

Internship Report

Role of Surgical Margin on Disease Progression
of High-Grade Soft Tissue Sarcoma of the
Extremities

Report by: Anja Juana Rüten-Budde
Student number: S1433792
Supervisor: Marta Fiocco
Leiden University Medical Center and
Mathematical Institute Leiden University
University: Leiden University

Contents

Introduction	2
Goal	2
Medical background	2
Research background	2
Choice of covariates	3
Patients and methods	3
Data	4
Statistical analysis	4
Coxmodel with time-dependent covariates	4
Multi-state analysis	6
Test Assumptions	10
Limb sparing operation	10
Results	11
Cox analysis	11
Multi-state analysis	14
Prediction Multi-state	14
Conclusion and discussion	17
Website to estimate patient specific probabilities of LR and OS at fixed time points	18
Motivation	18
Models	18
Time-fixed Cox model	18
Competing risk model	18
Implementation	21
Discussion	23
References	23

Introduction

Goal

The goal of this study was to evaluate the association of prognostic factors, with special interest on surgical margin, on local recurrence (LR), distant metastasis (DM) and overall survival (OS) for patients with high-grade soft tissue sarcoma of the extremities. Earlier work has been done by Willeumier et al. (2015) and Xinru Li during her Internship. Since more data was collected a more detailed analysis was possible.

Based on the models proposed in this study an app for clinicians was developed, where 5 and 10 years overall survival, as well as the cumulative incidence of LR at 5 and 10 years can be estimated based on patient specific characteristics.

Medical background

Soft tissue sarcomas are a rare and lifethreatening type of cancer that develops in the soft tissues like fat, muscle, nerves, fibrous tissues, bloodvessels, or deep skin tissues. They occur in a variety of subtypes and in every part of the body. Given this heterogeneity of the patient population combined with a low incidence, it is a challenge to make inference on prognostic factors.

Surgery is the standard treatment of primary soft tissue sarcoma. When removing the tumor it is the goal of the surgeon to take out the entire tumor with a rim of normal tissue around it. The smallest distance of the tumor cells to the edge is the margin of interest here. Generally wider margins are preferred, but depending on the location and involvement in critical structures a wide margin can result in a major loss of mobility. Therefore it is important to investigate the influence of surgical margin on OS, to make decisions when planning surgery.

Research background

Several studies have researched the influence of prognostic factors on LR, DM and OS for patients with soft tissue sarcoma in different settings (Willeumier et al. (2015), Maretty-Nielsen et al. (2014), Italiano et al. (2014), Toulmonde et al. (2014) (long-term), O'Donnell et al. (2014), Trovik et al. (2000)).

Positive margins were consistently linked to a higher risk of LR (Willeumier et al. (2015), Maretty-Nielsen et al. (2014), Trovik et al. (2000)). However the relation to DM and OS is not clear.

Apart from margin, several other possible prognostic factors were investigated in previous studies. Some factors such as tumor size and grade have consistently proven

to be of prognostic value for LR, DM and OS (Willeumier et al. (2015) only size, Maretty-Nielsen et al. (2014), Italiano et al. (2014), Toulmonde et al. (2014), Trovik et al. (2000)). Others risk factors were investigated, like age (Maretty-Nielsen et al. (2014), Italiano et al. (2014)), depth (Italiano et al. (2014)), tumor localization (Maretty-Nielsen et al. (2014), only internal trunc), histological-type (Italiano et al. (2014)) and radiotherapy (Maretty-Nielsen et al. (2014)). LR was studied as time-dependent covariate for DM and OS in Willeumier et al. (2015) and found to be an adverse event for OS. In Trovik et al. (2000) LR was associated with an increased risk of metastasis. These studies did investigate different factors and they differ wildly in sample size as well as statistical methods applied, which explains there different findings. Also most studies were performed on a highly heterogenous patient population.

Choice of covariates

When consulting with the clinician, apart from the previously mentioned covariates (age, depth, size, radiotherapy at baseline, margin, localization, type) 2 other possible risk factors emerged. Tumor presentation (primary, whoops, irradiated resection / residue) and a limb sparing operation (amputation, no amputation).

The presence of an adjuvant treatment of chemotherapy was not included as covariate, even though information was available. Clinicians were neither interested nor suspecting an effect and only a small percentage of patients received chemotherapy.

Originally tumor location was given in 6 categories, those were collapsed to form two categories (upper, lower extremity). Size was initially proposed as a categorical covariate but due to proportionality issues was finally included as continuous.

Patients and methods

Data of patients surgically treated in Leiden, Toronto, London and Birmingham was collected, forming a dataset containing a total of 709 patients. Only patients with a high-grade and tumor location in the extremities were included to form a more homogenous patient population. Exclusion criteria were (1) metastatic disease at time of diagnosis, (2) primary presentation with recurrent disease, (3) no operative treatment with curative intent at our centers, and (4) an unknown margin status. Margin status was determined by independent experienced musculoskeletal pathologists defining the closest surgical margin after inking and sectioning of the specimen. The closest identified margin of resection was classified as intralesional (positive margin with tumor cells visible at the inked margin), marginal (negative margin of 2mm or less), or wide (negative margin of more than 2mm).

Data

Two datasets were thoroughly checked and merged regards to the variables of interest. In case of inconsistent data entries, a clinician was consulted and the original information was retrieved. Time to event is rounded to whole months.

The first dataset contained 333 patients from Leiden, London and Toronto. For two patients the date of surgery was missing, for one patient the event time was missing, one patient was censored three days (0 months) after surgery with no events, one patient experienced distant metastasis before surgery and one patient had a missing margin status, which are both exclusion criteria. Those patients were excluded from the analysis, leaving a total of 327 patients.

The second dataset contained 376 patients from Birmingham. For two patients the date of surgery was missing, for one an event time was missing, one patient was censored with no event after two days (0 months), two patients experienced DM before surgery, one LR before surgery, which are both exclusion criteria. Those patients were excluded leaving a total of 370.

One patient died two days after surgery and three experienced DM immediately after surgery (5, 8 and 14 days). They were included with fractioned event times of 0.1, 0.2, 0.3 and 0.5 respectively. This leaves a total of 697 patients in the final dataset. A total of 687 patients are used in the analysis, due to some missing values in the regular covariates. Tumor and patient characteristics are shown in Table 1.

Statistical analysis

To investigate the effect of surgical margin and the other possible prognostic factors on overall survival a multivariate Cox regression analysis with LR and DM as time-dependent covariates, as well as a multi-state analysis was conducted. To control for unmeasured possible differences in study populations, the center of treatment was incorporated in all models as a factor. To test the proportional hazards assumption a p value < 0.01 was considered significant. For covariate effects a p value of < 0.05 was considered significant.

Coxmodel with time-dependent covariates

To analyse LR and DM as time-dependent covariates it was necessary to transform the data into a specific long format. A start and stop variable were created to indicate the intervals a person is at risk, each in a separate row, as well as event indicators for LR, DM and death. The covariates are repeated over the rows of a patient. The following output shows three patients with different event histories. Only some of the covariates are shown.

Table 1: Patients, tumor and treatment characteristics.

Characteristic	Mean/Total	SD/Percentage
Age (continuous)	57.9	19.8
Sex		
Male	389	56.6%
Female	298	43.4%
Tumor presentation		
Primary	555	80.8%
Whoops	109	15.9%
Irradical resection / residue	25	3.3%
Tumor location		
Upper extremity	162	23.6%
Lower extremity	525	76.4%
Tumor size		
<5cm	103	15%
5-9.99cm	273	39.7%
≥ 10 cm	311	45.3%
Tumor size in cm (continuous)	10.0	6.2
Depth		
Deep	531	77.3%
Superficial	115	16.7%
Deep & Superficial	41	6%
Histopathology		
Angiosarcoma	19	2.8%
MPNST	81	11.8%
Myxofibrosarcoma	217	31.6%
Synoviosarcoma	134	19.5%
Sarcoma nos	17	2.5%
Spindle cell sarcoma	165	24.0%
MFH/UPS	54	7.9%
Surgical margin		
Intralesional	114	16.6%
Marginal	325	47.3%
Wide	248	36.1%
Type of surgery		
Amputation	76	11.1%
Limb-sparing	611	88.9%
Radiotherapy		
Yes	513	74.7%
No	174	25.3%

##	id	age	size.cm	RT	margin.class	tstart	tstop	LR	DM	death
## 4	17	86	3.2	yes	0.1-2mm	0	15	0	0	1
## 5	18	31	9.0	yes	0.1-2mm	0	2	0	0	0
## 6	18	31	9.0	yes	0.1-2mm	2	14	0	1	1
## 49	257	29	12.5	yes	intralesional	0	17	0	0	0
## 50	257	29	12.5	yes	intralesional	17	21	0	1	0
## 51	257	29	12.5	yes	intralesional	21	61	1	1	1

Patient 17 experiences neither LR nor DM and dies after 15 months. Patient 18 experiences LR after 2 months and dies after 14 months, without DM. Patient 257 experiences first DM after 17 months, then LR after 21 months and dies after 61 months.

Several models were estimated, a full model containing all possible factors, a reduced model, a full model excluding DM and a reduced model excluding DM.

Since DM is an immediate predecessor of death and for almost all patients experiencing DM death was observed (88%), it might overshadow effects of other risk factors, models excluding DM as a predictor were estimated.

The reduced models are obtained by a stepwise backwards elimination looking at the AIC. Margin is kept in the model as it is the variable of interest.

All models significantly violate the proportional hazards assumption for radiotherapy and are therefore stratified by it. In Table 2 and 3 the estimated hazards ratios and their corresponding 95% confidence intervals of the models are shown.

Multi-state analysis

Multi-state models are a generalization of competing risk models by describing transitions to intermediate events. Through modeling the intermediate events, multi-state models are able to describe the disease/recovery process of patients in more detail. The added value of a multi-state model is, that it allows for different effects of covariates on different transitions. This gives a better insight about risk factors and their effect on different events. The multi-state model used is given in Figure 1. Four states representing disease progression were discerned. After surgery a patient may be alive without disease (AWD), he may experience local recurrence (LR), he may develop distant metastasis (DM) or he might die. The direction of arrows in Figure 1 indicate the transitions between the four states that are logically possible. The time scale used is time since surgery, corresponding to the ‘Clock forward’ approach. In this approach the time refers to the time since the patient entered the initial state AWD. The clock keeps moving forward for the patient, even when intermediate events occur. In comparison the ‘Clock reset’ approach resets the clock to 0 each time the patient enters a new state.

Table 2: Potential prognostic factors in different models: (A) full model, (B) reduced model

Variable	A. P value	HR	.95 CI	B. P value	HR	.95 CI
Age in years (continuous)	7e-15	1.03	1.02-1.04	3e-15	1.03	1.02-1.04
Presentation						
Primary	1					
Whoops	0.759624	1.06	0.74-1.50			
Irradical resection	0.662854	1.19	0.54-2.63			
Location (lower vs. upper extremity)	0.827546	0.97	0.73-1.28			
Size in cm (continuous)	0.000002	1.05	1.03-1.08	3e-08	1.05	1.03-1.07
Depth						
Deep	1					
Superficial	0.595337	0.91	0.63-1.30			
Deep and superficial	0.639038	1.12	0.70-1.78			
Histopathology						
Angiosarcoma	1				1	
MPNST	0.040002	2.43	1.04-5.68	0.021816	2.57	1.15-5.76
Myxofibrosarcoma	0.875043	1.07	0.46-2.47	0.766829	1.13	0.51-2.50
Synoviosarcoma	0.239728	1.68	0.71-4.02	0.165689	1.79	0.79-4.07
Sarcoma nos	0.562888	1.37	0.47-3.93	0.496240	1.43	0.51-4.01
Spindle cell sarcoma	0.306593	1.56	0.66-3.69	0.203214	1.67	0.76-3.67
MFH/UPS	0.561147	1.31	0.53-3.21	0.458632	1.37	0.60-3.14
Surgical margin						
Intralesional	1				1	
Marginal (≤ 2 mm)	0.065515	0.74	0.54-1.02	0.066450	0.75	0.54-1.02
Wide (> 2 mm)	0.040325	0.68	0.46-0.98	0.035266	0.68	0.48-0.97
Limb sparing operation (yes vs. no)	0.915264	0.98	0.62-1.54			
Radiotherapy (yes vs. no)						
Local recurrence (yes vs. no) ^a	0.002935	1.57	1.17-2.11	0.003246	1.55	1.16-2.08
Distant metastasis (yes vs. no) ^a	0e+00	15.18	11.72-19.65	0e+00	15.19	11.75-19.65

^a Time-dependent variable

Table 3: Potential prognostic factors in different models: (C) full model without DM, (D) reduced model without DM Variable

Variable	C. P value	HR	.95 CI	D. P value	HR	.95 CI
Age in years (continuous)	2e-07	1.02	1.01-1.03	3e-07	1.02	1.01-1.02
Presentation						
Primary		1				
Whoops	0.985788	1.00	0.71-1.42			
Irradical resection	0.723227	0.87	0.40-1.89			
Location (lower vs. upper extremity)	0.213073	1.19	0.91-1.55			
Size in cm (continuous)	1e-10	1.06	1.04-1.08	0e+00	1.07	1.05-1.09
Depth						
Deep		1				
Superficial	0.495898	0.89	0.62-1.26			
Deep and superficial	0.679694	1.11	0.68-1.81			
Histopathology						
Angiosarcoma		1			1	
MPNST	0.002098	3.66	1.60-8.38	0.001857	3.56	1.60-7.90
Myxofibrosarcoma	0.051777	2.20	0.99-4.87	0.054710	2.11	0.99-4.51
Synoviosarcoma	0.010462	2.98	1.29-6.89	0.008975	2.93	1.31-6.57
Sarcoma nos	0.056318	2.66	0.97-7.28	0.045565	2.74	1.02-7.35
Spindle cell sarcoma	0.030912	2.49	1.09-5.72	0.026120	2.46	1.11-5.43
MFH/UPS	0.024264	2.69	1.14-6.35	0.021244	2.64	1.16-6.03
Surgical margin						
Intralesional		1			1	
Marginal (≤ 2 mm)	0.813608	0.96	0.71-1.30	0.899294	0.98	0.73-1.32
Wide (> 2 mm)	0.740644	0.94	0.65-1.35	0.949360	0.99	0.70-1.40
Limb sparing operation (yes vs. no)	0.455498	0.85	0.55-1.30			
Radiotherapy (yes vs. no)						
Local recurrence (yes vs. no) ^a	0e+00	3.60	2.69-4.82	0e+00	3.52	2.64-4.71
Distant metastasis (yes vs. no) ^a						

^a Time-dependent variable

All analysis have been performed using the R-package `mstate` Wreede, Fiocco & Putter (2010), Wreede, Fiocco & Putter (2011).

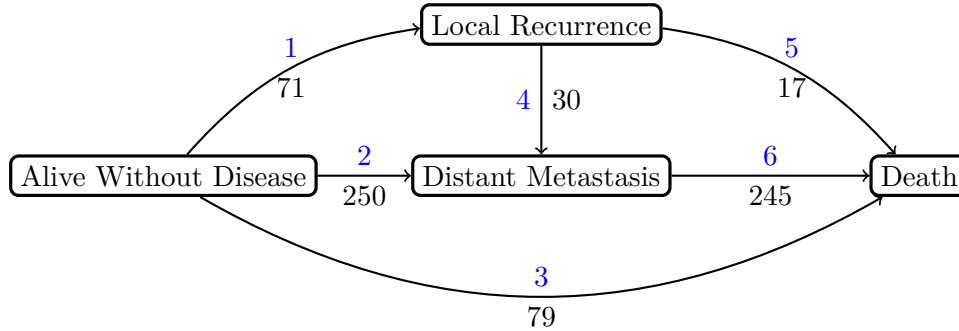


Figure 1: Disease progression of soft tissue sarcoma represented in a multi-state model. Blue: transition numbers. Black: number of patients moving from one state to the other. Total number of patients is 687.

Before performing the analysis the data must be transformed into a different long format. In this format each transition for which a patient is at risk is represented in a row. A start and stop variable indicate the intervals a patient is at risk for the corresponding transition and a status variable indicates whether or not a transition was made.

```

## An object of class 'msdata'
##
## Data:
##      id trans Tstart Tstop status age size.cm RT margin.class
## 9    17     1      0     15      0  86   3.2 yes    0.1-2mm
## 10   17     2      0     15      0  86   3.2 yes    0.1-2mm
## 11   17     3      0     15      1  86   3.2 yes    0.1-2mm
## 12   18     1      0      2      0  31   9.0 yes    0.1-2mm
## 13   18     2      0      2      1  31   9.0 yes    0.1-2mm
## 14   18     3      0      2      0  31   9.0 yes    0.1-2mm
## 15   18     6      2     14      1  31   9.0 yes    0.1-2mm
## 117 257     1      0     17      0  29  12.5 yes intralesional
## 118 257     2      0     17      1  29  12.5 yes intralesional
## 119 257     3      0     17      0  29  12.5 yes intralesional
## 120 257     6     17     61      1  29  12.5 yes intralesional
  
```

The previously shown patients are now represented in the long format for the multi-state analysis. Patient 17 is at risk for transition 1, 2 and 3 in the interval [0,

15) and moves at time 15 to the state death. Patient 18 is at risk for transition 1, 2 and 3 in the interval [0, 2) and experiences DM at time 2. He then is at risk for transition 6 in the interval [2,14) and moves at time 14 to the state death . Patient 257 is at risk for transition 1, 2 and 3 in the interval [0, 17) and moves to the state DM at time 17. He is then at risk for transition 6 in the interval [17, 61) and moves at time 61. Since he experienced LR only after DM he does not transition to the state LR.

Test Assumptions

The Markov assumption was tested for transitions from an intermediate state. For transitions from LR it was tested by including the time of LR as covariate into the model and for the transition from DM by including time of DM. For all transitions there was no significant effect of the time variable ($p < 0.05$). The proportional hazards assumption was tested for each transition ($p < 0.01$). In some transitions violation for some covariates was found and adjusted for as specified in Table 4.

Transition	Complete separation	Violates prop.	Adjustment
1			
2			
3		depth, RT	Stratify by depth:RT
4	depth, limb_sparing and type		Remove depth, limb_sparing, type
5	depth and type		Remove depth, type
6	presentation and type		Remove presentation, type

Table 4: Specifying adjustments for each transition.

Limb sparing operation

A limb sparing operation is one of the possible prognostic factors mentioned by clinicians and it is highly correlated with margin. Almost all patients with amputation have also a wide margin. It is not possible to estimate it for transition 4 and it is not significant for the other transitions. Including/excluding this covariate has an effect on margin in transition 5 and it changes the direction of its effect for wide margin, see Table 5. Only 17 patients make this transition and therefore the interpretation must be made carefully. Generally one could argue that the true effect of margin should be adjusted for the extra risk of amputation. But in the case of transition 5 where the number of patients at risk in the amputated group is only 4, with 3 patients transitioning, a limb sparing operation was excluded as a covariate.

Transition	Margin	Limb sparing	P value	HR	.95 CI
5	Marginal	with	0.443133	1.70	0.44-6.60
		without	0.416831	1.75	0.45-6.73
	Wide	with	0.692003	0.57	0.04-9.14
		without	0.653625	1.53	0.24-9.65

Table 5: Difference in margin including/excluding limb_sparing

Results

Cox analysis

The estimates of the 4 different models are shown in Table 2 and 3. There is almost no difference in estimates between the full models and the corresponding reduced models. Therefore only the full models are discussed.

The model including DM shows a significant effect of age, size, type MPNST, wide margin, LR and DM. The hazard ratios for age are equal to 1.03 (1.02-1.04), size 1.05 (1.03-1.08), hystopathology MPNST 2.43 (1.04-5.68), LR 1.57 (1.17-2.11) and DM 15.18 (11.72-19.65) indicate their negative effect on survival.

The effect of radiotherapy can not be estimated, because the model is stratified by it. A wide margin has a significant protective effect with hazard ratio equal to 0.68 (0.46-0.98) compared to an intralesional margin. A marginal margin has a nonsignificant effect, with hazard ratio equal to 0.74 (0.54-1.02).

The model excluding DM shows a significant effect of age, size, some types and LR. The hazard ratio for age is equal to 1.02 (1.01-1.03) and for size equal to 1.06 (1.04-1.08) are very similar to the full model. The hazard ratio of LR is twice as big as before with HR equal to 3.60 (2.69-4.82) compared to 1.57 (1.17-2.11) for the full model. Several types now show a significantly different effect from the baseline type Angiosarcoma, with HRs ranging between 2.20 and 3.66. This was not detectable in the model including DM.

The protective effect of margin in the previous model is not seen anymore. The HRs are equal to 0.96 (0.71-1.30) and 0.94 (0.65-1.35) for marginal and wide margin, respectively.

Table 6: Potential prognostic factors in different transitions.

Variable	AWD \rightarrow LR		AWD \rightarrow DM		AWD \rightarrow Death				
	P value	HR	.95 CI	P value	HR	.95 CI			
Age in years (continuous)	0.134566	1.01	1.00-1.03	0.791973	1.00	0.99-1.01	8e-09	1.06	1.04-1.08
Presentation									
Primary	1			1				1	
Whoops	0.877531	1.07	0.47-2.43	0.864611	0.97	0.66-1.43	0.492881	1.29	0.62-2.66
Irradical resection	0.225471	1.90	0.67-5.38	0.169903	0.49	0.18-1.36	0.977279	1.03	0.13-8.15
Location (lower vs. upper)	0.902267	0.96	0.54-1.72	0.867101	1.03	0.75-1.40	0.372523	0.78	0.45-1.35
Size in cm (continuous)	0.016239	1.05	1.01-1.10	4e-08	1.06	1.04-1.08	0.000502	1.07	1.03-1.11
Depth									
Deep	1			1					
Superficial	0.326194	0.70	0.35-1.42	0.893937	0.97	0.66-1.44			
Deep and superficial	0.125802	0.19	0.02-1.59	0.692581	1.12	0.64-1.94			
Histopathology									
Angiosarcoma	1			1				1	
MPNST	0.153687	0.38	0.10-1.43	0.759038	0.88	0.38-2.04	0.509309	1.69	0.36-8.03
Myxofibrosarcoma	0.141240	0.41	0.12-1.35	0.812844	1.10	0.49-2.46	0.436017	0.54	0.12-2.53
Synoviosarcoma	0.111118	0.35	0.10-1.27	0.830800	1.09	0.48-2.51	0.368218	0.43	0.07-2.71
Sarcoma nos	0.908661	0.91	0.17-4.73	0.587837	0.71	0.20-2.48	0.277197	0.31	0.04-2.54
Spindle cell sarcoma	0.241711	0.48	0.14-1.65	0.736808	0.87	0.38-1.98	0.872938	0.88	0.19-4.07
MFH/UPS	0.205112	0.41	0.10-1.63	0.759908	1.15	0.47-2.78	0.980503	1.02	0.20-5.23
Surgical margin									
Intralesional	1			1				1	
Marginal (≤ 2 mm)	0.014616	0.51	0.30-0.88	0.794398	0.95	0.65-1.39	0.170923	0.66	0.37-1.19
Wide (> 2 mm)	0.000004	0.13	0.06-0.31	0.618878	1.11	0.73-1.70	0.072695	0.51	0.25-1.06
Limb sparing (yes vs. no)	0.489309	1.52	0.47-4.95	0.732487	1.09	0.67-1.77	0.552169	0.76	0.31-1.88
Radiotherapy (yes vs. no)	0.000034	0.29	0.16-0.52	0.745227	0.94	0.67-1.34			

Table 7: Potential prognostic factors in different transitions.

Variable	LR \rightarrow DM		LR \rightarrow Death		DM \rightarrow Death				
	P value	HR	.95 CI	P value	HR	.95 CI			
Age in years (continuous)	0.071733	0.98	0.96-1.00	0.006356	1.06	1.02-1.11	0.000028	1.02	1.01-1.02
Presentation									
Primary		1			1				
Whoops	0.121912	0.29	0.06-1.39	0.408996	2.29	0.32-16.32			
Irradical resection	0.201688	0.25	0.03-2.11	0.640249	1.78	0.16-20.02			
Location (lower vs. upper)	0.044473	3.31	1.03-10.66	0.984641	0.99	0.26-3.69	0.537412	1.11	0.80-1.55
Size in cm (continuous)	0.949159	1.00	0.9-1.08	0.057914	1.11	1.00-1.24	0.019938	1.03	1.00-1.05
Depth									
Deep								1	
Superficial							0.966162	0.99	0.68-1.45
Deep and superficial							0.168057	0.66	0.36-1.19
Histopathology									
Angiosarcoma									
MPNST									
Myxofibrosarcoma									
Synoviosarcoma									
Sarcoma nos									
Spindle cell sarcoma									
MFH/UPS									
Surgical margin									
Intralesional		1			1			1	
Marginal (≤ 2 mm)	0.647160	1.22	0.52-2.87	0.413912	1.75	0.46-6.75	0.022192	0.66	0.46-0.94
Wide (> 2 mm)	0.595968	1.57	0.29-8.38	0.649029	1.53	0.24-9.71	0.062210	0.67	0.44-1.02
Limb sparing (yes vs. no)							0.164530	1.45	0.86-2.45
Radiotherapy (yes vs. no)	0.011005	3.93	1.37-11.31	0.891064	1.10	0.29-4.09	0.006770	0.61	0.43-0.87

Multi-state analysis

The estimates resulting from the multi-state analysis illustrated in Figure 1 are shown in the Tables 6 and 7. Bigger tumor size increases the risk of moving from one state to another for all transitions significantly, but for transition 4 (LR \rightarrow DM) and 5 (LR \rightarrow Death). Older age increases the risk of transitioning significantly for transition 3 (AWD \rightarrow Death), 5 (LR \rightarrow Death) and 6 (DM \rightarrow Death).

Location in the lower extremity is a significant risk factor for transition 4 (LR \rightarrow DM).

Radiotherapy has a significant protective effect for transition 1 (AWD \rightarrow LR) and 6 (DM \rightarrow Death), with HRs equal to 0.29 (0.16-0.52) and 0.61 (0.43-0.87), respectively. It is not possible to estimate the effect for transition 3 (AWD \rightarrow Death) since it was stratified by radiotherapy. In transition 4 radiotherapy increases the risk of moving from (LR \rightarrow DM) with HR equal to 3.93 (1.37-11.31)

Both marginal and wide margin have a significant protective effect compared to an intralesional margin for transition 1 (AWD \rightarrow LR) with HRs equal to 0.51 (0.30-0.88) and 0.13 (0.06-0.31). For transition 2 (AWD \rightarrow DM) both HRs are nonsignificant and close to 1, 0.95 (0.65-1.39) and 1.11 (0.73-1.70). For transition 3 (AWD \rightarrow Death) for both categories the HRs are nonsignificantly below 1, 0.66 (0.46-0.94) and 0.51 (0.25-1.06), for marginal and wide margin respectively. For transition 4 (LR \rightarrow DM) and 5 (LR \rightarrow Death) both categories are not significant but their HRs in both cases are above 1, 1.22 (0.52-2.87), 1.57 (0.29-8.38) and 1.75 (0.46-6.75), 1.53 (0.24-9.71), respectively. For transition 6 (DM \rightarrow Death) a significant protective effect of a marginal margin was found with a HR of 0.66 (0.46-0.94), while for a wide margin (HR:0.67 (0.44-1.02)) no significant effect could be detected.

For all other risk factors no significant effect could be detected.

Prediction Multi-state

In the previous section the effects of covariates on the transition hazards have been modelled. In this section the prediction aspect will be investigated. Future disease progression can be estimated not only based on the baseline characteristics of a patient, but also on his disease progression. In this way prediction can easily be updated in time when more information becomes available. Predictions are made by estimating the conditional probabilities of some clinical future events, given an (event) history, and a set of values for prognostic factors of a patient.

Some notation to explain the prediction procedure is introduced, following (Putter, M.Fiocco & Geskus (2007)). Let u be the time at which the prediction is made measured from the time origin of the patient. Let H_u denote the event history of a patient, containing the times of all events recorded and the event types. Let Z be the values for prognostic factors of the patient. Denote E_t some future event

evaluated at time t , e.g. the event of surviving until $t = 5$ years after surgery. Our interest lies now on the conditional probability $Prob(E_t|H_u, Z)$. Given multi-state model in Figure 1, these probabilities can be estimated by appropriately combining the estimated baseline hazards and regression coefficients.

Suppose a patient is 50 years old with a tumor of size 4 cm. He experiences LR after 13 months. One could be interested in his 5 year survival probabilities at baseline and after diagnosing LR at 13 months. Let R denote the time of entering the current state and T the time of death. Then the corresponding event histories and conditional probabilities are:

1. From state 1 at time 0:

$$H_{1,0}(0) = \{R = 0, T > 0\}$$

$$P_{14,0}(0, 120) = Prob(T > t|H_{1,0}(0), Z)$$

2. From state 2 at time 13:

$$H_{2,13}(13) = \{R = 13, T > 13\}$$

$$P_{24,13}(13, 120) = Prob(T > t|H_{2,13}(13), Z)$$

The transition probabilities are estimated with the function `probtrans()` from the `mstate` package using the Aalen-Johansen estimator (Wreede, Fiocco & Putter (2011)). The plots in Figure 2 are made by using the `fixedhorizon` option of `probtran()`, which returns probabilities for a fixed time point from the current time and specified state. The plot on the left shows for each point in time the probabilities of being in the corresponding state 5 years after surgery, conditional on being in state AWD at that time. The plot on the right shows the changed 5 year probabilities after the patient experienced LR.

We shall make predictions for three patients with different margin status, 50 years old with a tumor of size 4 cm. Figure 3 displays their different transition intensities.

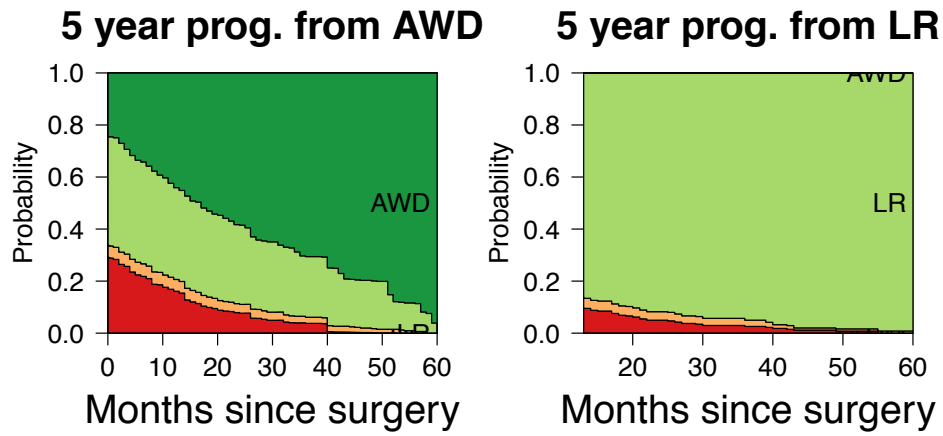


Figure 2: 5 year prognosis. Left: from state AWD. Right: from state LR.

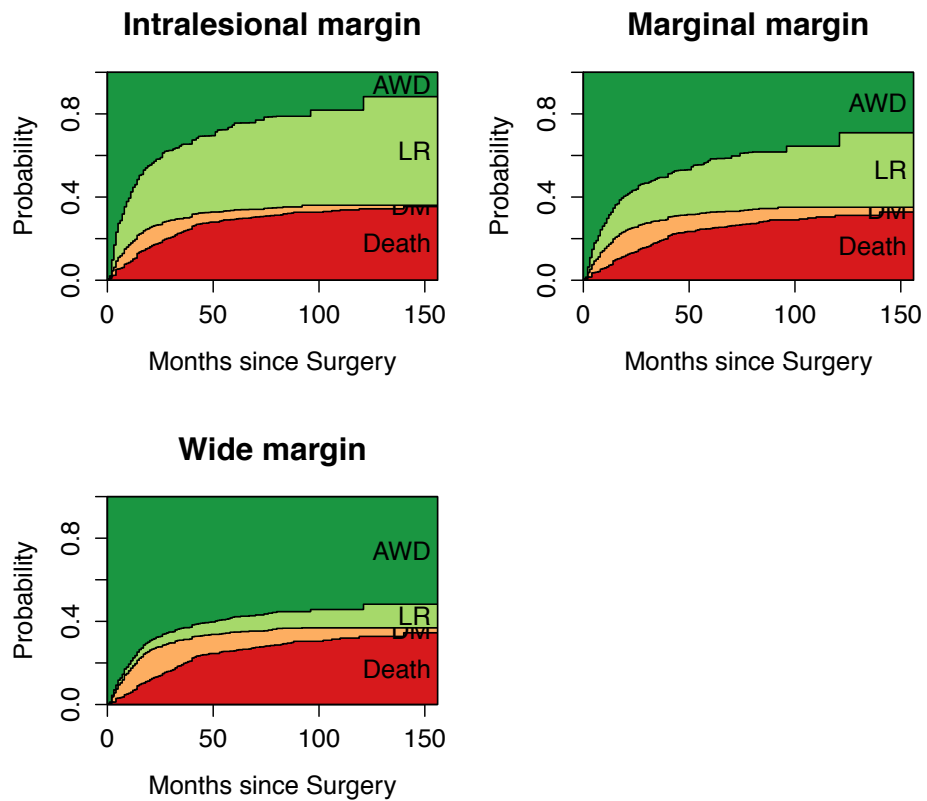


Figure 3: Stacked transition probabilities at time 0 for 3 patients with different margin status.

Conclusion and discussion

The results from the time-dependent Cox analysis including DM confirmed the negative effect of age, size, LR and DM. In this model a wider margin could be linked significantly to a higher survival probability.

Interesting is the analysis regarding the differences of the models including and excluding DM. The increase of effect of LR when excluding DM could indicate a to some degree causal relationship of LR and DM. This aspect was not investigated, since it is beyond the aim in the study.

Regarding margin, its effect is only present in the model including DM. This might be due to the fact that if not adjusting for DM the effect of margin, is too small to be detectable.

The results from the multi-state analysis give some more insight on the effect of the covariates on the transition rates. Age has a significant adverse effect on all transition rates to the state death. Age has no significant effect on the transition rates to LR and DM. Size has as expected a significant adverse effect on all transition rates, but transition 4 and 5, with transition 5 being marginally significant.

Tumor location in the lower extremity compared to the upper extremity increased the risk of transitioning from LR to DM.

Interesting is the effect of radiotherapy on the different transition rates in the multi-state model. For transitions 1 (AWD \rightarrow LR) and 6 (DM \rightarrow Death) it has a significant protective effect. For transition 3 (AWD \rightarrow Death) the effect could not be estimated because of stratification. For transition 4 (LR \rightarrow DM) radiotherapy has a highly significant negative effect. It is important to note that here only radiotherapy at baseline is considered. An interpretation of the strong negative effect for transition 4 (LR \rightarrow DM) could be that if a patient experiences LR *even though* radiotherapy was administered it could be a rather aggressive tumor with consequence on the transition rate to DM.

As expected both marginal and wide margin have a significant protective effect on the transition rate 1 (AWD \rightarrow LR). There is also a significant protective effect for marginal margin on the transition rate 6 (DM \rightarrow Death), which was not expected. No association between margin and transition rates 2 (AWD \rightarrow DM) and 3 (AWD \rightarrow Death) could be found. For transition 4 and 5, which are the transitions from LR also no significant effect could be found, but it is interesting that the HRs for both transitions and categories are bigger than 1. These results are similar to the ones of radiotherapy and might be explained in a similar fashion.

Website to estimate patient specific probabilities of LR and OS at fixed time points

Motivation

The website was developed to help clinicians in their decision making process regarding treatment choices and to inform patients about different scenarios concerning disease progression.

The models used for the website are fitted on the data set described above. This website provides estimates of 5 and 10 year survival probability, as well as 5 and 10 year probability of LR, from surgery. It also returns the patient specific survival curve as well as the patient specific cumulative incidence function of LR.

Models

The covariates included in the models were suggested by clinicians: age, sex, size, margin and type. Tumor size is used as a categorical variable with three levels (0-49.9mm, 50-99.9mm, ≥ 100). Age is used categorical with two levels (\leq median, $>$ median; 62 years).

Time-fixed Cox model

Since estimation is done from surgery, at or before time of surgery, no information concerning LR or DM is available. Therefore in this model only time-fixed covariates were included and based on it the prediction of OS is done. Margin violates the proportional hazard assumption, therefore it is incorporated in the model as stratified risk factor. The estimates for this model are shown in Table 8.

The confidence interval given alongside the estimate of survival on the website is retrieved from the `survfit.object`.

Competing risk model

The probability of LR was estimated in a competing risk setting, with death as competing event.

If a patient dies, then he is no longer at risk of experiencing LR and if not taken into account the probability of LR will be overestimated. In the case that LR and death happened at the same time (5 patients) death was censored, since LR must have happened first. The competing risk model for this analysis is shown in Figure 4.

Table 8: Time-fixed Cox model, stratified by margin.

Variable	P value	HR	.95 CI
Age (above median 62 years)	0.000166	1.56	1.24-1.98
Sex (female vs. male)	0.294600	0.89	0.71-1.11
Size			
<5cm		1	
5-9.99cm	0.008927	1.76	1.15-2.69
≥10cm	2e-09	3.50	2.32-5.27
Histopathology			
Angiosarcoma		1	
MPNST	0.357367	1.44	0.66-3.15
Myxofibrosarcoma	0.780683	1.11	0.53-2.33
Synoviosarcoma	0.735928	1.14	0.53-2.49
Sarcoma nos	0.396238	1.51	0.58-3.95
Spindle cell sarcoma	0.5549758	1.26	0.59-2.68
MFH/UPS	3.127502e-01	1.51	0.68-3.36

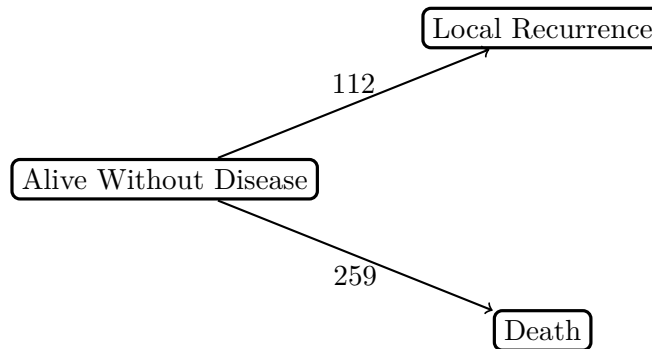


Figure 4: Competing Risk analysis, 687 patients at risk.

The data is transformed into a long format similar to the one for the multi-state analysis performed earlier. Here there are 2 competing events, each individual needs 2 rows in the long format data file, one for each possible cause of failure. The status variable indicates whether or not a patient experienced the specific transition. A start and stop variable indicate the interval a patient is at risk.

```
## An object of class 'msdata'
```

Table 9: Competing risk analysis.

Variable	AWD → LR			AWD → Death		
	P value	HR	.95 CI	P value	HR	.95 CI
Age (above median 62 years)	0.292017	1.24	0.83-1.85	0.000320	1.64	1.25-2.15
Sex (female vs. male)	0.428991	1.16	0.80-1.70	0.194821	0.85	0.66-1.09
Size						
<5cm		1			1	
5-9.99cm	0.689154	1.14	0.59-2.21	0.013763	1.81	1.13-2.89
≥10cm	0.042409	1.93	1.02-3.66	1e-07	3.42	2.17-5.42
Histopathology						
Angiosarcoma		1			1	
MPNST	0.056012	0.36	0.13-1.03	0.816416	0.90	0.37-2.19
Myxofibrosarcoma	0.006830	0.26	0.10-0.69	0.504742	0.75	0.32-1.74
Synoviosarcoma	0.005076	0.21	0.07-0.63	0.619523	0.80	0.34-1.91
Sarcoma nos	0.079538	0.27	0.06-1.17	0.892599	0.93	0.31-2.81
Spindle cell sarcoma	0.041334	0.36	0.13-0.96	0.513094	0.75	0.32-1.77
MFH/UPS	0.293174	0.57	0.20-1.63	0.546698	0.75	0.29-1.91
Surgical margin						
Intralesional		1			1	
Marginal (≤2mm)	0.000022	0.40	0.27-0.61	0.207209	0.79	0.55-1.14
Wide (>2mm)	1e-08	0.19	0.11-0.33	0.697766	0.93	0.63-1.36

```
##
## Data:
##      id trans Tstart Tstop status age.class      sex size.class  margin.class
## 5   17     1     0    15     0  >median  male    0-49.9    0.1-2mm
## 6   17     2     0    15     1  >median  male    0-49.9    0.1-2mm
## 7   18     1     0    14     0 <=median male    50-99.9   0.1-2mm
## 8   18     2     0    14     1 <=median male    50-99.9   0.1-2mm
## 65 257     1     0    21     1 <=median female  >=100  intralesional
## 66 257     2     0    21     0 <=median female  >=100  intralesional
```

Patient 17 is at risk for transition 1,2 in the interval [0,15) and dies at time 15.
 Patient 18 is at risk for both transitions in the interval [0,14) and dies at time 14.
 Patient 257 is at risk for both transitions in the interval [0,21) and gets LR at time 21.

The estimates of the transition specific covariates are shown in Table 9. The confidence

intervals given alongside the probability of LR on the website are calculated from the standard errors available in the `probtrans.object`.

Implementation

The website was implemented using the R-package `Shiny`. It is a package from RStudio to build interactive web applications with R.

Shiny apps have two components, the user-interface script and the server script. The user-interface script controls layout and appearance of the app. The server script contains the computational part in the background. Model based prediction are estimated and handed over to the user-interface script for visualization.

When loading the app, on the left hand side a side-bar with empty scroll down menus is visible, as in Figure 5. The user can now select patient specific characteristics. After clicking on the ‘calculate’ button an estimation of 5 and 10 year OS and probability of LR is returned, as in Figure 6. On the ride hand side through selection of the top tabs the plot of overall suvival (Figure 7) and the cumulative incidence function of LR (Figure 8) can be selected.

Prognostic tool for highgrade soft-tissue sarcoma patients

5 and 10 year prognosis

ESTIMATES OVERALL SURVIVAL LR

Enter your information

Age:
Select age

Sex:
Select sex

Tumorsize:
Select tumorsize

Margin:
Select margin

Tumortype:
Select tumortype

CALCULATE

Figure 5: Empty website.

Prognostic tool for highgrade soft-tissue sarcoma patients

5 and 10 year prognosis

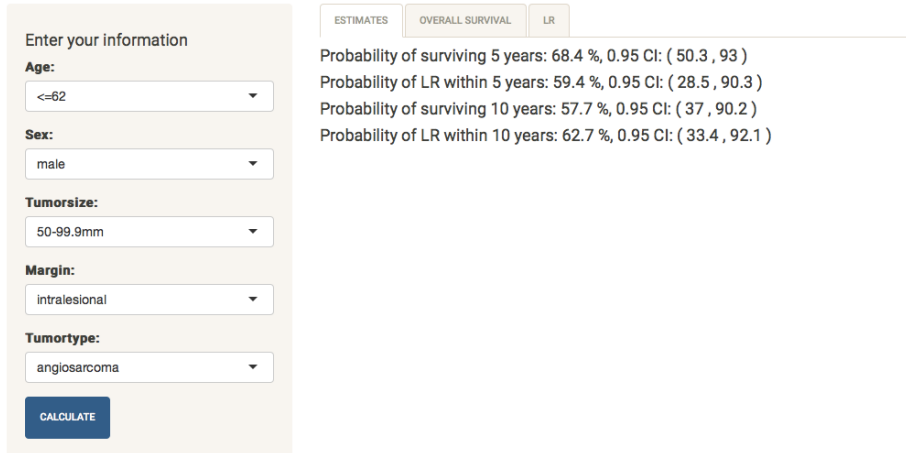


Figure 6: Estimates for 5 and 10 year survival and cumulative incidence of LR.

Prognostic tool for highgrade soft-tissue sarcoma patients

5 and 10 year prognosis

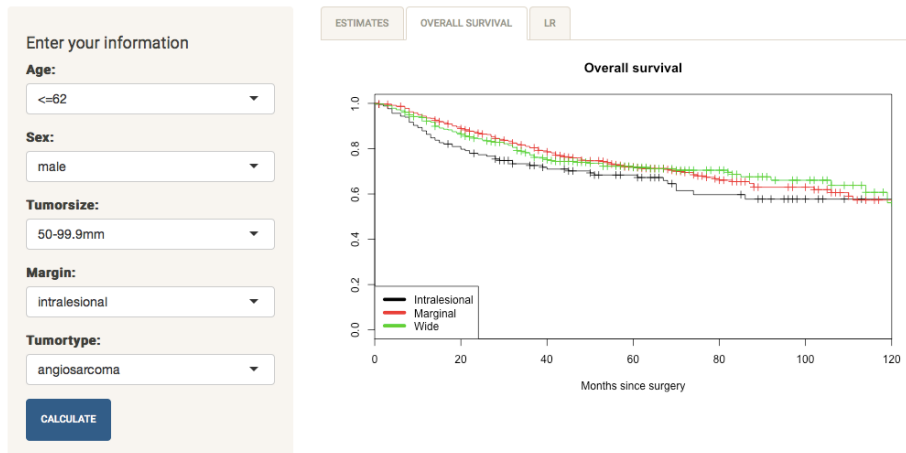


Figure 7: Survival curves stratified by margin.

Prognostic tool for highgrade soft-tissue sarcoma patients

5 and 10 year prognosis

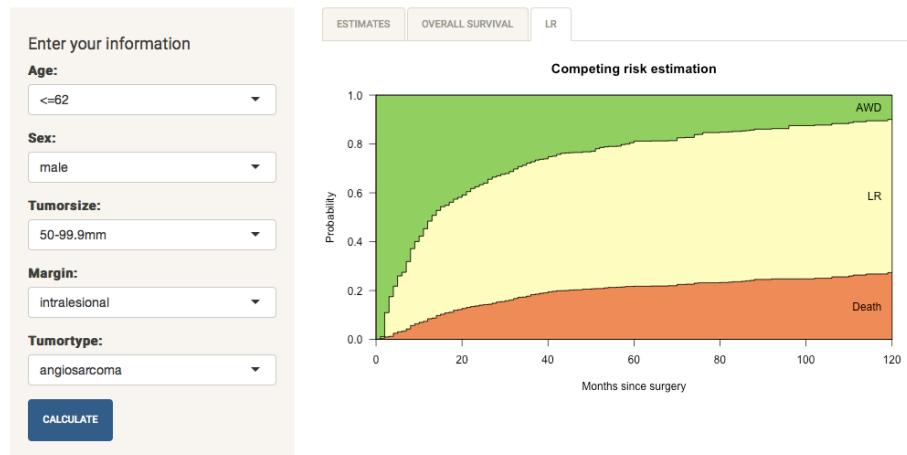


Figure 8: Cumulative Incidence of LR.

Discussion

A user friendly website was created, which immediately returns patient specific estimates. The prediction models proposed in the app need to be validated. This piece of work will be done in the future.

References

- Italiano A, Cesne AL, Mendiboure J, et al. (2014): Prognostic factors and impact of adjuvant treatments on local and metastatic relapse of soft-tissue sarcoma patients in the competing risks setting. *Cancer* **120**: 3361–3369.
- Marety-Nielsen K, Aggerholm-Pedersen N, Safwat A, Jørgensen PH, Hansen BH, Baerentzen S & Keller ABP and Johnny (2014): Prognostic factors for local recurrence and mortality in adult soft tissue sarcoma of the extremities and trunk wall. *Acta Orthopaedica* **85**: 323–332.
- O'Donnell PW, Griffin AM, Eward WC, et al. (2014): The effect of the setting of a positive surgical margin in soft tissue sarcoma. *Cancer* **120**: 2866–2875.
- Putter H, M.Fiocco & Geskus RB (2007): Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* **26**: 2389–430.

- Toulmonde M, Cesne AL, Mendiboure J, et al. (2014): Long-term recurrence of soft tissue sarcomas: prognostic factors and implications for prolonged follow-up. *Cancer* **120**: 3003–3006.
- Trovik C, Bauer H, Alvegård T, et al. (2000): Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma Group Register. *Eur J Cancer* **36**: 710–716.
- Willeumier J, Fiocco M, Nout R, et al. (2015): High-grade soft tissue sarcomas of the extremities: surgical margins influence only local recurrence not overall survival. *International Orthopaedics* **39**: 935–941.
- Wreede LC de, Fiocco M & Putter H (2010): The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* **99**: 261–274.
- Wreede LC de, Fiocco M & Putter H (2011): mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. *Journal of Statistical Software* **38**: 1–30.

Appendix

Data checking and merging

```
#functions
library(Hmisc)
library(survival)
library(mstate)
setwd("/Users/anjarutenbudde/Desktop/Uni Leiden/Internship/OwnRcode")

#####
#written for first dataset
#####
getWideData <- function(stringOfDataName){
# Returns dataset with relevant covariates in wide format
  raw<-spss.get(stringOfDataName,
                to.data.frame=T,
                datevars = c("DoB",
                             "dateofsurgery",
                             "LASTKNOWNdate",
                             "dateoffirstlocalrecurrence",
                             "dateoffirstdistantmetastasis",
                             "dateof2ndlocalrecurrence",
                             "dateof2nddistantmetastasis"))

  # quickFixMarginLeiden
  margin.class <- factor(ifelse(raw$marginmm<=0,0,ifelse(raw$marginmm<=2,1,2)),
                        labels=c("intralesional","0.1-2mm", ">2mm"))
  diff <- (raw$Margins!=margin.class|is.na(raw$Margins))&!is.na(margin.class)
  raw$Margins[diff] <- margin.class[diff]

  daysInYear <- 365.242199
  daysInMonth <- 30.4368499
  data <- data.frame(id = as.numeric(apply(cbind(raw$Center,raw$IDnr),
                                          1, function(x) paste(x, collapse="")),
                                          center = raw$Center,
                                          time_surv = as.numeric(round(difftime(raw$LASTKNOWNdate,
                                          raw$dateofsurgery,
                                          unit = "days")/daysInMonth)),
                                          event_surv = ifelse(raw$Status.dead.alive=="dead", 1, 0),
                                          time_LR = as.numeric(round(difftime(raw$dateoffirstlocalrecurrence,
                                          raw$dateofsurgery,
                                          unit = "days")/daysInMonth)),
                                          event_LR = raw$LR,
                                          time_DM = as.numeric(round(difftime(raw$dateoffirstdistantmetastasis,
                                          raw$dateofsurgery,
                                          unit = "days")/daysInMonth)),
                                          event_DM = ifelse(raw$distantmetastasis=="yes", 1, 0),
                                          age = as.numeric(trunc(difftime(raw$dateofsurgery,
                                          raw$DoB, units="days")/daysInYear)),
                                          depth = factor(ifelse(raw$Depth==1, 0,
                                          ifelse(raw$Depth==2, 1,
```

```

        ifelse(raw$Depth==3, 2, NA))),
        labels = c("deep","superfiscial","deep&superfiscial")),
size.class = factor(raw$tumor.size.groups,
        labels=c("0-49.9","50-99.9",">=100")),
size.cm= raw$maxdiametertmm/10,
RT = raw$RT,
chemo=raw$Chemo,
margin.class = raw$Margins,
limb_sparing = factor(ifelse(raw$operation=="amputation",0,1),
        labels=c("no","yes")),
presentation = factor(ifelse(raw$tumorpresentation=="primary",0,
        ifelse(raw$tumorpresentation=="whoops",1,2)),
        labels=c("primary", "whoops",
                "irradical resection / residue")),
localization = factor(raw$tumorlokalization),
type = raw$tumor.type,
sex = raw$gender)
print(c("Nbr of missclassified margins:", sum(diff)))
return(data)
}

#####
#written for second dataset
#####
loadRaw <- function(stringOfDataName){
  raw<-spss.get(stringOfDataName,
    to.data.frame=T,
    datevars = c("DoB",
      "dateofsurgery",
      "Lastknowndate",
      "dateoffirstlocalrecurrence",
      "dateoffirstdistantmetastasis"))
# Missing values in dates are imported as 0 and transformed to 1582-10-14, correct:
raw$DoB[raw$DoB=="1582-10-14"] <- NA
raw$dateofsurgery[raw$dateofsurgery=="1582-10-14"] <- NA
raw$dateoffirstlocalrecurrence[raw$dateoffirstlocalrecurrence=="1582-10-14"] <- NA
raw$dateoffirstdistantmetastasis[raw$dateoffirstdistantmetastasis=="1582-10-14"] <- NA
  return(raw)
}

checkRaw <- function(raw){
#First check before creating the working data:
#Missing data
#1. Missing Last.known.outcome:      give id
#Check illogical order of events
#2. Birthday after surgery:          give id
#3. Lastknowndate before surgery:    give id
#4. No date for LR,DM, missing event: censor event
#5. No date LR,DM, event:            give id
#6. Margin missclassification:      corect

# Quick fix of margin classification
# Margmm is margin in mm (76 NA), Resectionmarg is classified (0 NA),

```

```

# but some contradictions. I correct Resectionmarg with the information of Margmm:
margin.class <- factor(ifelse(is.na(raw$Margmm),0,ifelse(raw$Margmm<=2,1,2)),
                      labels=c("Intralesional","Marginal 0.1-2mm","Wide > 2mm"))
diff <- (raw$Resectionmarg!=margin.class|is.na(raw$Resectionmarg))&!is.na(margin.class)
raw$Resectionmarg[diff] <- margin.class[diff]

# Output table:
t <- t(matrix(c("Missing Last.known.outcome, ID:",
               paste(na.omit(raw$IDnr[is.na(raw$Last.known.outcome)]),
                     collapse=" "),
               "Birthday after surgery, ID:",
               paste(na.omit(raw$IDnr[raw$DoB >=raw$dateofsurgery]),
                     collapse=" "),
               "Lastknowndate before surgery, ID:",
               paste(na.omit(raw$IDnr[raw$Lastknowndate<raw$dateofsurgery]),
                     collapse=" "),
               "Patients with no date for LR,DM, and missing event",
               paste(na.omit(raw$IDnr[is.na(raw$dateoffirstlocalrecurrence)&
                                     is.na(raw$LR)|
                                     is.na(raw$dateoffirstdistantmetastasis)&
                                     is.na(raw$distantmetastasis)]),collapse=" "),
               "No date LR,DM, but event, ID:",
               paste(na.omit(raw$IDnr[is.na(raw$dateoffirstlocalrecurrence)&
                                     !is.na(raw$LR)&
                                     (raw$LR=="Local recurrence")|
                                     is.na(raw$dateoffirstdistantmetastasis)&
                                     !is.na(raw$distantmetastasis)&
                                     (raw$distantmetastasis=="Distant metastasis")])),
               collapse=" "),
            "Nbr of missclassified margins:",
            sum(diff)),
        ,ncol=6))

#4. No date for LR and no status, censor event
raw$LR[is.na(raw$dateoffirstlocalrecurrence)&
       is.na(raw$LR)] <- "No local recurrence"
#4. No date for DM and no status, censor event
raw$distantmetastasis[is.na(raw$dateoffirstdistantmetastasis)&
                     is.na(raw$distantmetastasis)] <- "No distant metastasis"

print(t)
return(raw)
}

quickFix <- function(checkedRaw){
  #missing Last.known.outcome
  raw$Last.known.outcome[raw$IDnr==250] <- "Alive,m disease unknown"
  #wrong DoB
  checkedRaw$DoB[checkedRaw$IDnr==269] <- as.Date("1917-05-20", "%Y-%m-%d")
  #wrong DoB
  checkedRaw$DoB[checkedRaw$IDnr==166] <- as.Date("1912-09-18", "%Y-%m-%d")
  #wrong dateofsurgery
  checkedRaw$dateofsurgery[checkedRaw$IDnr==244] <- as.Date("2007-02-09", "%Y-%m-%d")
  #Remove because of missing event time
  checkedRaw <- checkedRaw[checkedRaw$IDnr!=199,]

```

```

#corrected date of surgery (julies mail)
checkedRaw$dateofsurgery[checkedRaw$IDnr==263] <- as.Date("2007-11-15", "%Y-%m-%d")
return(checkedRaw)
}

#use after checkData
quickFix2 <- function(data){
  #exclude patient with DM at same day as surgery
  data <- data[data$id!=3260,]
  #Set time to fraction for p. dying 2 after 2 days, set censoring time for LR/DM
  data$time_surv[data$id==3165] = 0.1
  data$time_LR[data$id==3165] = 0.1
  data$time_DM[data$id==3165] = 0.1
  #Set time to fraction for patients with DM immediately after surgery
  data$time_DM[data$id==3316] = 0.2
  data$time_DM[data$id==3289] = 0.3
  data$time_DM[data$id==3111] = 0.5
  #data <- data[data$id!=3111&data$id!=3260&data$id!=3289&data$id!=3316,]
  return(data)
}

workingData <- function(firstCheck){
  # Retrieves all variables of interest in correct format from raw (checked) data
  # New id is combination of Center nbr and old id
  daysInYear <- 365.242199
  daysInMonth <- 30.4368499
  data <- data.frame(id = as.numeric(apply(cbind(firstCheck$Center,firstCheck$IDnr),
    1, function(x) paste(x, collapse="")),
    center = firstCheck$Center,
    time_surv = as.numeric(round(difftime(firstCheck$Lastknowndate,
      firstCheck$dateofsurgery,
      unit = "days")/daysInMonth)),
    event_surv = ifelse(firstCheck$Last.known.outcome %in%
      c("Dead, of disease","Dead, other cause",
      "Dead, of unknown cause"), 1, 0),
    time_LR = as.numeric(round(difftime(firstCheck$dateoffirstlocalrecurrence,
      firstCheck$dateofsurgery,
      unit = "days")/daysInMonth)),
    event_LR = ifelse(firstCheck$LR=="Local recurrence", 1, 0),
    time_DM = as.numeric(round(difftime(firstCheck$dateoffirstdistantmetastasis,
      firstCheck$dateofsurgery,
      unit = "days")/daysInMonth)),
    event_DM = ifelse(firstCheck$distantmetastasis=="Distant metastasis", 1, 0),
    age = as.numeric(trunc(difftime(firstCheck$dateofsurgery,
      firstCheck$DoB, units="days")/daysInYear)),
    depth = firstCheck$depth,
    size.class = factor(ifelse(firstCheck$maxdiametermm<5, 0,
      ifelse(firstCheck$maxdiametermm<10, 1,2)),
      labels=c("0-49.9", "50-99.9", ">=100")),
    size.cm = firstCheck$maxdiametermm,
    RT = factor(firstCheck$Radiotherapy, labels=c("no","yes")),
    chemo = factor(firstCheck$Chemo, labels=c("no","yes")),
    margin.class = factor(ifelse(is.na(firstCheck$Margmm),0,
      ifelse(firstCheck$Margmm<=2,1,2)),

```

```

                                labels=c("intralesional","0.1-2mm",
                                             ">2mm")),
operation = firstCheck$Operation,
limb_sparing = factor(ifelse(firstCheck$Operation=="Amputation",0,1),
                        labels=c("no","yes")),
presentation = factor(firstCheck$tumorpresentation,
                       labels=c("primary", "whoops")),
localization = factor(firstCheck$tumorlocalization,
                       labels=c("arm, upper incl shoulder",
                                "arm, lower incl elbow",
                                "hand, incl wrist",
                                "leg, upper incl hip and buttock",
                                "leg, lower incl knee",
                                "foot, incl ankle")),
type = factor(firstCheck$tumor.type, labels=c("angiosarcoma",
                                              "mpnst",
                                              "myxofibrosarcoma",
                                              "synoviosarcoma",
                                              "mfh/ups",
                                              "spindle cell sarcoma")),

sex = firstCheck$gender)
# missing values in time_LR/DM variables correspond to no event:
data$time_LR[is.na(data$time_LR)] <- data$time_surv[is.na(data$time_LR)]
data$time_DM[is.na(data$time_DM)] <- data$time_surv[is.na(data$time_DM)]
# depth has an extra empty level "Unknow", remove here:
levels(data$depth) <- c("deep","superficial","deep&superficial",
                       "deep&superficial")
return(data)
}

#####
#written for both
#####

checkData <- function(wideData){
  # Deals with problematic situations in the following way:
  # 1. Missing values in time variables (LR,DM,death): remove
  # 2. Missing values in event variables (LR,DM,death): remove
  # 3. Censored at 0, no events: remove
  # 4. Patients with time_DM, time_LR < 0: remove
  # 5. Patients with LR/DM/death at time 0: give ID
  # 6. LR or DM after death: set corr. time to time_surv
  # 7. No margin information: remove
  # 8. Events censored before time_surv Set censoring times to time_surv
  t <- t(matrix(c(c("Patients with missing time variables",
                    sum((is.na(wideData$time_LR) |
                          is.na(wideData$time_DM) |
                          is.na(wideData$time_surv))))),
                c("Patients with missing event variables",
                    sum((is.na(wideData$event_LR) |
                          is.na(wideData$event_DM) |
                          is.na(wideData$event_surv))))),

```

```

c("Censored at 0, no events",sum((wideData$time_surv == 0)&
!(wideData$event_surv)&
!(wideData$event_LR)&
!(wideData$event_DM), na.rm=T)),
c("Patients with time_DM or time_LR < 0",
sum(wideData$time_DM < 0 |
wideData$time_LR < 0, na.rm=T)),
c("Patients with LR/DM/death at time 0, ID:",
paste(na.omit(wideData$id[wideData$time_DM == 0 & wideData$event_DM |
wideData$time_LR == 0 & wideData$event_LR |
wideData$time_surv == 0 & wideData$event_surv]),
collapse=" ")),
c("Patients with time_LR, time_DM >time_surv:",
sum(wideData$time_surv < wideData$time_LR |
wideData$time_surv < wideData$time_DM,na.rm=T)),
c("Patients with no margin information:",
sum(is.na(wideData$margin.class), na.rm=T)),
c("Patients censored before time_surv:",
sum(wideData$time_DM < wideData$time_surv &
!wideData$event_DM |
wideData$time_LR < wideData$time_surv &
!wideData$event_LR, na.rm=T)),ncol=8))
# 1. Remove patients with missing values in time variables (LR,DM,death)
wideData <- wideData[!(is.na(wideData$time_LR) |
is.na(wideData$time_DM) |
is.na(wideData$time_surv)),]
# 2. Remove patients with missing values in event variables (LR,DM,death)
wideData <- wideData[!(is.na(wideData$event_LR) |
is.na(wideData$event_DM) |
is.na(wideData$event_surv)),]
# 3. Remove patients censored at 0, no events
wideData <- wideData[!((wideData$time_surv == 0)&
!(wideData$event_surv)&
!(wideData$event_LR)&
!(wideData$event_DM)),]
# 4. Remove patients with time_DM, time_LR < 0
wideData <- wideData[wideData$time_DM >= 0 &
wideData$time_LR >= 0,]
# 6. Set time to time_surv if time_LR or time_DM after death
checkLR <- ifelse(wideData$time_surv >= wideData$time_LR,0,1)
checkDM <- ifelse(wideData$time_surv >= wideData$time_DM, 0, 1)
wideData$time_LR[checkLR==1]<-wideData$time_surv[checkLR==1]
wideData$time_DM[checkDM==1]<-wideData$time_surv[checkDM==1]
# 7. Remove patients with no margin information:
wideData <- wideData[!is.na(wideData$margin.class),]
# 8. Censoring times to time_surv, if events censored before time_surv
checkLR <- (wideData$time_LR < wideData$time_surv & !wideData$event_LR)
checkDM <- (wideData$time_DM < wideData$time_surv & !wideData$event_DM)
wideData$time_LR[checkLR]<-wideData$time_surv[checkLR]
wideData$time_DM[checkDM]<-wideData$time_surv[checkDM]

print(rbind(t, c("Remaining Patients",nrow(wideData))))
return(wideData)

```

```

}

#1. Dataset
data1 <- getWideData("working_Toronto_LUMC_for_MS_4.sav")
data1 <- checkData(data1)
data1 <- data1[data1$id!=42052,]# DM before surgery

#2. Dataset
raw <- loadRaw("working_Birmingham for MS.sav")
checkedRaw <- checkRaw(raw)
fixed <- quickFix(checkedRaw)
data2 <- workingData(fixed)
data2 <- checkData(data2)
data2 <- quickFix2(data2)

#Merge two datasets
data <- data.frame(id = c(data1$id, data2$id),
                  center = factor(c(data1$center,data2$center),
                                labels = c("Leiden","London",
                                           "Birmingham","Toronto")),
                  time_surv = c(data1$time_surv, data2$time_surv),
                  event_surv = c(data1$event_surv, data2$event_surv),
                  time_LR = c(data1$time_LR, data2$time_LR),
                  event_LR = c(data1$event_LR, data2$event_LR),
                  time_DM = c(data1$time_DM, data2$time_DM),
                  event_DM = c(data1$event_DM, data2$event_DM),
                  age = c(data1$age, data2$age),
                  depth = factor(c(data1$depth, data2$depth),
                                labels=c("deep","superfiscial","deep&superfiscial")),
                  size.class = factor(c(data1$size.class, data2$size.class),
                                     labels=c("0-49.9","50-99.9",">=100")),
                  size.cm = c(data1$size.cm, data2$size.cm),
                  RT = factor(c(data1$RT, data2$RT),
                              labels=c("no","yes")),
                  chemo = factor(c(data1$chemo, data2$chemo),
                                 labels=c("no","yes")),
                  margin.class = factor(c(data1$margin.class, data2$margin.class),
                                       labels=c("intralesional","0.1-2mm",">2mm")),
                  limb_sparing = factor(c(data1$limb_sparing,data2$limb_sparing),
                                       labels=c("no","yes")),
                  presentation = factor(c(data1$presentation,data2$presentation),
                                       labels=c("primary","whoops",
                                               "irradical resection / residue")),
                  localization = factor(c(data1$localization,data2$localization),
                                       labels=c("arm, upper incl shoulder",
                                               "arm, lower incl elbow",
                                               "hand, incl wrist",
                                               "leg, upper incl hip and buttock",
                                               "leg, lower incl knee",
                                               "foot, incl ankle")),
                  type = factor(c(data1$type,ifelse(data2$type=="mfh/ups",
                                                    as.numeric(data2$type)+2,
                                                    as.numeric(data2$type))),

```



```

                                labels=c("angiosarcoma", "mpnst", "myxofibrosarcoma",
                                "synoviosarcoma", "sarcoma - nos",
                                "spindle cell sarcoma", "mfh/ups")),
sex = factor(c(data1$sex, data2$sex),
              labels=c("male", "female"))
)
data$extremity <- data$localization
levels(data$extremity) <- c("upper", "upper", "upper", "lower", "lower", "lower")
#save(data, file="MergedData.Rda")

```

Time-dependent Cox analysis

```

load("MergedData.Rda")
data <- na.omit(data)

longCox <- function(data){
  # transformes data to long format for coxph with time dependant covariates
  # 1. censure events when at the same time as death:
  data$event_DM[data$time_DM == data$time_surv &
                data$event_DM &
                data$event_surv] = 0
  data$event_LR[data$time_LR == data$time_surv &
                data$event_LR &
                data$event_surv] = 0

  n <- nrow(data)
  tstart <- tstop <- NULL
  for(i in 1:n){
    times <- c(data$time_LR[i],
               data$time_DM[i],
               data$time_surv[i])
    events <- c(data$event_LR[i],
                data$event_DM[i], 1)
    tied <- (data$time_LR[i]==data$time_DM[i] &
             data$event_LR[i] &
             data$event_DM[i])
    tiedZeroLR <- (data$time_LR[i]==0 &
                  data$event_LR[i])
    tiedZeroDM <- (data$time_DM[i]==0 &
                  data$event_DM[i])
    tiedSurvLR <- (data$time_LR[i]==data$time_surv[i] &
                  data$event_LR[i])
    tiedSurvDM <- (data$time_DM[i]==data$time_surv[i] &
                  data$event_DM[i])
    if(tied) events[1] = 0
    if(tiedZeroLR) events[1] = 0
    if(tiedZeroDM) events[2] = 0
    if(tiedSurvLR) events[1] = 0
    if(tiedSurvDM) events[2] = 0
    eventTimes <- sort(c(0, times[events==1]))
    tstart <- c(tstart, eventTimes[-length(eventTimes)])
    tstop <- c(tstop, eventTimes[-1])
  }
}

```

```

}
rows <- 1 + (checkedData$event_LR & checkedData$time_LR!=0) +
           (checkedData$event_DM &
            checkedData$time_DM!=0 &
            checkedData$time_DM!=checkedData$time_LR)
d <- checkedData[rep(1:n, rows),]

death <- as.numeric(tstop==d$time_surv & d$event_surv)
LR <- as.numeric(tstop>d$time_LR & d$event_LR)
DM <- as.numeric(tstop>d$time_DM & d$event_DM)

longCoxData <- data.frame(id=d$id,
                          center=d$center,
                          age=d$age, depth=d$depth,
                          size.class=d$size.class,
                          size.cm=d$size.cm,
                          RT=d$RT, #chemo=d$chemo,
                          sex=d$sex,
                          margin.class=d$margin.class,
                          limb_sparing=d$limb_sparing,
                          presentation=d$presentation,
                          localization=d$localization,
                          extremity=d$extremity,
                          type=d$type,
                          tstart=tstart, tstop=tstop, LR=LR, DM=DM, death=death)

return(longCoxData)
}

longCoxData <- longCox(data)

#Model selection
###full model

full <- coxph(Surv(tstart,tstop,death)~center+age+depth+size.cm+RT+
              margin.class+limb_sparing+presentation+extremity+
              type+LR+DM, data=longCoxData)
cox.zph(full)
#Test proportional hazards assumption and violation (0.01) for RT.

full <- coxph(Surv(tstart,tstop,death)~center+age+depth+size.cm+strata(RT)+
              margin.class+limb_sparing+presentation+extremity+
              type+LR+DM, data=longCoxData)
cox.zph(full)
summary(full)
#no violation (0.01) anymore and reasonable hrs and cis.

###full, no DM
noDM <- coxph(Surv(tstart,tstop,death)~center+age+depth+size.cm+RT+
              margin.class+limb_sparing+presentation+extremity+
              type+LR, data=longCoxData)

cox.zph(noDM)
summary(noDM)

```

```

#Violation of proportionality (0.01) for: RT and type sarcoma nos. I stratify by RT.
noDM <- coxph(Surv(tstart,tstop,death)~center+age+depth+size.cm+
             margin.class+limb_sparing+presentation+extremity+
             type+LR+strata(RT), data=longCoxData)

cox.zph(noDM)
summary(noDM)

#Model selection with aic only:
##Model with DM
###Step 1
#remove depth
lrtData <- na.omit(longCoxData[,-c(5,8,12)])#-size.class,sex,localization
full <- coxph(Surv(tstart,tstop,death)~., data=lrtData[,-1])
m1 <- coxph(Surv(tstart,tstop,death)~.-age, data=lrtData[,-1])
m2 <- coxph(Surv(tstart,tstop,death)~.-depth, data=lrtData[,-1])
m3 <- coxph(Surv(tstart,tstop,death)~.-size.cm, data=lrtData[,-1])
m4 <- coxph(Surv(tstart,tstop,death)~.-RT, data=lrtData[,-1])
m5 <- coxph(Surv(tstart,tstop,death)~.-margin.class, data=lrtData[,-1])
m6 <- coxph(Surv(tstart,tstop,death)~.-limb_sparing, data=lrtData[,-1])
m7 <- coxph(Surv(tstart,tstop,death)~.-presentation, data=lrtData[,-1])
m8 <- coxph(Surv(tstart,tstop,death)~.-extremity, data=lrtData[,-1])
m9 <- coxph(Surv(tstart,tstop,death)~.-type, data=lrtData[,-1])
m10 <- coxph(Surv(tstart,tstop,death)~.-LR, data=lrtData[,-1])
m11 <- coxph(Surv(tstart,tstop,death)~.-DM, data=lrtData[,-1])
l <- lapply(list(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m2)#not significant

###Step 2
#Remove presentation.
m21 <- coxph(Surv(tstart,tstop,death)~.-depth-age, data=lrtData[,-1])
m22 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation, data=lrtData[,-1])
m23 <- coxph(Surv(tstart,tstop,death)~.-depth-size.cm, data=lrtData[,-1])
m24 <- coxph(Surv(tstart,tstop,death)~.-depth-RT, data=lrtData[,-1])
m25 <- coxph(Surv(tstart,tstop,death)~.-depth-limb_sparing, data=lrtData[,-1])
m26 <- coxph(Surv(tstart,tstop,death)~.-depth-extremity, data=lrtData[,-1])
m27 <- coxph(Surv(tstart,tstop,death)~.-depth-type, data=lrtData[,-1])
m28 <- coxph(Surv(tstart,tstop,death)~.-depth-LR, data=lrtData[,-1])
m29 <- coxph(Surv(tstart,tstop,death)~.-depth-DM, data=lrtData[,-1])
l <- lapply(list(m21,m22,m23,m24,m25,m26,m27,m28,m29),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m22)#not significant

###Step 3
#Remove limb_sparing.
m31 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-age,
             data=lrtData[,-1])
m32 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity,
             data=lrtData[,-1])
m33 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-size.cm,
             data=lrtData[,-1])

```

```

m34 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-RT,
             data=lrtData[,-1])
m35 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing,
             data=lrtData[,-1])
m36 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-type,
             data=lrtData[,-1])
m37 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-LR,
             data=lrtData[,-1])
m38 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-DM,
             data=lrtData[,-1])
l <- lapply(list(m31,m32,m33,m34,m35,m36,m37,m38,m22),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m35)#not significant

###Step 4
#Remove extremity.
m41 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-age,
             data=lrtData[,-1])
m42 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-size.cm,
             data=lrtData[,-1])
m43 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-RT,
             data=lrtData[,-1])
m44 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-extremity,
             data=lrtData[,-1])
m45 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-type,
             data=lrtData[,-1])
m46 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-LR,
             data=lrtData[,-1])
m47 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-DM,
             data=lrtData[,-1])
l <- lapply(list(m41,m42,m43,m44,m45,m46,m47,m35),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m44)#not significant

###Step 5
#final model is model of step 4.
m51 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-age, data=lrtData[,-1])
m52 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-size.cm, data=lrtData[,-1])
m53 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-RT, data=lrtData[,-1])
m54 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-type, data=lrtData[,-1])
m55 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-LR, data=lrtData[,-1])
m56 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing
             -extremity-DM, data=lrtData[,-1])
l <- lapply(list(m51,m52,m53,m54,m55,m56,m44),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)#smallest aic is from step 4

```

```

###proportional hazards assumption
#Violation of proportionality (0.01) for RT. Therefore stratify by it.
final.full <- coxph(Surv(tstart,tstop,death)~center+age+size.cm+RT+
  margin.class+type+LR+DM, data=longCoxData[,-1])

cox.zph(final.full)
summary(final.full)
final.full <- coxph(Surv(tstart,tstop,death)~center+age+size.cm+
  margin.class+type+LR+DM+strata(RT),
  data=longCoxData[,-1])

cox.zph(final.full)
summary(final.full)

##Model without DM
###Step 1
#Remove presentation.
lrtData <- na.omit(longCoxData[,-c(5,8,12)])#-size.class,sex,localization
full <- coxph(Surv(tstart,tstop,death)~.-DM, data=lrtData[,-1])
m1 <- coxph(Surv(tstart,tstop,death)~.-DM-age, data=lrtData[,-1])
m2 <- coxph(Surv(tstart,tstop,death)~.-DM-depth, data=lrtData[,-1])
m3 <- coxph(Surv(tstart,tstop,death)~.-DM-size.cm, data=lrtData[,-1])
m4 <- coxph(Surv(tstart,tstop,death)~.-DM-RT, data=lrtData[,-1])
m5 <- coxph(Surv(tstart,tstop,death)~.-DM-limb_sparing, data=lrtData[,-1])
m6 <- coxph(Surv(tstart,tstop,death)~.-DM-presentation, data=lrtData[,-1])
m7 <- coxph(Surv(tstart,tstop,death)~.-DM-extremity, data=lrtData[,-1])
m8 <- coxph(Surv(tstart,tstop,death)~.-DM-type, data=lrtData[,-1])
m9 <- coxph(Surv(tstart,tstop,death)~.-DM-LR, data=lrtData[,-1])
l <- lapply(list(m1,m2,m3,m4,m5,m6,m7,m8,m9,full),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m6)#not significant

###Step 2
#Remove depth.
m21 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-age, data=lrtData[,-1])
m22 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-depth,
  data=lrtData[,-1])
m23 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-size.cm, data=lrtData[,-1])
m24 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-RT, data=lrtData[,-1])
m25 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-limb_sparing,
  data=lrtData[,-1])
m26 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-extremity,
  data=lrtData[,-1])
m27 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-type, data=lrtData[,-1])
m28 <- coxph(Surv(tstart,tstop,death)~.-presentation-DM-LR, data=lrtData[,-1])
l <- lapply(list(m21,m22,m23,m24,m25,m26,m27,m28,m6),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m22)#not significant

###Step 3
#Remove limb_sparing.
m31 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM-age,
  data=lrtData[,-1])

```

```

m32 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM-size.cm,
             data=lrtData[,-1])
m33 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM-RT,
             data=lrtData[,-1])
m34 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM
             -limb_sparing, data=lrtData[,-1])
m35 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM
             -extremity, data=lrtData[,-1])
m36 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM-type,
             data=lrtData[,-1])
m37 <- coxph(Surv(tstart,tstop,death)~.-presentation-depth-DM-LR,
             data=lrtData[,-1])
l <- lapply(list(m31,m32,m33,m34,m35,m36,m37,m22),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m34)#not significant

###Step 4
#Remove extremity.
m41 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-age, data=lrtData[,-1])
m42 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-size.cm, data=lrtData[,-1])
m43 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-RT, data=lrtData[,-1])
m44 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-extremity, data=lrtData[,-1])
m45 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-type, data=lrtData[,-1])
m46 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-limb_sparing-
             DM-LR, data=lrtData[,-1])
l <- lapply(list(m41,m42,m43,m44,m45,m46,m34),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)
anova(full,m44)#not significant

###Step 5
#smallest aic is model of step 4
m51 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity-limb_sparing-DM-age,
             data=lrtData[,-1])
m52 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity-limb_sparing-DM-size.cm,
             data=lrtData[,-1])
m53 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity-limb_sparing-DM-RT,
             data=lrtData[,-1])
m54 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity-limb_sparing-DM
             -type, data=lrtData[,-1])
m55 <- coxph(Surv(tstart,tstop,death)~.-depth-presentation-extremity-limb_sparing-DM-LR,
             data=lrtData[,-1])
l <- lapply(list(m51,m52,m53,m54,m55,m44),extractAIC)
aic <- lapply(l,function(x)tail(x,1))
which.min(aic)# smallest aic is model of step 4

###proportional hazards assumption

```

```

#Violation of proportionality for RT, so stratify by it.
final.noDM <- coxph(Surv(tstart,tstop,death)~center+age+size.cm+RT+
                    margin.class+type+LR, data=longCoxData[,-1])
cox.zph(final.noDM)
final.noDM <- coxph(Surv(tstart,tstop,death)~center+age+size.cm+
                    strata(RT)+margin.class+type+LR,
                    data=longCoxData[,-1])
cox.zph(final.noDM)
summary(final.noDM)

```

Multi-state analysis

```

load("MergedData.Rda")
data<-na.omit(data)

longMstate <- function(wideData){
  # creates long format to do a multi-state analysis.
  # first censors LR if LR and DM happened at the same time
  wideData$event_LR[(wideData$time_LR==wideData$time_DM) &
                    wideData$event_LR & wideData$event_DM]=0
  # Now censors LR/DM when at the same time as death:
  wideData$event_DM[wideData$time_DM == wideData$time_surv &
                    wideData$event_DM &
                    wideData$event_surv] = 0
  wideData$event_LR[wideData$time_LR == wideData$time_surv &
                    wideData$event_LR &
                    wideData$event_surv] = 0

  tmat<-transMat(x=list(c(2,3,4),c(3,4),c(4),c()),names=c("OK","LR","DM","Death"))
  covs=c("id","center","age","depth", "size.class", "size.cm","RT",
         "margin.class", "limb_sparing", "presentation", "localization",
         "extremity","sex", "type","time_LR","time_DM")

  longMstateData <- msprep(data = wideData, trans = tmat,
                          time = c(NA, "time_LR", "time_DM", "time_surv"),
                          status = c(NA, "event_LR", "event_DM", "event_surv"),
                          keep = covs)

  return(longMstateData)
}
longMstateData <- longMstate(data)
longMstateData <- longMstateData[longMstateData$Tstart<longMstateData$Tstop,]
covs=c("id","center","age", "depth", "size.cm", "RT",
      "margin.class","limb_sparing","presentation",
      "extremity", "type", "sex","time_LR","time_DM")
expandedData<-expand.covs(longMstateData,covs,longnames=F)

#All transitions in one model
#only violation for transition 3: depth (0.002) and RT (0.001)
#now stratified by RT.3 and depth.3
#no limb_sparing in trans5

```

```

expandedData$depth.3 <- ifelse(expandedData$depth1.3,1,ifelse(expandedData$depth2.3,2,0))
#create stratification variable
expandedData$str <- interaction(expandedData$trans,
                               expandedData$RT.3,
                               expandedData$depth.3)
c.final <- coxph(Surv(Tstart,Tstop,status)~#trans1
                center+
                age.1+depth1.1+depth2.1+size.cm.1+RT.1+
                margin.class1.1+margin.class2.1+limb_sparing.1+presentation1.1+
                presentation2.1+extremity.1+
                type1.1+type2.1+type3.1+type4.1+
                type5.1+type6.1+
                #trans2
                age.2+depth1.2+depth2.2+size.cm.2+RT.2+
                margin.class1.2+margin.class2.2+limb_sparing.2+presentation1.2+
                presentation2.2+extremity.2+
                type1.2+type2.2+type3.2+type4.2+
                type5.2+type6.2+
                #trans3
                age.3+size.cm.3+
                margin.class1.3+margin.class2.3+limb_sparing.3+presentation1.3+
                presentation2.3+extremity.3+
                type1.3+type2.3+type3.3+type4.3+
                type5.3+type6.3+
                #trans4
                age.4+size.cm.4+RT.4+
                margin.class1.4+margin.class2.4+presentation1.4+
                presentation2.4+extremity.4+
                #trans5
                age.5+size.cm.5+RT.5+
                margin.class1.5+margin.class2.5+#limb_sparing.5+
                presentation1.5+
                presentation2.5+extremity.5+
                #trans6
                age.6+depth1.6+depth2.6+size.cm.6+RT.6+
                margin.class1.6+margin.class2.6+limb_sparing.6+extremity.6+
                strata(str)
                ,data=expandedData)

cox.zph(c.final)
summary(c.final)

#MARKOV ASSUMPTION
#Because center is chosen not transition specific tests for
#proportionality and markov must be done in full model
#Markov assumption transition 4
markov.4 <- coxph(Surv(Tstart,Tstop,status)~#trans1
                 center+
                 age.1+depth1.1+depth2.1+size.cm.1+RT.1+
                 margin.class1.1+margin.class2.1+limb_sparing.1+presentation1.1+
                 presentation2.1+extremity.1+
                 type1.1+type2.1+type3.1+type4.1+
                 type5.1+type6.1+
                 #trans2

```



```

age.2+depth1.2+depth2.2+size.cm.2+RT.2+
margin.class1.2+margin.class2.2+limb_sparing.2+presentation1.2+
presentation2.2+extremity.2+
type1.2+type2.2+type3.2+type4.2+
type5.2+type6.2+
#trans3
age.3+size.cm.3+
margin.class1.3+margin.class2.3+limb_sparing.3+presentation1.3+
presentation2.3+extremity.3+
type1.3+type2.3+type3.3+type4.3+
type5.3+type6.3+
#trans4
age.4+size.cm.4+RT.4+
margin.class1.4+margin.class2.4+presentation1.4+
presentation2.4+extremity.4+
#trans5
age.5+size.cm.5+RT.5+
margin.class1.5+margin.class2.5+#limb_sparing.5+
presentation1.5+
presentation2.5+extremity.5+
#trans6
age.6+depth1.6+depth2.6+size.cm.6+RT.6+
margin.class1.6+margin.class2.6+limb_sparing.6+extremity.6+
strata(str)+time_LR.4
,data=expandedData)
summary(markov.4)
#Markov assumption transition 5
markov.5 <- coxph(Surv(Tstart,Tstop,status)~#trans1
center+
age.1+depth1.1+depth2.1+size.cm.1+RT.1+
margin.class1.1+margin.class2.1+limb_sparing.1+presentation1.1+
presentation2.1+extremity.1+
type1.1+type2.1+type3.1+type4.1+
type5.1+type6.1+
#trans2
age.2+depth1.2+depth2.2+size.cm.2+RT.2+
margin.class1.2+margin.class2.2+limb_sparing.2+presentation1.2+
presentation2.2+extremity.2+
type1.2+type2.2+type3.2+type4.2+
type5.2+type6.2+
#trans3
age.3+size.cm.3+
margin.class1.3+margin.class2.3+limb_sparing.3+presentation1.3+
presentation2.3+extremity.3+
type1.3+type2.3+type3.3+type4.3+
type5.3+type6.3+
#trans4
age.4+size.cm.4+RT.4+
margin.class1.4+margin.class2.4+presentation1.4+
presentation2.4+extremity.4+
#trans5
age.5+size.cm.5+RT.5+
margin.class1.5+margin.class2.5+#limb_sparing.5+

```

```

presentation1.5+
presentation2.5+extremity.5+
#trans6
age.6+depth1.6+depth2.6+size.cm.6+RT.6+
margin.class1.6+margin.class2.6+limb_sparing.6+extremity.6+
strata(str)+time_LR.5
, data=expandedData)

summary(markov.5)
#Markov assumption transition 6
markov.6 <- coxph(Surv(Tstart,Tstop,status)~#trans1
center+
age.1+depth1.1+depth2.1+size.cm.1+RT.1+
margin.class1.1+margin.class2.1+limb_sparing.1+presentation1.1+
presentation2.1+extremity.1+
type1.1+type2.1+type3.1+type4.1+
type5.1+type6.1+
#trans2
age.2+depth1.2+depth2.2+size.cm.2+RT.2+
margin.class1.2+margin.class2.2+limb_sparing.2+presentation1.2+
presentation2.2+extremity.2+
type1.2+type2.2+type3.2+type4.2+
type5.2+type6.2+
#trans3
age.3+size.cm.3+
margin.class1.3+margin.class2.3+limb_sparing.3+presentation1.3+
presentation2.3+extremity.3+
type1.3+type2.3+type3.3+type4.3+
type5.3+type6.3+
#trans4
age.4+size.cm.4+RT.4+
margin.class1.4+margin.class2.4+presentation1.4+
presentation2.4+extremity.4+
#trans5
age.5+size.cm.5+RT.5+
margin.class1.5+margin.class2.5+#limb_sparing.5+
presentation1.5+
presentation2.5+extremity.5+
#trans6
age.6+depth1.6+depth2.6+size.cm.6+RT.6+
margin.class1.6+margin.class2.6+limb_sparing.6+extremity.6+
strata(str)+time_DM.6
, data=expandedData)

summary(markov.6)

#Prediction
##Create new data
newData <- function(center=0,age=0,depth=0,size.cm=0,RT=0,margin.class=0,
limb_sparing=0,presentation=0,extremity=0,
type=0,sex=0){
data <- data.frame(trans=1:6,age.1=0,age.2=0,age.3=0,age.4=0,age.5=0,age.6=0,
depth1.1=0,depth1.2=0,depth1.3=0,depth1.4=0,depth1.5=0,depth1.6=0,
depth2.1=0,depth2.2=0,depth2.3=0,depth2.4=0,depth2.5=0,depth2.6=0,
depth.3=0,

```

```

size.cm.1=0,size.cm.2=0,size.cm.3=0,size.cm.4=0,
size.cm.5=0,size.cm.6=0,
RT.1=0,RT.2=0,RT.3=0,RT.4=0,RT.5=0,RT.6=0,
margin.class1.1=0,margin.class1.2=0,margin.class1.3=0,margin.class1.4=0,
margin.class1.5=0,margin.class1.6=0,margin.class2.1=0,margin.class2.2=0,
margin.class2.3=0,margin.class2.4=0,margin.class2.5=0,margin.class2.6=0,
limb_sparing.1=0,limb_sparing.2=0,limb_sparing.3=0,limb_sparing.4=0,
limb_sparing.5=0,limb_sparing.6=0,
presentation1.1=0,presentation1.2=0,presentation1.3=0,presentation1.4=0,
presentation1.5=0,presentation1.6=0,presentation2.1=0,presentation2.2=0,
presentation2.3=0,presentation2.4=0,presentation2.5=0,presentation2.6=0,
extremity.1=0,extremity.2=0,extremity.3=0,extremity.4=0,
extremity.5=0,extremity.6=0,
type1.1=0,type1.2=0,type1.3=0,type1.4=0,type1.5=0,type1.6=0,type2.1=0,
type2.2=0,type2.3=0,type2.4=0,type2.5=0,type2.6=0,type3.1=0,type3.2=0,
type3.3=0,type3.4=0,type3.5=0,type3.6=0,type4.1=0,type4.2=0,type4.3=0,
type4.4=0,type4.5=0,type4.6=0,type5.1=0,type5.2=0,type5.3=0,type5.4=0,
type5.5=0,type5.6=0,type6.1=0,type6.2=0,type6.3=0,type6.4=0,type6.5=0,
type6.6=0,
sex.1=0,sex.2=0,sex.3=0,sex.4=0,sex.5=0,sex.6=0,
center1.1=0,center1.2=0,center1.3=0,
center1.4=0,center1.5=0,center1.6=0,
center2.1=0,center2.2=0,center2.3=0,
center2.4=0,center2.5=0,center2.6=0,
center3.1=0,center3.2=0,center3.3=0,
center3.4=0,center3.5=0,center3.6=0,
strata=1:6)
p <- as.matrix(data)
diag(p[,2:7]) <- age
if(depth==1) diag(p[,8:13]) <- 1
if(depth==2) diag(p[,14:19]) <- 1
#eine spalte für depth.3
diag(p[,21:26]) <- size.cm
if(RT==1) diag(p[,27:32]) <- 1
if(margin.class==1) diag(p[,33:38]) <- 1
if(margin.class==2) diag(p[,39:44]) <- 1
if(limb_sparing==1) diag(p[,45:50]) <- 1
if(presentation==1) diag(p[,51:56]) <- 1
if(presentation==2) diag(p[,57:62]) <- 1
if(extremity==1) diag(p[,63:68]) <- 1
if(type==1) diag(p[,69:74]) <- 1
if(type==2) diag(p[,75:80]) <- 1
if(type==3) diag(p[,81:86]) <- 1
if(type==4) diag(p[,87:92]) <- 1
if(type==5) diag(p[,93:98]) <- 1
if(type==6) diag(p[,99:104]) <- 1
if(sex==1) diag(p[,105:110]) <- 1
if(center==1) diag(p[,111:116]) <- 1
if(center==2) diag(p[,117:122]) <- 1
if(center==3) diag(p[,123:128]) <- 1
data <- as.data.frame(p)
data$depth.3 <- ifelse(data$depth1.3,1,ifelse(data$depth2.3,2,0))
data$strata <- interaction(data$trans,data$depth.3,data$RT.3)

```

```

  ifelse(center==0, data$center <- "Leiden",
    ifelse(center==1, data$center <- "London",
      ifelse(center==2, data$center <- "Birmingham",
        data$center <- "Toronto")))
  return(data)
}

p1 <- newData(center=0, age=50, depth=0, size.cm=5, RT=0, margin.class=0,
  limb_sparing=0, presentation=0, extremity=0,
  type=0, sex=0)
p2 <- newData(center=0, age=50, depth=0, size.cm=5, RT=0, margin.class=1,
  limb_sparing=0, presentation=0, extremity=0,
  type=0, sex=0)
p3 <- newData(center=0, age=50, depth=0, size.cm=5, RT=0, margin.class=2,
  limb_sparing=0, presentation=0, extremity=0,
  type=0, sex=0)

ord <- c(4:1)
tmat<-transMat(x=list(c(2,3,4),c(3,4),c(4),c()),names=c("AWD","LR","DM","Death"))
par(mfrow=c(2,2))
msfitcox1<-msfit(object=c.final,newdata=p1,vartype="aalen",trans=tmat )
ptcox1 <- probtrans(msfitcox1,pretd = 0, method = "aalen")
plot(ptcox1, main="Intralesional margin",type="filled", ord=ord, xlab="Time since Surgery")

msfitcox2<-msfit(object=c.final,newdata=p2,vartype="aalen",trans=tmat )
ptcox2 <- probtrans(msfitcox2,pretd = 0, method = "aalen")
plot(ptcox2, main="Marginal margin",type="filled", ord=ord, xlab="Time since Surgery")
par(mfrow=c(1,1))

```

Website

Models

```

load("MergedData.Rda")
data<-na.omit(data)
#Add age above median indicator for website:
m <- median(data$age)
data$age.class <- factor(ifelse(data$age>m,1,0), labels=c("<=median",">median"))

longCox <- function(checkedData){
# transformes data to long format for coxph with time dependant covariates
# 1. censure events when at the same time as death:
  checkedData$event_DM[checkedData$time_DM == checkedData$time_surv &
    checkedData$event_DM &
    checkedData$event_surv] = 0
  checkedData$event_LR[checkedData$time_LR == checkedData$time_surv &
    checkedData$event_LR &
    checkedData$event_surv] = 0

  n <- nrow(checkedData)
  tstart <- tstop <- NULL
  for(i in 1:n){

```

```

times <- c(checkedData$time_LR[i],
          checkedData$time_surv[i])
events <- c(checkedData$event_LR[i], 1)
tiedZeroLR <- (checkedData$time_LR[i]==0 &
              checkedData$event_LR[i])
tiedSurvLR <- (checkedData$time_LR[i]==checkedData$time_surv[i] &
              checkedData$event_LR[i])
if(tiedZeroLR) events[1] = 0
if(tiedSurvLR) events[1] = 0
eventTimes <- sort(c(0, times[events==1]))
tstart <- c(tstart, eventTimes[-length(eventTimes)])
tstop <- c(tstop, eventTimes[-1])
}
rows <- 1 + (checkedData$event_LR & checkedData$time_LR!=0)
d <- checkedData[rep(1:n, rows),]

death <- as.numeric(tstop==d$time_surv & d$event_surv)
LR <- as.numeric(tstop>d$time_LR & d$event_LR)

longCoxData <- data.frame(id=d$id, age=d$age,
                          sex=d$sex, depth=d$depth,
                          size.class=d$size.class,
                          RT=d$RT, #chemo=d$chemo,
                          margin.class=d$margin.class,
                          limb_sparing=d$limb_sparing,
                          presentation=d$presentation,
                          localization=d$localization,
                          type=d$type,
                          tstart=tstart, tstop=tstop, LR=LR, death=death)

return(longCoxData)
}
longCoxData <- longCox(data)

#Survival Time-fixed analysis

#Test propotional hazards assumption
m <- coxph(Surv(time_surv,event_surv)~age.class+sex+size.class+margin.class+type, data=data)
cox.zph(m)
#violation for margin, stratify
m <- coxph(Surv(time_surv,event_surv)~age.class+sex+size.class+strata(margin.class)+type, data=data)
cox.zph(m)

##Prediction
p1 <- data.frame(age.class="<=median", sex="male",
                 size.class="0-49.9",
                 margin.class="intralesional",
                 type="angiosarcoma", LR=0)
p2 <- data.frame(age.class=">median", sex="male",
                 size.class="0-49.9",
                 margin.class=">2mm",
                 type="angiosarcoma",LR=0)
par(mfrow=c(1,2))
plot (survfit(m,p1), main="Young patient", xlab="Time since Surgery", col=1:3)

```

```

legend("bottomleft", c("Intralesional", "Marginal", "Wide"), col=1:3, lwd=5)
plot (survfit(m,p2),main="Older patient", xlab="Time since Surgery", col=1:3)
legend("bottomleft", c("Intralesional", "Marginal", "Wide"), col=1:3, lwd=5)

OSeestimate <- function(model,patient, months){
  # get estimates for os, difficult
  # because of stratification
  s <- survfit(model,patient)
  margin <- ifelse(patient$margin.class=="intralesional",1,
                  ifelse(patient$margin.class=="0.1-2mm",2,3))
  index <- cumsum(c(0,s$strata))[margin:(margin+1)]
  t <- which.max(which(months-s$time[(index[1]+1):index[2]]>=0))+index[1]
  estimate <- c(s$surv[t], s$lower[t], s$upper[t])
  return(estimate)
}

#LR in competing risk
longMstate <- function(wideData){
  # creates long format to do a multi-state analysis.
  # first censors Death if LR and Death happened at the same time
  wideData$event_surv[(wideData$time_LR==wideData$time_surv) &
                      wideData$event_LR & wideData$event_surv]=0
  tmat <- transMat(x=list(c(2,3),c(),c()),names=c("OK","LR","Death"))
  covs=c("id","age.class","sex", "size.class", "margin.class", "type","RT")

  longMstateData <- msprep(data = wideData, trans = tmat,
                          time = c(NA, "time_LR", "time_surv"),
                          status = c(NA, "event_LR", "event_surv"),
                          keep = covs)
  return(longMstateData)
}

longMstateData <- longMstate(data)
tmat <- transMat(x=list(c(2,3),c(),c()),names=c("AWD","LR","Death"))
covs <- c("id","age.class","sex", "size.class", "margin.class", "type","RT")
expandedData <- expand.covs(longMstateData, covs, longnames=FALSE)

#Fit a model for which all covariates are transition specific.
c.final <- coxph(Surv(time,status)~age.class.1+age.class.2+sex.1+sex.2+
                size.class1.1+size.class1.2+size.class2.1+size.class2.2+
                margin.class1.1+margin.class1.2+
                margin.class2.1+margin.class2.2+
                type1.1+type1.2+type2.1+type2.2+
                type3.1+type3.2+type4.1+type4.2+
                type5.1+type5.2+type6.1+type6.2+
                strata(trans), data=expandedData)

##Test proportional hazards assumption
cox.zph(c.final)#no violation (0.01)

##Prediction for New Data
newData <- function(age.class=0,sex=0,size.class=0,margin.class=0,type=0){
  data <- data.frame(trans=1:2, age.class.1=0,age.class.2=0,sex.1=0,sex.2=0,
                    size.class1.1=0,size.class1.2=0,size.class2.1=0,size.class2.2=0,

```

```

margin.class1.1=0,margin.class1.2=0,margin.class2.1=0,margin.class2.2=0,
type1.1=0,type1.2=0,type2.1=0,type2.2=0,type3.1=0,type3.2=0,type4.1=0,
type4.2=0,type5.1=0,type5.2=0,type6.1=0,type6.2=0,
strata=1:2)
p <- as.matrix(data)
diag(p[,2:3]) <- age.class # age
diag(p[,4:5]) <- sex # female
if(size.class==1) diag(p[,6:7]) <- 1
if(size.class==2) diag(p[,8:9]) <- 1
if(margin.class==1) diag(p[,10:11]) <- 1
if(margin.class==2) diag(p[,12:13]) <- 1
if(type==1) diag(p[,14:15]) <- 1
if(type==2) diag(p[,16:17]) <- 1
if(type==3) diag(p[,18:19]) <- 1
if(type==4) diag(p[,20:21]) <- 1
if(type==5) diag(p[,22:23]) <- 1
if(type==6) diag(p[,24:25]) <- 1
p <- as.data.frame(p)
return(p)
}

par(mfrow=c(1,2))
ord<-c(3,2,1)
p1 <- newData(age.class=0)
p1.0<-msfit(object=c.final,newdata=p1,vartype= "aalen",trans=tmat)
pt1.0<-probrans(p1.0, predt = 0, method = "aalen")
plot(pt1.0,ord=ord,type="filled",las=1,main="margin:intralesional",
      xlab="Months since surgery",xlim=c(0,120))

p2 <- newData(age.class=0, margin.class=2)
p2.0<-msfit(object=c.final,newdata=p2,vartype= "aalen",trans=tmat)
pt2.0<-probrans(p2.0, predt = 0, method = "aalen")
plot(pt2.0,ord=ord,type="filled",las=1,main="margin:wide", xlab="Months since surgery",xlim=c(0,120))

#five year cumulative incidence of LR:
#probability from 1. state to second state at 5 years general
rbind(pt1.0[[1]]$pstate2[which.max(which(60-pt1.0[[1]]$time>=0))],
pt2.0[[1]]$pstate2[which.max(which(60-pt1.0[[1]]$time>=0))])

#10 cum inc. problem: no event at exactly 10 years
pt1.0[[1]]$pstate2[which.max(which(120-pt1.0[[1]]$time>=0))]

#get CI:
index <- which.max(which(60-pt1.0[[1]]$time>=0))
est <- pt1.0[[1]]$pstate2[index]
se <- pt1.0[[1]]$se2[index]
c(est-1.959964*se,est+1.959964*se)

#cumulative incidence of death 5 yrs:
rbind(pt1.0[[1]]$pstate3[pt1.0[[1]]$time==60],
pt2.0[[1]]$pstate3[pt2.0[[1]]$time==60])

```

Server script

```
library(shiny)
library(mstate)
library(survival)
load("/Users/anjarutenbudde/Desktop/Uni Leiden/Internship/OwnRcode/MergedData.Rda")
data<-na.omit(data)
#Add age above median indicator for website:
m <- median(data$age)
data$age.class <- factor(ifelse(data$age>m,1,0), labels=c("<=median",">median"))

#OS
m <- coxph(Surv(time_surv,event_surv)~age.class+sex+size.class+
           strata(margin.class)+type, data=data)
OSestimate <- function(model,patient, months){
  # get estimates for os, difficult
  # because of stratification
  s <- survfit(model,patient)
  margin <- ifelse(patient$margin.class=="intralesional",1,
                  ifelse(patient$margin.class=="0.1-2mm",2,3))
  index <- cumsum(c(0,s$strata))[margin:(margin+1)]
  t <- which.max(which(months-s$time[(index[1]+1):index[2]]>=0))+index[1]
  estimate <- c(s$surv[t], s$lower[t], s$upper[t])
  return(estimate)
}

#comp risk
longMstate <- function(wideData){
  # creates long format to do a multi-state analysis.
  # first censors Death if LR and Death happened at the same time
  wideData$event_surv[(wideData$time_LR==wideData$time_surv) &
                      wideData$event_LR & wideData$event_surv]=0
  tmat <- transMat(x=list(c(2,3),c(),c()),names=c("OK","LR","Death"))
  covs=c("id","age.class","sex", "size.class", "margin.class", "type","RT")

  longMstateData <- msprep(data = wideData, trans = tmat,
                           time = c(NA, "time_LR", "time_surv"),
                           status = c(NA, "event_LR", "event_surv"),
                           keep = covs)

  return(longMstateData)
}
longMstateData <- longMstate(data)
tmat <- transMat(x=list(c(2,3),c(),c()),names=c("AWD","LR","Death"))
covs <- c("id","age.class","sex", "size.class", "margin.class", "type","RT")
expandedData <- expand.covs(longMstateData, covs, longnames=FALSE)
c.final <- coxph(Surv(time,status)~age.class.1+age.class.2+sex.1+sex.2+
                size.class1.1+size.class1.2+size.class2.1+size.class2.2+
                margin.class1.1+margin.class1.2+
                margin.class2.1+margin.class2.2+
                type1.1+type1.2+type2.1+type2.2+
                type3.1+type3.2+type4.1+type4.2+
                type5.1+type5.2+type6.1+type6.2+
                strata(trans), data=expandedData)
ord<-c(3:1)
```



```

newData <- function(age.class=0,sex=0,size.class=0,margin.class=0,type=0){
  data <- data.frame(trans=1:2, age.class.1=0,age.class.2=0,sex.1=0,sex.2=0,
                    size.class1.1=0,size.class1.2=0,size.class2.1=0,size.class2.2=0,
                    margin.class1.1=0,margin.class1.2=0,margin.class2.1=0,margin.class2.2=0,
                    type1.1=0,type1.2=0,type2.1=0,type2.2=0,type3.1=0,type3.2=0,type4.1=0,
                    type4.2=0,type5.1=0,type5.2=0,type6.1=0,type6.2=0,
                    strata=1:2)

  p <- as.matrix(data)
  diag(p[,2:3]) <- age.class # age
  diag(p[,4:5]) <- sex # female
  if(size.class==1) diag(p[,6:7]) <- 1
  if(size.class==2) diag(p[,8:9]) <- 1
  if(margin.class==1) diag(p[,10:11]) <- 1
  if(margin.class==2) diag(p[,12:13]) <- 1
  if(type==1) diag(p[,14:15]) <- 1
  if(type==2) diag(p[,16:17]) <- 1
  if(type==3) diag(p[,18:19]) <- 1
  if(type==4) diag(p[,20:21]) <- 1
  if(type==5) diag(p[,22:23]) <- 1
  if(type==6) diag(p[,24:25]) <- 1
  p <- as.data.frame(p)
  return(p)
}

# Define server logic
shinyServer(function(input, output) {
# creates patient data.frame for time-dependent analysis from user input
  p1<-reactive({
    age <- ifelse(input$age=="<=62", "<=median", ">median")
    size <- ifelse(input$size.class=="0-49.9mm", "0-49.9",
                  ifelse(input$size.class=="50-99.9mm", "50-99.9", ">=100"))
    data.frame(age.class=age, sex=input$sex,
              size.class=size, margin.class=input$margin.class,
              type=input$type)
  })
# creates patient data.frame for comp. risk analysis from user input
  p1.comp <- reactive({
    newData(age.class=ifelse(input$age=="<=62",0,1),
            sex=ifelse(input$sex=="male",0,1),
            size.class=ifelse(input$size.class=="0-49.9mm",0,
                              ifelse(input$size.class=="50-99.9mm",1,2)),
            margin.class=ifelse(input$margin.class=="intralesional",0,
                                ifelse(input$margin.class=="0.1-2mm",1,2)),
            type=ifelse(input$type=="angiosarcoma",0,
                        ifelse(input$type=="mpnst",1,
                                ifelse(input$type=="myxofibrosarcoma",2,
                                        ifelse(input$type=="synoviosarcoma",3,
                                                ifelse(input$type=="sarcoma - nos",4,
                                                        ifelse(input$type=="spindle cell sarcoma",
                                                                5,6))))))))))
  })
# OS plot

```

```

output$survivalPlot <- renderPlot({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
input$type=='')){
    plot(survfit(m,newdata=p1()),main="Overall survival",
      xlab="Months since surgery",col=1:3,xlim=c(0,120))
    legend("bottomleft", c("Intralesional", "Marginal","Wide"), col=1:3, lwd=5)
  }
})
# comp. risk plot
output$comriskPlot <- renderPlot({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
input$type=='')){
    ms.1 <- msfit(object=c.final,newdata=p1.comp(),vartype= "aalen",trans=tmat)
    pt.1<-probtrans(ms.1, predt = 0, method = "aalen")
    plot(pt.1,ord=ord,type="filled",las=1,main="Competing risk estimation",
      xlab="Months since surgery",xlim=c(0,120))
  }
})
# 5 year survival estimate
output$estimates <- renderText({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
input$type=='')){
    paste("Probability of surviving 5 years:",
      round(OSestimate(m,p1(),60)[1],3)*100,"%", 0.95 CI: ("",
      round(OSestimate(m,p1(),60)[2],3)*100,"%",",",
      round(OSestimate(m,p1(),60)[3],3)*100,"%") )
  }
})
# 5 year cum inc LR
output$estimates.comp <- renderText({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
input$type=='')){
    ms.1 <- msfit(object=c.final,newdata=p1.comp(),vartype= "aalen",trans=tmat)
    pt.1<-probtrans(ms.1, predt = 0, method = "aalen")
    t <- which.max(which(60-pt.1[[1]]$time>=0))
    est <- pt.1[[1]]$pstate2[t]
    se <- pt.1[[1]]$se2[t]
    #probability from 1. state to 2. state at 5 years (60 months)
    paste("Probability of LR within 5 years:", round(est,3)*100,"%", 0.95 CI: ("",
      max(round(est-1.959964*se,3)*100,0),",",",
      min(round(est+1.959964*se,3)*100,100),")" )
  }
})
# 10 year OS
output$estimates10 <- renderText({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
input$type=='')){
    s <- survfit(m,p1())
    t <- which.max(which(120-s$time>=0))
    paste("Probability of surviving 10 years:",
      round(OSestimate(m,p1(),120)[1],3)*100,"%", 0.95 CI: ("",
      round(OSestimate(m,p1(),120)[2],3)*100,"%",",",

```

```

        round(OSestimate(m,p1(),120)[3],3)*100,")" )
    }
  })
# 10 cum. inc. LR
output$estimates.comp10 <- renderText({
  if(!(input$sex==''|input$size.class=='|input$margin.class=='|
        input$type=='')){
    ms.1 <- msfit(object=c.final,newdata=p1.comp(),vartype= "aalen",trans=tmat)
    pt.1<-probtrans(ms.1, predt = 0, method = "aalen")
    t <- which.max(which(120-pt.1[[1]]$time>=0))
    est <- pt.1[[1]]$pstate2[t]
    se <- pt.1[[1]]$se2[t]
    #probability from 1. state to 2. state at 5 years (60 months)
    paste("Probability of LR within 10 years:", round(est,3)*100,"%", 0.95 CI: ("
          max(round(est-1.959964*se,3)*100,0),",",
          min(round(est+1.959964*se,3)*100,100),")")
  }
})

#runApp("shinyapp")
})

```

User-interface script

```

library(shiny)

# Define UI
shinyUI(fluidPage(theme="bootstrap.min.css",
  fluidRow(
    column(12,h3("Prognostic tool for highgrade soft-tissue sarcoma patients"),
      p("")),
  ),
  # Application title
  headerPanel("5 and 10 year prognosis"),

  # Sidebar with controls to select the patient characteristics
  sidebarPanel(
    h4("Enter your information"),
    #numericInput("age", "Age:", 10, min=3, max=96),
    selectizeInput("age", label = "Age:",
      choices = c("<=62", ">62"),
      options = list(
        placeholder = "Select age",
        onInitialize = I('function() { this.setValue(""); }')
      )
    ),
    selectizeInput("sex", label = "Sex:",
      choices = c("male", "female"),
      options = list(
        placeholder = "Select sex",
        onInitialize = I('function() { this.setValue(""); }')
      )
    ),
    selectizeInput("size.class", label = "Tumorsize:",

```

```

    choices = c("0-49.9mm", "50-99.9mm", ">=100mm"),
    options = list(
      placeholder = "Select tumorsize",
      onInitialize = I('function() { this.setValue(""); }')
    ),
  selectizeInput("margin.class", label = "Margin:",
    choices = c("intralesional", "0.1-2mm", ">2mm"),
    options = list(
      placeholder = "Select margin",
      onInitialize = I('function() { this.setValue(""); }')
    ),
  selectizeInput("type", label = "Tumortype:",
    choices = c("angiosarcoma", "mpnst",
      "myxofibrosarcoma", "synoviosarcoma",
      "sarcoma - nos", "spindle cell sarcoma",
      "mfh/ups"),
    options = list(
      placeholder = "Select tumortype",
      onInitialize = I('function() { this.setValue(""); }')
    ),

  submitButton("Calculate"),
  width=4
),

mainPanel(
  tabsetPanel(
    tabPanel("estimates", h4(textOutput("estimates")),
      h4(textOutput("estimates.comp")),
      h4(textOutput("estimates10")),
      h4(textOutput("estimates.comp10"))),
    tabPanel("Overall Survival", plotOutput("survivalPlot")),
    tabPanel("LR", plotOutput("comriskPlot"))
  )
),
fluidRow(
  column(12, h2("Basis of estimation"),
    p("This estimation is based on a retrospective study conducted at LUMC,
with data collected from several hospitals. The probability of overall
survival is estimated by a model with only time-fixed covariates.
The probability of LR is estimated by a model, accounting for the competing risk of death."
))
)

```