed*, 99(3):261–274, 2010.
- [50] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J Royal Stat Soc*, 39(1):1–38, 1977.
- [51] S. Dixon, G. Darlington, and A. Desmond. A competing risks model for correlated data based on the subdistribution hazard. *Lifetime Data Anal*, 17(4):473–495, 2011.
- [52] K. R. Duchman, Y. Gao, and B. J. Miller. Prognostic factors for survival in patients with ewing’s sarcoma using the surveillance, epidemiology, and end results (seer) program database. *Cancer Epidemiol*, 39(2):189–195, 2015.
- [53] F. Eilber, G. Rosen, S. Nelson, M. Selch, F. Dorey, J. Eckardt, and F. Eilber. High-grade extremity soft tissue sarcomas: Factors predictive of local recurrence and its effect on morbidity and mortality. *Ann Surg*, 237(2):218–226, 2003.
- [54] J. Engellau, V. Samuelsson, H. Anderson, B. Bjerkehagen, P. Rissler, K. Sundby-Hall, and A. Rydholm. Identification of low-risk tumours in histological high-grade soft tissue sarcomas. *Eur J Cancer*, 43(13):1927–1934, 2007.
- [55] W. Enneking, S. Spanier, Goodman, and MA. A system for the surgical staging of musculoskeletal sarcoma. *Clin Orthop Relat Res*, 153:106–120, 1980.
- [56] ESMO/European Sarcoma Network Working Group. Soft tissue and visceral sarcomas: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 25(suppl_3):iii102–iii112, 2014.
- [57] P. Fine and R. Gray. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*, 94(446):496–509, 1999.
- [58] M. Fiocco, H. Putter, and J. van Houwelingen. Meta-analysis of pairs of survival curves under heterogeneity: a poisson correlated gamma-frailty approach. *Stat Med*, 28(30):3782–3797, 2009.
- [59] M. Fiocco, H. Putter, and J. van Houwelingen. A new serially correlated gamma-frailty process for longitudinal count data. *Biostatistics*, 10(2):245–257, 2009.
- [60] T. R. Fleming and D. Y. Lin. Survival analysis in clinical trials : past developments and future directions. *Biometrics*, 56(4):971–983, 2000.

- [61] C. Fletcher, P. H. J.A. Bridge, and F. Mertens. *WHO Classification of Tumours of Soft Tissue and Bone*. WHO, 2013.
- [62] D. B. Y. Fontein, M. Klinten Grand, J. W. R. Nortier, C. Seynaeve, E. Meershoek-Klein Kranenbarg, L. Y. Dirix, C. J. H. van de Velde, and H. Putter. Dynamic prediction in breast cancer: proving feasibility in clinical practice using the team trial. *Ann Oncol*, 26(6):1254–1262, 2015.
- [63] S. Foulon, B. Brennan, N. Gaspar, U. Dirksen, L. Jeys, A. Cassoni, L. Claude, B. Seddon, P. Marec-berard, J. Whelan, M. Paulussen, A. Streitbuenger, O. Oberlin, H. Juergens, R. Grimer, and M.-C. Le Deley. Can postoperative radiotherapy be omitted in localised standard-risk ewing sarcoma? an observational study of the euro-e.w.i.n.g group. *Eur J Cancer*, 61:128–136, 2016.
- [64] H. Frydman. Nonparametric estimation of a markov ‘illness-death’ process from interval-censored observations, with application to diabetes survival data. *Biometrika*, 82(4):773–789, 1995.
- [65] H. Frydman and M. Szarek. Nonparametric estimation in a markov ‘illness-death’ process from interval censored observations with missing intermediate transition status. *Biometrics*, 65(1):143–151, 2009.
- [66] P. G Casali, N. Abecassis, S. Bauer, R. Biagini, S. Bielack, S. Bonvalot, I. Boukovinas, J. Bovee, T. Brodowicz, J. Martin-Broto, A. Buonadonna, E. De Álava, A. Tos, X. G Del Muro, P. Dileo, M. Eriksson, A. Fedenko, V. Ferraresi, A. Ferrari, and J.-Y. Blay. Soft tissue and visceral sarcomas: Esmo-euracan clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 29(Supplement_4):iv51–iv67, 2018.
- [67] N. Gaspar, A. Rey, P. Marec-berard, J. Michon, J.-C. Gentet, M.-D. Tabone, H. Roché, A. Defachelles, O. Lejars, E. Plouvier, C. Schmitt, B. Bui, P. Boutard, S. Taque, M. Munzer, J.-P. Vannier, D. Plantaz, N. Entz-Werle, and O. Oberlin. Risk adapted chemotherapy for localised ewing’s sarcoma of bone: the french ew93 study. *Eur J Cancer*, 48(9):1376–1385, 2012.
- [68] N. Gaspar, D. S Hawkins, U. Dirksen, I. J Lewis, S. Ferrari, M.-C. Le Deley, H. Kovar, R. Grimer, J. Whelan, L. Claude, O. delattre, M. Paulussen, P. Picci, K. Hall, H. van den Berg, R. Ladenstein, J. Michon, L. Hjorth, I. Judson, and O. Oberlin. Ewing sarcoma: current management and future approaches through collaboration. *J Clin Oncol*, 33(27):3036–3046, 2015.
- [69] P. Gilbert and R. Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2016. R package version 2016.8-1.
- [70] J. P. Ginsberg, P. Goodman, W. Leisenring, K. K. Ness, P. A. Meyers, S. L. Wolden, S. M. Smith, M. Stovall, S. Hammond, L. L. Robison, and K. C. Oeffinger. Long-term survivors of childhood ewing sarcoma: report from the childhood cancer survivor study. *J Natl Cancer Inst*, 102(16):1272–1283, 2010.

-
- [71] H. Goldstein and D. Spiegelhalter. League tables and their limitations: statistical issues in comparisons of institutional performance. *J Royal Stat Soc*, 159:385–443, 1996.
- [72] M. Gorfine and L. Hsu. Frailty-based competing risks model for multivariate survival data. *Biometrics*, 67(2):415–426, 2011.
- [73] A. Gronchi, P. Casali, L. Mariani, R. Miceli, M. Fiore, S. Lo Vullo, R. Bertulli, P. Collini, L. Lozza, P. Olmi, and J. Rosai. Status of surgical margins and prognosis in adult soft tissue sarcomas of the extremities: A series of patients treated at a single institution. *J Clin Oncol*, 23(1):96–104, 2005.
- [74] A. Gronchi, S. Lo Vullo, C. Colombo, P. Collini, S. Stacchiotti, L. Mariani, M. Fiore, and P. G. Casali. Extremity soft tissue sarcoma in a series of patients treated at a single institution: local control directly impacts survival. *Ann Surg*, 251(3):506–511, 2010.
- [75] W. M. Guerrero and J. L. Deneve. Local recurrence of extremity soft tissue sarcoma. *Surg Clin North Am*, 96(5):1157–1174, 2016.
- [76] K. R. Gundle, L. Kafchinski, S. Gupta, A. M. Griffin, B. C. Dickson, P. W. Chung, C. N. Catton, B. O’Sullivan, J. S. Wunder, and P. C. Ferguson. Analysis of margin classification systems for assessing the risk of local recurrence after soft tissue sarcoma resection. *J Clin Oncol*, 36(7):704–709, 2018.
- [77] P. Gustafson, M. Åkerman, T. Alvegård, J.-M. Coindre, C. Fletcher, A. Rydholm, and H. Willén. Prognostic information in soft tissue sarcoma using tumour size, vascular invasion and microscopic tumour necrosis-the sin-system. *Eur J Cancer*, 39(11):1568–1576, 2003.
- [78] F. Harrell, K. Lee, and D. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15:361–387, 1996.
- [79] P. Heagerty, T. Lumley, and M. Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56:337–344, 2000.
- [80] P. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61(1):92–105, 2005.
- [81] K. Hoang, . Yubo Gao, PhD, and M. Benjamin J. Miller, MD. The variability in surgical margin reporting in limb salvage surgery for sarcoma. *Iowa Orthop J*, 35(1):181–186, 2015.
- [82] J. Honaker, G. King, and M. Blackwell. Amelia ii: a program for missing data. *J Stat Softw*, 45(7):1–54, 2011.
- [83] A. Italiano, A. Le Cesne, J. Mendiboure, J. Blay, S. Piperno-Neumann, C. Chevreau, C. Delcambre, N. Penel, P. Terrier, D. Ranchere-Vince, M. Lae,

- S. Le Guellec, J. Michels, Y. Robin, C. Bellera, and S. Bonvalot. Prognostic factors and impact of adjuvant treatments on local and metastatic relapse of soft-tissue sarcoma patients in the competing risks setting. *Cancer*, 120(21):3361–3369, 214.
- [84] B. J Miller, Y. Gao, and K. R Duchman. Does surgery or radiation provide the best overall survival in ewing’s sarcoma? a review of the national cancer data base. *J Surg Oncol*, 116(3):384–390, 2017.
- [85] C. Jackson. Multi-state models for panel data: The msm package for r. *J Stat Softw*, 38(8), 2011.
- [86] P. Joly, D. Commenges, C. Helmer, and L. Letenneur. A penalized likelihood approach for an illness-death model with interval-censored data : application to age-specific incidence of dementia. *Biostatistics*, 3(3):433–443, 2002.
- [87] V. Kainhofer, M. Smolle, J. Szkandera, B. Liegl-Atzwanger, W. Maurer-Ertl, A. Gerger, J. Riedl, and A. Leithner. The width of resection margins influences local recurrence in soft tissue sarcoma patients. *Eur J Surg Oncol*, 42(6):899–906, 2016.
- [88] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*, 53(282):457–481, 1958.
- [89] S. Katsahian, S. C. M Resche-Rigon, and R. Porcher. Analysing multicentre competing risks data with a mixed proportional hazards model for the subdistribution. *Stat Med*, 25(24):4267–4278, 2006.
- [90] N. Kawaguchi, A. Ahmed, S. Matsumoto, J. Manabe, and Y. Matsushita. The concept of curative margin in surgery for bone and soft tissue sarcoma. *Clin Orthop Relat Res*, 419:165–172, 2004.
- [91] N. Kawaguchi, S. Matumoto, and J. Manabe. New method of evaluating the surgical margin and safety margin for musculoskeletal sarcoma, analysed on the basis of 457 surgical cases", journal="journal of cancer research and clinical oncology. *Journal of Cancer Research and Clinical Oncology*, 121(9):555–563, 1995.
- [92] J. P. Klein and M. L. Moeschberger. *Survival Analysis*. Springer, 1997.
- [93] M. T. Koller, H. Raatz, E. W. Steyerberg, and M. Wolbers. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*, 31:1089–1097, 2012.
- [94] T. Kuklo, H. Temple, B. Owens, J. Juliano, R. Islinger, Y. Andejas, D. Frassica, and B. Berrey. Preoperative versus postoperative radiation therapy for soft-tissue sarcomas. *Am J Orthop (Belle Mead NJ)*, 34(2):75–80, 2005.

- [95] R. Ladenstein, U. Pötschger, M. Cécile Le Deley, J. Whelan, M. Paulussen, O. Oberlin, H. van den Berg, U. Dirksen, L. Hjorth, J. Michon, I. Lewis, A. Craft, and H. Juergens. Primary disseminated multifocal ewing sarcoma: results of the euro-ewing 99 trial. *J Clin Oncol*, 28(20):3284–3291, 2010.
- [96] J. Lee, B. H. Hoang, A. Ziogas, and J. A. Zell. Analysis of prognostic factors in ewing sarcoma using a population-based cancer registry. *Cancer*, 116(8):1964–1973, 2010.
- [97] K. Leffondré, C. Touraine, C. Helmer, and P. Joly. Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the cox model? *Int J Epidemiol*, 42(4):1177–1186, 2013.
- [98] J. J. Lewis, D. Leung, M. Heslin, J. M. Woodruff, and M. F. Brennan. Association of local recurrence with subsequent survival in extremity soft tissue sarcoma. *J Clin Oncol*, 15(2):646–652, 1997.
- [99] B. Liqueet, J. Timsit, and V. Rondeau. Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC Med Res Methodol*, 12(79):doi: 10.1186/1471–2288–12–79, 2012.
- [100] R. Little and D. Rubin. *Multiple Imputation for Nonresponse in Surveys, 2nd Edition*. Wiley, 2002.
- [101] T. Louis. Finding the observed information matrix when using the em algorithm. *J Royal Stat Soc*, 44(2):226–233, 1982.
- [102] K. Maretty-Nielsen, N. Aggerholm-Pedersen, A. Safwat, P. H. J. rgensen, B. H. Hansen, S. Baerentzen, A. B. Pedersen, and J. Keller. Prognostic factors for local recurrence and mortality in adult soft tissue sarcoma of the extremities and trunk wall: a cohort study of 922 consecutive patients. *Acta Orthop*, 85(3):323–332, 2014.
- [103] L. Mariani, R. Miceli, M. W. Kattan, M. F. Brennan, M. Colecchia, M. Fiore, P. G. Casali, and A. Gronchi. Validation and adaptation of a nomogram for predicting the survival of patients with extremity soft tissue sarcoma using a three-grade system. *Cancer*, 103(2):402–408, 2005.
- [104] M. Mastboom, E. Staals, F. Verspoor, **A. J. Rueten-Budde**, S. Stacchiotti, E. Palmerini, G. Schaap, P. Jutte, W. Aston, A. Leithner, D. Dammerer, A. Takeuchi, Q. Thio, X. Niu, J. Wunder, TGCT study group, and M. van de Sande. Surgical treatment of localized-type tenosynovial giant cell tumours of large joints. *J Bone Joint Surg*, 101(14):1309–1318, 2019.
- [105] M. J. L. Mastboom, E. Palmerini, F. G. M. Verspoor, **A. J. Rueten-Budde**, S. Stacchiotti, E. L. Staals, G. R. Schaap, P. C. Jutte, W. Aston, H. Gelderblom, A. Leithner, D. Dammerer, A. Takeuchi, Q. Thio, X. Niu, J. S. Wunder, TGCT Study Group, and M. A. J. van de Sande. Surgical outcomes of patients with

- diffuse-type tenosynovial giant-cell tumours: an international, retrospective, cohort study. *Lancet Oncol*, 20(6):877–886, 2019.
- [106] M. D. Mc Kee, D. F. Liu, J. J. Brooks, J. F. Gibbs, D. L. Driscoll, and W. G. Kraybill. The prognostic significance of margin width for extremity and trunk sarcoma. *J Surg Oncol*, 85(2):68–76, 2004.
- [107] W. Mendenhall, D. Indelicato, M. Scarborough, R. Zlotecki, C. Gibbs, N. Mendenhall, C. Mendenhall, and W. Enneking. The management of adult soft tissue sarcomas. *Am J Clin Oncol*, 32(4):436–442, 2009.
- [108] D. Müller, G. Beltrami, G. Scoccianti, F. Frenos, and R. Capanna. Combining limb-sparing surgery with radiation therapy in high-grade soft tissue sarcoma of extremities - is it effective? *Eur J Surg Oncol*, 42(7):1057–1063, 2016.
- [109] G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand J Stat*, 19(1):25–43, 1992.
- [110] E. N. Novais, B. Demiralp, J. Alderete, M. C. Larson, P. S. Rose, and F. H. Sim. Do surgical margin and local recurrence influence survival in soft tissue sarcomas? *Clin Orthop Relat Res*, 468(11):3003–3011, 2010.
- [111] P. W. O’Donnell, A. M. Griffin, W. C. Eward, A. Sternheim, C. N. Catton, P. W. Chung, B. O’Sullivan, P. C. Ferguson, and J. S. Wunder. The effect of the setting of a positive surgical margin in soft tissue sarcoma. *Cancer*, 120(18):2866–2875, 2014.
- [112] B. O’Sullivan, A. M. Davis, R. Turcotte, R. Bell, C. Catton, P. Chabot, J. Wunder, R. Kandel, K. Goddard, A. Sadura, J. Pater, and B. Zee. Preoperative versus postoperative radiotherapy in soft-tissue sarcoma of the limbs: a randomised trial. *Lancet*, 359(9325):2235–2241, 2002.
- [113] T. Ozaki M.D, A. Hillmann, C. Hoffmann, C. Rube M.D, S. Blasius M.D, J. Dunst, H. Jürgens, and W. Winkelmann M.D. Significance of surgical margin on the prognosis of patients with ewing’s sarcoma: a report from the cooperative ewing’s sarcoma study. *Cancer*, 78(4):892–900, 1996.
- [114] S. Pasquali, C. Colombo, S. Pizzamiglio, P. Verderio, D. Callegaro, S. Stacchiotti, J. Martin Broto, A. Lopez-Pousa, S. Ferrari, A. Poveda, A. De Paoli, V. Quagliuolo, J. C. Jurado, A. Comandone, G. Grignani, R. De Sanctis, E. Palassini, A. Llomboart-Bosch, A. P. Dei Tos, P. G. Casali, P. Picci, and A. Gronchi. High-risk soft tissue sarcomas treated with perioperative chemotherapy: improving prognostic classification in a randomised clinical trial. *Eur J Cancer*, 93:28–36, 2018.
- [115] S. Pasquali, S. Pizzamiglio, N. Touati, S. Litiere, S. Marreaud, B. Kasper, H. Gelderblom, S. Stacchiotti, I. Judson, A. P. Dei Tos, P. Verderio, P. G. Casali, P. J. Woll, A. Gronchi, E.-S. Tissue, and B. S. Group. The impact of

- chemotherapy on survival of patients with extremity and trunk wall soft tissue sarcoma: revisiting the results of the eortc-stbsg 62931 randomised trial. *Eur J Cancer*, 109:51–60, 2019.
- [116] J. Petersen, P. Andersen, and R. Gill. Variance components models for survival data. *Stat Neerl*, 50:193–211, 1996.
- [117] P. Pisters, D. Leung, J. Woodruff, W. Shi, and M. Brennan. Analysis of prognostic factors in 1,041 patients with localized soft tissue sarcomas of the extremities. *J Clin Oncol*, 14(5):1679–1689, 1996.
- [118] E. M. Pullenayegum, R. W. Platt, M. Barwick, B. M. Feldman, M. Offringa, and L. Thabane. Knowledge translation in biostatistics: a survey of current practices, preferences, and barriers to the dissemination and uptake of new statistical methods. *Stat Med*, 35(6):805–818, 2016.
- [119] H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*, 26(11):2389–2430, 2007.
- [120] H. Putter and J. van Houwelingen. Dynamic frailty models based on compound birth-death processes. *Biostatistics*, 16(3):550–564, 2015.
- [121] H. Putter and J. van Houwelingen. Frailties in multi-state models: are they identifiable? do we need them? *Stat Methods Med Res*, 24(6):675–692, 2015.
- [122] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [123] F. Rotolo, V. Rondeau, and C. Legrand. Incorporation of nested frailties into semiparametric multi-state models. *Stat Med*, 35(4):609–621, 2016.
- [124] P. Royston and D. G. Altman. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol*, 13:33, 2013.
- [125] D. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, 1987.
- [126] D. B. Rubin. Multiple imputation after 18+ years. *J Am Stat Assoc*, 91(434):473–489, 1996.
- [127] S. Sampath, T. E. Schultheiss, Y. J. Hitchcock, R. L. Randall, D. C. Shrieve, and J. Y. Wong. Preoperative versus postoperative radiotherapy in soft-tissue sarcoma: multi-institutional analysis of 821 patients. *Int J Radiat Oncol Biol Phys*, 81(2):498–505, 2011.
- [128] M. Sampo, M. Tarkkanen, E. Tukiainen, P. Popov, P. Gustafson, M. Lundin, T. Böhling, C. Blomqvist, and J. Lundin. A web-based prognostic tool for extremity and trunk wall soft tissue sarcomas and its external validation. *Br J Cancer*, 106(6):1076–1082, 2012.
- [129] J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall, 1997.

- [130] J. L. Schafer and M. K. Olsen. Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivar Behav Res*, 33(4):545–571, 1998.
- [131] T. Scheike, Y. Sun, M. Zhang, and T. Jensen. A semiparametric random effects model for multivariate competing risks data. *Biometrika*, 97(1):133–145, 2010.
- [132] M. J. Schell. Identifying key statistical papers from 1985 to 2002 using citation data for applied biostatisticians. *Am Stat*, 64(4):310–317, 2010.
- [133] M. Schemper and T. L. Smith. A note on quantifying follow-up in studies of failure time. *Controlled Clin Trials*, 17(4):343–346, 1996.
- [134] A. Schuck, S. Ahrens, M. Paulussen, M. Kuhlen, S. Könemann, C. Rübe, W. Winkelmann, R. Kotz, J. Dunst, N. Willich, and H. Jürgens. Local therapy in localized ewing tumors: results of 1058 patients treated in the ccess 81, ccess 86, and eicess 92 trials. *Int J Radiat Oncol Biol Phys*, 55(1):168–177, 2003.
- [135] E. Schutgens, P. Picci, D. Baumhoer, R. Pollock, J. Bovee, P. Hogendoorn, P. Dijkstra, **A. J. Rueten-Budde**, P. Jutte, F. Traub, A. Leithner, P. Tunn, P. Funovics, G. Sys, M. Julian, G. Schaap, H. Dürr, J. Harges, J. Healy, R. Capanna, D. Biau, A. Gomez-Brouchet, J. Wunder, T. Cosker, M. Laitinen, X. Niu, V. Kostjuk, Adamantinoma research group, and M. van de Sande. Surgical outcome and oncological survival of osteofibrous dysplasia-like- and classic adamantinoma: an international multicenter study of 318 cases. *Submitted*.
- [136] G. Shankar, C. R. Pinkerton, A. Atra, S. Ashley, I. Lewis, D. Spooner, S. Cannon, R. Grimer, S. Cotterill, and A. Craft. Local therapy and other factors influencing site of relapse in patients with localised ewing's sarcoma. united kingdom children's cancer study group (ukccsg). *Eur J Cancer*, 35(12):1698–704, 1999.
- [137] R. Siegel, D. Naishadham, and A. Jemal. Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, 63(1):11–30, 2013.
- [138] J. A. C. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338:b2393, 2009.
- [139] A. Stojadinovic, D. H. Y. Leung, A. Hoos, D. P. Jaques, J. J. Lewis, and M. F. Brennan. Analysis of the prognostic significance of microscopic margins in 2,084 localized primary adult soft tissue sarcomas. *Ann Surg*, 235(3):424–434, 2002.
- [140] K. K. Tanabe, R. E. Pollock, L. M. Ellis, A. Murphy, N. Sherman, and M. M. Romsdahl. Influence of surgical margins on outcome in patients with preoperatively irradiated extremity soft tissue sarcomas. *Cancer*, 73(6):1652–1659, 1994.

-
- [141] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000.
- [142] T. Therneau. *A package for survival analysis in s*, 2015. R package version 2.38.
- [143] N. Thomas, N. Longford, and J. Rolph. Empirical bayes methods for estimating hospital-specific mortality rates. *Stat Med*, 13(9):889–903, 1994.
- [144] C. Touraine, T. Gerds, and P. Joly. Smoothhazard: An r package for fitting regression models to interval-censored observations of illness-death models. *J Stat Softw*, 79(7), 2017.
- [145] M. Trojani, G. Contesso, J. M. Coindre, J. Rouesse, N. B. Bui, A. De Mascarel, J. F. Goussot, M. David, F. Bonichon, and C. Lagarde. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *Int J Cancer*, 33(1):37–42, 1984.
- [146] C. Trovik, H. Bauer, T. Alvegård, H. Anderson, C. Blomqvist, O. Berlin, P. Gustafson, G. Sæter, and A. Walløe. Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the scandinavian sarcoma group register. *Eur J Cancer*, 36(6):710–716, 2000.
- [147] J. van der Hage, C. van de Velde, J. Julien, J. Floiras, T. Delozier, C. Vandervelden, and L. Duchateau. Improved survival after one course of perioperative chemotherapy in early breast cancer patients. long-term results from the european organization for research and treatment of cancer (eortc) trial 10854. *Eur J Cancer*, 37(17):2184–2193, 2001.
- [148] J. C. van Gaal, E. Bastiaannet, M. Schaapveld, R. Otter, J. C. Kluin-Nelemans, E. S. J. M. de Bont, and W. T. A. van der Graaf. Cancer in adolescents and young adults in north netherlands (1989–2003): increased incidence, stable survival and high incidence of second primary tumours. *Ann Oncol*, 20(2):365–373, 2008.
- [149] H. van Houwelingen and H. Putter. *Dynamic Prediction in Clinical Survival Analysis*. CRC Press/Chapman & Hall, 2011.
- [150] H. C. van Houwelingen. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*, 19(24):3401–3415, 2000.
- [151] H. C. van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scand J Stat*, 34(1):70–85, 2007.
- [152] J. van Houwelingen. The role of empirical bayes methodology as a leading principle in modern medical statistics. *Biom J*, 56(6):919–932, 2014.
- [153] J. van Houwelingen, R. Brand, and T. Louis. Empirical Bayes methods for monitoring health care quality. <https://www.lumc.nl/sub/3020/att/EmpiricalBayes.pdf>, 2000.

- [154] J. C. van Houwelingen and S. Le Cessie. Predictive value of statistical models. *Stat Med*, 9(11):1303–1325, 1990.
- [155] V. van Praag, **A. J. Rueten-Budde**, V. Ho, P. Dijkstra, I. C. van der Geest, J. A. Bramer, G. R. Schaap, P. C. Jutte, H. B. Schreuder, J. Ploegmakers, M. Fiocco, and M. van de Sande. Incidence, outcomes and prognostic factors during 25 years of treatment of chondrosarcomas. *Surg Oncol*, 27(3):402–408, 2018.
- [156] C. van Walraven and F. A. McAlister. Competing risk bias was common in kaplan-meier risk estimates published in prominent medical journals. *J Clin Epidemiol*, 69:170–173, 2016.
- [157] J. Vaupel, K. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.
- [158] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenberg. The strengthening the reporting of observational studies in epidemiology (strobe) statement: guidelines for reporting observational studies. *J Clin Epidemiol*, 61(4):344–349, 2008.
- [159] A. E. Wahlquist, L. N. Muhammad, T. L. Herbert, V. Ramakrishnan, and P. J. Nietert. Dissemination of novel biostatistics methods: impact of programming code availability and other characteristics on article citations. *PLOS ONE*, 13(8):1–12, 2018.
- [160] J. Werier, X. Yao, J.-M. Caudrelier, G. Di Primio, M. Ghert, A. A. Gupta, R. Kandel, and S. Verma. Evidence-based guideline recommendations on treatment strategies for localized ewing’s sarcoma of bone following neo-adjuvant chemotherapy. *Surg Oncol*, 25(2):92–97, 2016.
- [161] A. Wienke, K. Christensen, N. Holm, and A. Yashin. Heritability of death from respiratory diseases: an analysis of danish twin survival data using a correlated frailty model. *Stud Health Technol Inform*, 77(77):407–411, 2000.
- [162] A. Wienke, K. Christensen, A. Skytthe, and A. Yashin. Genetic analysis of cause of death in a mixture model of bivariate lifetime data. *Stat Modelling*, 2:89–102, 2002.
- [163] A. Wienke, N. Holm, A. Skytthe, and A. Yashin. The heritability of mortality due to heart diseases: a correlated frailty model applied to danish twins. *Twin Res*, 4(4):266–274, 2001.
- [164] J. Willeumier, M. Fiocco, R. Nout, S. Dijkstra, W. Aston, R. Pollock, H. Hartgrink, J. Bovée, and M. van de Sande. High-grade soft tissue sarcomas of the extremities: surgical margins influence only local recurrence not overall survival. *Int Orthop*, 39(5):935–941, 2015.

-
- [165] M. Wolbers, M. T. Koller, J. C. Witteman, and E. W. Steyerberg. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiol*, 20(4):555–561, 2009.
- [166] A. Yashin, J. Vaupel, and I. Iachine. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Math Popul Stud*, 5(2):145–159, 1995.
- [167] G. Young and R. Smith. *Essentials of Statistical Inference*. Cambridge University Press, 2005.
- [168] B. Yu, J. Saczynski, and L. Launer. Multiple imputation for estimating the risk of developing dementia and its impact on survival. *Biom J*, 52(5):616–627, 2010.
- [169] G. Zagars, M. Ballo, P. Pisters, R. Pollock, S. Patel, R. Benjamin, and H. Evans. Prognostic factors for patients with localized soft-tissue sarcoma treated with conservation surgery and radiation therapy. *Cancer*, 97(10):2530–2543, 2003.
- [170] G. K. Zagars, M. T. Ballo, P. W. Pisters, R. E. Pollock, S. R. Patel, and R. S. Benjamin. Preoperative vs. postoperative radiation therapy for soft tissue sarcoma: a retrospective comparative evaluation of disease outcome. *Int J Radiat Oncol Biol Phys*, 56(2):482–488, 2003.
- [171] Y. Zheng and P. Heagerty. Prospective accuracy for longitudinal markers. *Biometrics*, 63(2):332–341, 2007.
- [172] B. Zhou, J. Fine, A. Latouche, and M. Labopin. Competing risks regression for clustered data. *Biostatistics*, 13(3):371–383, 2012.

Summary

Statistical analysis aims to find data based answers to important research questions in a variety of research areas. The Field of statistics called survival analysis is where the topics of this thesis find their place. Survival analysis deals with life-time data. In this type of data the time from a specific starting point until an event of interest occurs are recorded. In medical research for example, time from diagnosis of disease until death could be studied. What characterizes life-time data, also called survival data, is that it is generally incomplete. Some individuals in the data might not have experienced the event of interest at the end of the study period or have dropped out of the study before the event has occurred. These data are called *right-censored*. The event time is unknown, it is known however, that the event had not occurred before the last observation time. To handle this particular type of missing data, and other similar types, special methodology is necessary summarized under the term survival analysis.

Survival analysis is used by clinicians to identify risk factors associated with the occurrence of a clinical event of interest. For example in cancer research, clinicians use survival models to investigate if a patient's age, sex, tumor size, and other clinically relevant variables are associated to the risk of death. To describe the evolution of disease complex mathematical models are required. Patients may experience several disease related events in different orders. *Multi-state models* can be applied in such context. Another extension of survival models is to add a random effect, also called *frailty*. Frailty terms are used to model unobserved covariates which might have an effect on the event of interest. In all studies not all relevant patient or disease characteristics can be collected and therefore the survival model is incomplete. Random effects quantify the so called unobserved heterogeneity resulting from an incomplete model.

Survival models may be used to investigate the effect of risk factors on clinical events of interest and to predict survival probabilities. Such predictions inform both patients and clinicians of a patient's prognosis and may help in the shared decision making process. Prediction models are available for a variety of diseases and there is a demand for more and more sophisticated models. Ordinary prediction models are often limited to a single prediction time point. This means that predictions can only be made at a particular time, such as at time of diagnosis of disease. When a patient comes back for a follow-up visit, such models are not able to provide accurate predictions. A patient may experience disease related events over time which are not taken into account by a model that considers only risk factors known at diagnosis or at start of treatment. *Dynamic prediction* models provide updated predictions from different time points during follow-up. They are able to include updated information

as it becomes available. A simple idea to create dynamic prediction models is through the *landmarking* approach. Predictions are made from a chosen landmark time point by using a subset of the data consisting of patients still alive at that time. Multiple landmark times can be chosen to make predictions from different time points during follow-up.

The main objective of this thesis was to develop clinically relevant survival models for patients with high-grade soft tissue sarcoma of the extremities, in particular the development and validation of prediction models for use in clinical practice. The interdisciplinary collaboration between the Mathematical Institute of Leiden University and the Leiden University Medical Center resulted in important contributions to the care of soft tissue sarcoma patients [2, 4, 5].

In Chapter 1 basic concepts of survival analysis are introduced as well as more complex models that are used in this thesis. After a short introduction of general concepts, such as the hazard and survival function, frailty models are discussed which add random effects to a survival model. Later on, the simple one end-point survival model for a single event of interest is extended to multiple end-points by introducing competing risks models. More complicated event structures are described thereafter using multi-state models, in which transitioning states where an individual can move through are allowed. Next, dynamic prediction models are introduced as well as measures of discrimination that assess the predictive accuracy of survival prediction models. Some information about the motivating soft tissue sarcoma data set are given. Finally, the developed prediction tool is discussed. An outline of this thesis ends the chapter.

In Chapter 2 a novel frailty model for multi-center data with two competing events is proposed. In practice not all relevant covariates to explain the variance of event times between subjects can be collected. Random effects, called frailty, quantify the unobserved heterogeneity resulting from an incomplete model. Frailty variables that are shared by individuals who were treated in the same hospital are used to model unobserved heterogeneity on the hospital level; they could be interpreted as the "hospital effect" on the competing events. The patients treated in some hospitals may, corrected for covariates, live longer than those treated in other hospitals. This "hospital effect" may be an interest of study. The novelty of the proposed frailty model lies in the construction of the frailty variables. Two frailty variables, one for each competing event, are constructed from three independent gamma distributed frailty components. Each frailty is the sum of two frailty components, a cause-specific and a shared frailty component. This allows for the two frailties to be correlated. The model is estimated using the expectation-maximization algorithm which additionally provides empirical Bayes estimates for each hospital's frailties.

In Chapter 3 the effect of interval censoring is studied on the predicted accuracy of a binary disease marker. The motivation comes from cancer care. After surgery a patient is regularly screened for local recurrence and distant metastasis. Once a recurrence is diagnosed, however, it is only known that it occurred between the last negative and the first positive screening. Additionally, if a patient dies after a negative recurrence screening, then it is unknown whether he developed a recurrence between

the last screening and death. The predictive value of this time-dependent recurrence variable can be summarized by time-specific Area Under the receiver operating characteristics Curve (AUC) measures. The effect that ignoring the interval-censored nature of the observation time has on the time-specific AUC in both incident/dynamic and cumulative/dynamic definition is studied through simulations. AUC estimates derived from different methods for fitting two types of models are compared: the Cox model with time-dependent covariate, which ignores interval-censoring and the illness-death model for interval-censored data.

Chapter 4 is the first in a series of publications based on the growing soft tissue sarcoma data set. A data set of 687 patients with high-grade soft tissue sarcoma of the extremities treated surgically was collected from 4 international tertiary centers. The effect of risk factors on local recurrence and distant metastasis/death was studied using a 3-state multi-state model. Multi-state models describe the evolution of the disease close to reality and allow detailed insights into the effect of risk factors on disease progression. After surgery a patient starts in the starting state "alive without evidence of disease", he can then move to the local recurrence state and subsequently to the distant metastasis/death state or move to distant metastasis/death directly. For each of the three transitions the effect of risk factors was studied allowing for the effects to differ between transitions. Of particular interest was the effect of surgical margin. Surgical margin describes the amount of healthy tissue surrounding the tumor that is dissected during tumor removal surgery. The association with survival and local recurrence was of great interest for clinicians as it impacts the functional outcome after surgery.

Chapter 5 is the continuation of the soft tissue sarcoma project, with a data set of 766 patients collected from 5 international tertiary centers. The motivation came from the need of clinicians for an easy to use prediction tool for patients with soft tissue sarcoma. Two prediction models one for survival and one for the probability of local recurrence were developed using Cox and Fine and Gray's methodology. The survival model is a simple one end-point model, the model for local recurrence however, needs to consider the competing risk of death. The models predict the probability of surviving 3, 5, and 10 years as well as the probability of developing a local recurrence within 3, 5, and 10 years from time of surgery respectively. The advantage of using Fine and Gray's model for competing risks to model covariate effects on the probability of developing local recurrence is that estimated regression coefficients are more intuitive to interpret for clinicians compared to the cause-specific hazards model. The prediction models were implemented in the PERSARC mobile application to be used by clinicians to improve patient care [4, 5]. An internal validation considering calibration plots and the C-index demonstrated good calibration and discrimination of the prediction models.

In Chapter 6 a dynamic prediction model based on the growing soft tissue sarcoma data set was developed. Data of 2232 soft tissue sarcoma patients was collected from a total of 14 international tertiary centers. The aim was to develop a prediction tool able to make updated survival predictions for patients during follow-up. After surgery a patient has scheduled follow-up visits to monitor him and to screen for

adverse events. Events like local recurrence and distant metastasis affect the future prognosis. Also the fact that a patient survived a length of time after surgery may give an insight into the future prognosis. This requires dynamic predictions for a patient to be updated over time. For this purpose a landmark supermodel was used to predict the probability of surviving an additional 5-years from different prediction time points during follow-up. Local recurrence and distant metastasis, are used to update predictions over time and covariates were investigated for time-varying effects. The model was internally validated.

In Chapter 7 the previously developed dynamic prediction model for soft tissue sarcoma patients is updated and externally validated. The updated model is based on 3826 patients collected from 17 international tertiary centers and a randomized controlled trial. Data for external validation consisted of 1111 patients from a single tertiary center. The updated dynamic prediction model now includes grade as additional covariate in the model. This important covariate was initially omitted because the previously collected data contained mainly grade III patients. During this research, the data set has been significantly augmented and now includes a large cohort of grade II patients. A successful external validation showed that the model was able to adequately predict the probability of surviving an additional 5-years from different prediction time points during follow-up. The model is implemented in the updated PERSARC mobile application [4, 5].

In Chapter 8 a multi-state model was developed for 982 Ewing sarcoma patients that were treated surgically according to the EURO-E.W.I.N.G99 protocol. The starting time of analysis is the time of surgery, from which a patient can move to different states corresponding to disease progression. Adverse events considered in the multi-state model were local recurrence, distant metastasis of the lungs, distant metastasis at other locations, and death. The effect of risk factors was studied on the transitions between disease states and the effect was allowed to differ between transitions. A particular interest lay in the effects of surgical margins, histological response, and radiotherapy treatment.

Samenvatting

Statistische analyse streeft ernaar om antwoorden te vinden op belangrijke onderzoeksvragen in verscheidene onderzoeksgebieden door gebruik te maken van data. De onderwerpen in dit proefschrift behoren tot het statistisch gebied *overlevingsanalyse*. Overlevingsanalyse houdt zich bezig met levensduurdata. Voor dit type data wordt de tijd van een specifiek beginpunt tot een gebeurtenis waar de interesse naar uitgaat geregistreerd. Bijvoorbeeld in medisch onderzoek kan men de tijd van diagnose van een ziekte tot overlijden bestuderen. Wat karakteristiek is voor levensduurdata, ook wel overlevingsdata genoemd, is dat het vaak onvolledig is. Voor sommige individuen in de data heeft de gebeurtenis nog niet plaatsgevonden aan het einde van de studie of zijn ze uit de studie weggevallen voordat de gebeurtenis heeft plaatsgevonden. Deze data wordt *rechtsgecensureerde data* genoemd. Hoewel het moment van de gebeurtenis niet bekend is, is het wel bekend dat de gebeurtenis nog niet heeft plaatsgevonden bij de laatste observatie. Om om te gaan met dit soort missende data, zijn speciale methoden nodig die zich laten samenvatten onder de noemer overlevingsanalyse.

Overlevingsanalyse wordt gebruikt door medici om risicofactoren te identificeren die worden geassocieerd met het plaatsvinden van een klinische gebeurtenis waar de interesse naar uitgaat. Bijvoorbeeld in kankeronderzoek waar medici overlevingsmodellen gebruiken om te onderzoeken of een patiënt zijn leeftijd, sex, tumor grootte en andere klinisch relevante variabelen geassocieerd kunnen worden met de kans op overlijden. Om het beloop van een ziekte te beschrijven zijn complexe wiskundige modellen nodig. Patiënten kunnen verschillende, met de ziekte geassocieerde, gebeurtenissen ervaren in verschillende volgorden. *Multi-state-modellen* kunnen worden toegepast in deze context. Een andere uitbreiding van overlevingsmodellen is om een willekeurig effect toe te voegen genaamd *fragiliteit*. Fragiliteitstermen worden gebruikt om niet geobserveerde covariaten te modelleren die misschien een effect hebben op de gebeurtenis waar de interesse naar uitgaat. Namelijk, niet alle relevante karakteristieken van de patiënt of ziekte worden in alle studies verzameld. Daarom is het overlevingsmodel incompleet. Willekeurige effecten kwantificeren de zogenoemde ongeobserveerde heterogeniteit als resultaat van een incompleet model.

Overlevingsmodellen kunnen worden gebruikt om de effecten van risicofactoren op klinische gebeurtenissen te onderzoeken en om te voorspellen wat de overlevingskansen zijn. Dergelijke voorspellingen geven zowel de patiënten als de medici informatie over de prognose en kunnen helpen in de gezamenlijke besluitvorming. Voorspellingsmodellen zijn bruikbaar voor een verscheidenheid aan ziektes en er is vraag naar steeds geavanceerdere modellen. Gebruikelijke voorspellingsmodellen zijn vaak beperkt tot een enkel voorspellingsmoment. Dit betekent dat voorspellingen alleen op één bepaald tijdstip, zoals het moment van diagnose van de ziekte, gemaakt kunnen

worden. Wanneer een patiënt terugkomt voor een vervolgbezoek, kunnen zulke modellen geen nauwkeurige voorspelling geven. Een patiënt kan over tijd ziekte-gerelateerde gebeurtenissen ervaren waarmee geen rekening wordt gehouden wanneer een model alleen risicofactoren meeneemt die gemeten zijn op het moment van diagnose of aan de start van de behandeling. *Dynamische voorspellingsmodellen* verschaffen bijgewerkte voorspellingen vanaf verschillende tijdstippen gedurende een vervolgbehandeling. Ze kunnen dus informatie gebruiken zodra het beschikbaar is. Een simpele manier om een dynamisch voorspellingsmodel te creëren is door middel van de *herkenningspunten* aanpak. Voorspellingen worden gemaakt vanaf een gekozen herkenningspunttijdstip door gebruikt te maken van het deel van de data bestaande uit patiënten die op dat tijdstip nog in leven zijn. Meerdere herkenningspunttijdstippen kunnen worden gekozen om voorspellingen te maken vanaf verschillende tijdstippen tijdens een vervolgbehandeling.

Het hoofddoel van dit proefschrift is om klinisch relevante overlevingsmodellen te ontwikkelen voor patiënten met een hoogwaardig zacht weefselsacroom op een ledemaat met in het bijzonder de ontwikkeling en valorisatie van overlevingsmodellen voor gebruik in de klinische praktijk. De interdisciplinaire samenwerking tussen het Mathematisch Instituut van Universiteit Leiden en het Leids Universitair Medisch Centrum heeft geleid tot belangrijke bijdragen in de verzorging van patiënten met zachte weefselsacroma's [2, 4, 5].

In hoofdstuk 1 worden de basisconcepten van overlevingsanalyse geïntroduceerd evenals de complexere modellen die worden gebruikt in dit proefschrift. Na een korte introductie van algemene concepten, zoals de hazard- en overlevingsfunctie, komen fragiliteitsmodellen aan de orde die willekeurige effecten toevoegen aan een overlevingsmodel. Later in het hoofdstuk wordt een simpel overlevingsmodel met een enkel eindpunt uitgebreid naar een overlevingsmodel met meerdere eindpunten door concurrerende-risicomodellen te introduceren. Daarna worden ingewikkeldere gebeurtenisstructuren beschreven die gebruikmaken van multi-state-modellen waarin transitietoestanden worden toegelaten waar een individu zich doorheen kan bewegen. Vervolgens worden zowel dynamische voorspellingmodellen geïntroduceerd als maatstaven van onderscheiding die de nauwkeurigheid van voorspellingen van overlevingsvoorspellingmodellen schatten. Er wordt informatie gegeven over de zacht weefselsacroom dataset wat de motivatie was voor dit onderzoek. De ontwikkelde voorspellingstool wordt ook besproken. Het hoofdstuk sluit af met een uitleg over de hoofdlijnen van dit proefschrift.

In hoofdstuk 2 wordt een vernieuwend fragiliteitsmodel voor meerdere-centra-data met twee concurrerende gebeurtenissen voorgelegd. In de praktijk kunnen niet alle relevante covariaten worden verzameld die de variantie tussen de tijdstippen waarop de gebeurtenissen van patiënten plaatsvinden verklaren. Willekeurige effecten, genaamd fragiliteit, kwantificeren de ongeobserveerde heterogeniteit die veroorzaakt wordt door een incompleet model.

Voor individuen die in hetzelfde ziekenhuis worden behandeld worden dezelfde fragiliteitsvariabelen gebruikt om ongeobserveerde heterogeniteit op ziekenhuisniveau te modelleren; ze kunnen worden geïnterpreteerd als het "ziekenhuiseffect" voor de

concurrerende gebeurtenissen. Patiënten die worden behandeld in bepaalde ziekenhuizen zouden, gecorrigeerd voor de covariaten, langer kunnen leven dan diegene die worden behandeld in een ander ziekenhuis. Dit "ziekenhuseffect" zou interessant kunnen zijn om te bestuderen. De vernieuwing van het voorgestelde fragiliteitsmodel zit in de constructie van de fragiliteitsvariabelen. Twee fragiliteitsvariabelen, één voor elke concurrerende gebeurtenis, worden geconstrueerd vanuit drie onafhankelijke gamma-verdeelde fragiliteitscomponenten. Elke fragiliteit is de som van twee fragiliteitscomponenten, één oorzakspecifieke en één gedeeld fragiliteitscomponent. Dit geeft de mogelijkheid voor de twee fragiliteiten om gecorreleerd te zijn. Het model wordt geschat door gebruik te maken van het verwachting-maximalisatie-algoritme wat ook empirische Bayes-schatters geeft voor de fragiliteiten van elk ziekenhuis.

In hoofdstuk 3 wordt het effect van intervalcensurering op de voorspelde nauwkeurigheid van een binaire ziektemarker bestudeerd. Dit is gemotiveerd vanuit de kankerzorg. Na een operatie wordt een patiënt regelmatig onderzocht voor lokale terugkeer en metastase op afstand. Wanneer er terugkeer wordt gediagnosticeerd is het alleen bekend dat het teruggekomen is tussen de laatste negatieve en de eerste positieve test. Daarbij komt dat wanneer een patiënt komt te overlijden na een negatieve herhalingstest het onbekend is of hij een herhaling heeft ontwikkeld tussen de laatste test en het moment van overlijden. De voorspellende waarde van deze tijdafhankelijke variable over terugkeer kan worden samengevat door tijdspecifieke *Area Under the receiver operating characteristics Curve* (AUC) maten. Wat het effect van het negeren van de intervalgecensureerde aard van de observatietijd heeft op de tijdspecifieke AUC in zowel de incidentele/dynamische en cumulatieve/dynamische definitie wordt bestudeerd door middel van simulaties. AUC schatters worden vergeleken die zijn afgeleid van verschillende methoden om twee soorten modellen te berekenen: het Cox model met tijdafhankelijke covariaten welke interval censuur negeert en het ziekte-dood model voor intervalgecensureerde data.

Hoofdstuk 4 is de eerste in een serie van publicaties gebaseerd op de groeiende zacht weefselsarcom dataset. Een dataset, bestaande uit data van 687 patiënten met hoogwaardige zacht weefselsarcoma's op ledematen die chirurgisch behandeld zijn, is verzameld door 4 internationale tertiaire centra. Het effect van risicofactoren op lokale terugkeer en verre metastasis/dood is bestudeerd met behulp van van een 3-toestanden multi-state-model. Multi-state-modellen beschrijven het beloop van de ziekte realistisch en geven gedetailleerd inzicht in het effect van de risicofactoren op dit beloop. Na de operatie begint een patiënt in de "levend zonder teken van ziekte-toestand. Vervolgens kan hij bewegen naar de "lokale terugkeer-toestand en hierop volgend naar de "verre metastasis/dood-toestand of direct naar de "metastasis/dood-toestand. Voor elk van de drie overgangen worden de effecten van de risicofactoren bestudeerd waarbij deze effecten verschillend mogen zijn bij elke overgang. In het bijzonder ging er interesse uit naar het effect van chirurgische marge. Chirurgische marge beschrijft de hoeveelheid gezond weefsel rondom de tumor dat weggehaald is tijdens de verwijderingsoperatie. De associatie met overleving en lokale terugkeer was erg interessant voor klinici, omdat het invloed heeft op de functionele uitkomst van een operatie.

Hoofdstuk 5 is de voortzetting van het zacht weefselsarcoma's project met een dataset van 766 patiënten wat verzameld is door 5 internationale tertiaire centra. De motivatie kwam van de vraag van klinici om een voorspellingstool voor patiënten met een zacht weefselsarcoma te maken die makkelijk is in gebruik. Twee voorspellingsmodellen, één voor overleving en één voor de kans op lokale terugkeer, zijn ontwikkeld met behulp van Cox en Fine en Gray's methodologie. Het overlevingsmodel is een simpel enkel eindpunt model. In het model voor lokale terugkeer moet echter ook het concurrerende risico van doodgaan meegenomen worden. De modellen voorspellen zowel de kans op het overleven voor 3, 5 en 10 jaar als de kans op lokale terugkeer binnen 3, 5 en 10 jaar vanaf het moment van opereren. Het voordeel van het gebruik van Fine en Gray's model voor concurrerende risico's voor het modelleren van de effecten van covariaten op de kans van de ontwikkeling van lokale terugkeer is dat de geschatte regressiecoëfficiënten intuïtiever zijn om te interpreteren voor klinici in vergelijking met het oorzaak-specifieke hazardmodel. De voorspellingsmodellen werden geïmplementeerd in de PERSARC mobiele applicatie wat gebruikt wordt door klinici om de zorg voor patiënten te verbeteren [4, 5]. Een interne validatie door calibratieplots en de C-index laten een goede calibratie en onderscheidend vermogen zien van de voorspellingsmodellen.

In hoofdstuk 6 wordt een dynamisch voorspellingsmodel gebaseerd op de groeiende zacht weefselsarcoma dataset ontwikkeld. De data van 2232 zacht weefselsarcoma patiënten werd verzameld door 14 internationale tertiaire centra. Het doel was om een voorspellingstool te ontwikkelen die het mogelijk maakt om bijgewerkte overlevingsvoorspellingen te maken voor patiënten tijdens een vervolgbehandeling. Na een operatie werd een vervolgbezoek ingepland om de patiënt te monitoren en te testen op bijwerkingen. Gebeurtenissen als lokale terugkeer en metastase op afstand hebben invloed op de prognose. Het feit dat een patiënt nog een bepaalde tijd leeft na een operatie kan inzicht geven in de prognose. Dit vergt dynamische voorspellingen voor de patiënt om de prognose over tijd aan te passen. Voor dit doeleinde is een herkenningspunt supermodel gebruikt om een voorspelling te geven van de kans op een extra 5 jaar overleven vanaf verschillende voorspellingstijdpunten tijdens een vervolgbehandeling. Lokale terugkeer en metastase op afstand worden gebruikt om voorspellingen over tijd bij te werken en covariaten werden onderzocht voor tijdsveranderlijke effecten. Het model is intern gevalideerd.

In hoofdstuk 7 wordt het eerder ontwikkelde dynamische voorspellingsmodel voor zacht weefselsarcoma patiënten aangepast en extern gevalideerd. Het aangepaste model is gebaseerd op data van 3826 patiënten wat verzameld is door 17 internationale tertiaire centra en een gerandomiseerde gecontroleerde studie. Data voor externe validatie bestond uit 1111 patiënten van een enkel tertiair centrum. Het aangepaste dynamische voorspellingsmodel bevat nu graad als extra covariaat in het model. Deze belangrijke covariaat was aanvankelijk weggelaten, omdat de eerder verzamelde data vooral graad III patiënten bevatte. Tijdens dit onderzoek is de dataset significant aangevuld en bevat het een groot cohort van graad II patiënten. Een succesvolle externe validatie laat zien dat het model in staat was om adequaat de kans op een extra 5 jaar overleven te voorspellen vanaf verschillende voorspellingstijdpunten tijdens vervolgbehandeling. Het model is geïmplementeerd in de aangepaste PERSARC

mobiele applicatie [4, 5].

In hoofdstuk 8 is een multi-state-model ontwikkeld voor 982 Ewing sarcoom patiënten die chirurgisch behandeld zijn volgens het EURO-E.W.I.N.G99 protocol. De starttijd van de analyse is het moment van de operatie waar vanuit de patiënt zich kan verplaatsten naar verschillende toestanden corresponderende met de ontwikkeling van de ziekte. Bijwerkingen die zijn meegenomen in het multi-state-model waren lokale terugkeer, metastasis op afstand van de longen, metastasis op afstand van andere locaties en overlijden. Het effect van risicofactoren op de transitie tussen zieke toestanden werd bestudeerd en daarbij mogen deze effecten verschillend zijn bij elke overgang. In het bijzonder is men geïnteresseerd in het effect van chirurgische marge, histologische reactie en radiotherapie.

List of Publications

- 1) **A. J. Rueten-Budde**, H. Putter, and M. Fiocco. Investigating hospital heterogeneity with a competing risks frailty model. *Stat Med*, 38(2):269–288, 2018
- 2) **A. J. Rueten-Budde**, H. Putter, and M. Fiocco. Assessment of predictive accuracy of an intermittently observed binary time-dependent marker. *Submitted*
- 3) **A. J. Rueten-Budde**, J. J. Willeumier, L. M. Jeys, M. Laitinen, R. Pollock, W. Aston, P. D. S. Dijkstra, P. C. Ferguson, A. M. Griffin, J. S. Wunder, M. Fiocco, and M. A. J. van de Sande. Individualised risk assessment for local recurrence and distant metastases in a retrospective transatlantic cohort of 687 patients with high-grade soft tissue sarcomas of the extremities: a multistate model. *BMJ Open*, 7(2):e012930, 2017
- 4) **A. J. Rueten-Budde**, V. M. van Praag, L. M. Jeys, M. K. Laitinen, R. Pollock, W. Aston, J. A. van der Hage, P. S. Dijkstra, P. C. Ferguson, A. M. Griffin, J. J. Willeumier, J. S. Wunder, M. A. van de Sande, and M. Fiocco. A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: personalised sarcoma care (persarc). *Eur J Cancer*, 83:313–323, 2017
- 5) **A. J. Rueten-Budde**, V. van Praag, PERSARC studygroup, M. van de Sande, and M. Fiocco. Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma. *Surg Oncol*, 27(4):695–701, 2018
- 6) **A. J. Rueten-Budde**, M. van de Sande, V. van Praag, PERSARC studygroup, and M. Fiocco. External validation and adaptation of a dynamic prediction model for patients with high-grade extremity soft tissue sarcoma. *Submitted*
- 7) **A. J. Rueten-Budde**, S. Bosma, C. Lancia, A. Ranft, U. Dirksen, Krol, H. Gelderblom, M. van de Sande, P. Dijkstra, and M. Fiocco. Individual risk evaluation for local recurrence and distant metastasis in ewing sarcoma: a multistate model. *Pediatr Blood Cancer*, 66(e27943):doi: 10.1002/pbc.27943, 2019
- 8) V. van Praag, **A. J. Rueten-Budde**, V. Ho, P. Dijkstra, I. C. van der Geest, J. A. Bramer, G. R. Schaap, P. C. Jutte, H. B. Schreuder, J. Ploegmakers, M. Fiocco, and M. van de Sande. Incidence, outcomes and prognostic factors during 25 years of treatment of chondrosarcomas. *Surg Oncol*, 27(3):402–408, 2018
- 9) M. J. L. Mastboom, E. Palmerini, F. G. M. Verspoor, **A. J. Rueten-Budde**, S. Stacchiotti, E. L. Staals, G. R. Schaap, P. C. Jutte, W. Aston, H. Gelderblom, A. Leithner, D. Dammerer, A. Takeuchi, Q. Thio, X. Niu, J. S. Wunder, TGCT Study Group, and M. A. J. van de Sande. Surgical outcomes of patients with

- diffuse-type tenosynovial giant-cell tumours: an international, retrospective, cohort study. *Lancet Oncol*, 20(6):877–886, 2019
- 10) M. Mastboom, E. Staals, F. Verspoor, **A. J. Rueten-Budde**, S. Stacchiotti, E. Palmerini, G. Schaap, P. Jutte, W. Aston, A. Leithner, D. Dammerer, A. Takeuchi, Q. Thio, X. Niu, J. Wunder, TGCT study group, and M. van de Sande. Surgical treatment of localized-type tenosynovial giant cell tumours of large joints. *J Bone Joint Surg*, 101(14):1309–1318, 2019
 - 11) S. Bosma, C. Lancia, **A. J. Rueten-Budde**, A. Ranft, H. Gelderblom, M. Fiocco, M. van de Sande, P. Dijkstra, and U. Dirksen. Easy-to-use clinical tool for survival estimation in ewing sarcoma at diagnosis and after surgery. *Sci Rep*, 9(11000), 2019
 - 12) E. Schutgens, P. Picci, D. Baumhoer, R. Pollock, J. Bovee, P. Hogendoorn, P. Dijkstra, **A. J. Rueten-Budde**, P. Jutte, F. Traub, A. Leithner, P. Tunn, P. Funovics, G. Sys, M. Julian, G. Schaap, H. Dürr, J. Harges, J. Healy, R. Capanna, D. Biau, A. Gomez-Brouchet, J. Wunder, T. Cosker, M. Laitinen, X. Niu, V. Kostiuik, Adamantinoma research group, and M. van de Sande. Surgical outcome and oncological survival of osteofibrous dysplasia-like- and classic adamantinoma: an international multicenter study of 318 cases. *Submitted*
 - 13) **A. J. Rueten-Budde**, C. Liu, A. Ranft, U. Dirksen, H. Gelderblom, and M. Fiocco. Dynamic prediction of overall survival for patients with ewing sarcoma. *Submitted*

Curriculum Vitae

Anja Juana Rueten-Budde was born on August 12th 1990 in Bremen, Germany. In 2013 she obtained a BSc degree in Mathematics at the University of Bremen. In 2013 she moved to the Netherlands to continue her studies at Leiden University where she completed the Master programme ‘Statistical Science for the Life & Behavioural Sciences’ in 2015.

During her Master studies she completed an internship collaborating with the Department of Orthopaedic Surgery of the Leiden University Medical Center under the supervision of dr. Marta Fiocco, which made the start of her research of soft tissue sarcoma. The results of this intership assisted in obtaining a grant from the KWF to fund part of her PhD research which she started in 2016 under the supervision of dr. Marta Fiocco at Leiden University. Her work focused on developing accurate prediction models for patients with high-grade extremity soft tissue sarcoma. Results of this research are presented in this thesis and were implemented in the PERSARC mobile application for use in clinical practice. The impact of the PERSARC application is currently studied by researchers at the Leiden University Medical Center. Their research is funded by a grant from the KWF.

She has been teaching assistant in various courses and was lecturer of Statistics at Leiden University College (The Hague). In the last year of her PhD research she was lecturer for the Survival Analysis course from the Statistical Science master programme. She has been involved in a number of consultations at the Leiden University Medical Center in which she helped clinicians with their data analysis. During her PhD research time she presented her research at conferences in the Netherlands, United Kingdom, Finland, Canada, and Denmark. She was co-organizer of the conference Survival Analysis for Junior Researchers 2018 in Leiden which was the first time that the conference took place outside of the United Kingdom.

In her free time she enjoys to knit and was part of the Mathematical Institute Knitting Club. In 2019 she published her first knitting pattern under the name ‘Spherical fractal scarf’ which was a collaboration with Niels Langeveld. At the moment of writing she is looking for a position as a statistical consultant.

Acknowledgements

I would like to take this opportunity to thank the people that were part of my PhD student life.

First of all, I would like to thank my supervisor dr. Marta Fiocco who guided me through the world of clinical biostatistics. She has supervised my internship, master thesis and my PhD research project. Marta, thank you so much for your guidance throughout my study and research time. Because of your constant support I was able to complete my PhD project in time which I am very grateful for. I learned from you how to do research and write articles but also how to communicate with clinicians. You invested a lot of time in reviewing my work and gave me valuable feedback in a timely manner. Apart from being an outstanding supervisor we have a very good personal relationship which I value a lot.

Secondly, I would like to thank Prof. dr. Hein Putter. Hein, I valued our collaboration that resulted in two chapters of this thesis. Apart from being a known expert in your field you are a humble person who is a pleasure to work with. Even outside of our direct collaboration you have been available to discuss and answer my questions. Thank you.

I thank Prof. dr. Eric Eliel for being of great help with the graduation procedure.

I would like to extend my sincere gratitude to Michiel van de Sande and Veroniek van Praag who formed collaborations between international sarcoma centers to make our research possible. Thank you Michiel, for the networking which resulted in the creation of the PERSARC studygroup. Thank you Veroniek, for collecting and combining all the different data sources and making sure that we have a reliable data set to work with. I am grateful to the members of the PERSARC studygroup for contributing their share of data to the project.

I thank Prof. dr. Jacqueline Meulman who was the main sponsor of the SAFJR 2018 conference which I co-organized. Thank you Jacqueline.

I am grateful for all my colleagues at the Mathematical Institute who created a beautiful atmosphere for work and life. We shared coffee breaks, did sports, crafted together and became friends in this gezellige institute. Thank you Sanne and Stéphanie for organizing the SAFJR 2018 conference with me, it was great fun. Thank you Sanne, also for helping me with the Dutch summary of this thesis. A special thanks goes to my office mates Carlo (we never forgot you), Magnus, Dirk and Lasse. I had a great time with you, I hope you keep dreaming big.

I thank my in-laws who welcomed me whole heartedly into their family. Thank you for making me feel right at home with you.

Thanks also to my family who allowed me to find my own way, even when it lead me abroad. I want to thank my parents who raised me to be independent which gave

Acknowledgements

me so many opportunities. I am grateful for my siblings Inge, Frank, and Pablo. Talking with you always enlightens my day.

Lastly, I want to thank my fantastic husband Niels. Apart from providing the beautiful artwork for the cover of this dissertation and translating the summary for me you are a pleasure to be around. Life with you is so easy and beautiful.