CHAPTER 6

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

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

Abstract

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

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

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

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

§6.1 Introduction

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

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

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

§6.2 Methods

§6.2.1 Study design

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

§6.2.2 Patients and variables

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

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

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

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

§6.2.3 Statistical analysis

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

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

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

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

close to one represent a good calibration.

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

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

§6.3 Results

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

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

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

Table 6.1: Patient demographics.

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

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

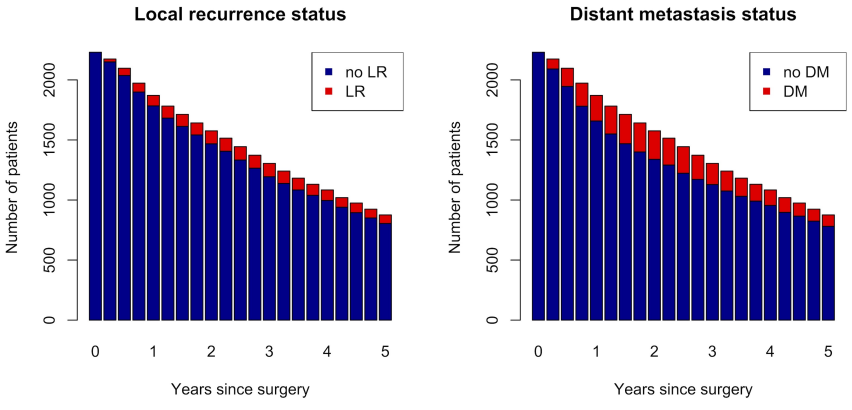

Figure 6.1: Number of patients at risk at each landmark time point t_{LM}. A) Red, patients with local recurrence; blue, patients without local recurrence. B) Red, patients with distant metastases; blue, patients without distant metastases.

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

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

Table 6.2: (continued)

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

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

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

$$\text{HR}_{lin}^{step} \times \text{HR}_{quad}^{step^2} = 1.444 \times 1.065 = 1.538$$

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

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

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

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

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

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

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

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

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

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

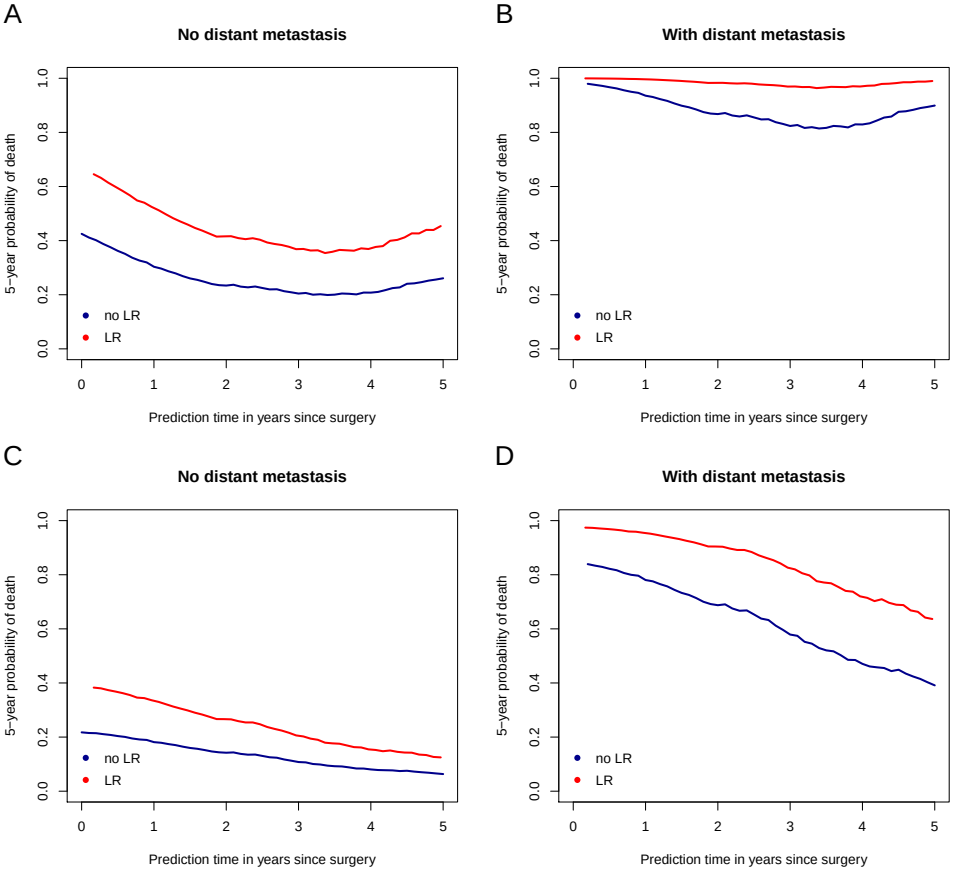

Figure 6.2: 5-year probability of death estimates for patients with different characteristics and at different states of disease progression. A and B: 61 years, tumor of 9 cm, deep myxofibrosarcoma, treated with an R0 margin and no radiotherapy. (A) Without DM at time of prediction (t_p). (B) diagnosed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

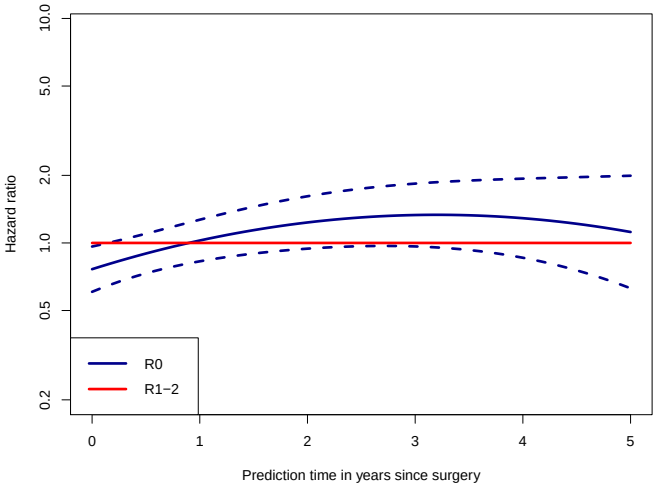

Figure 6.3: Time-varying hazard ratio for surgical margin. Blue: R0 margin; red: R1-2 margin (reference). Dashed line: pointwise confidence interval for HR of R0 margin.

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

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

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

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

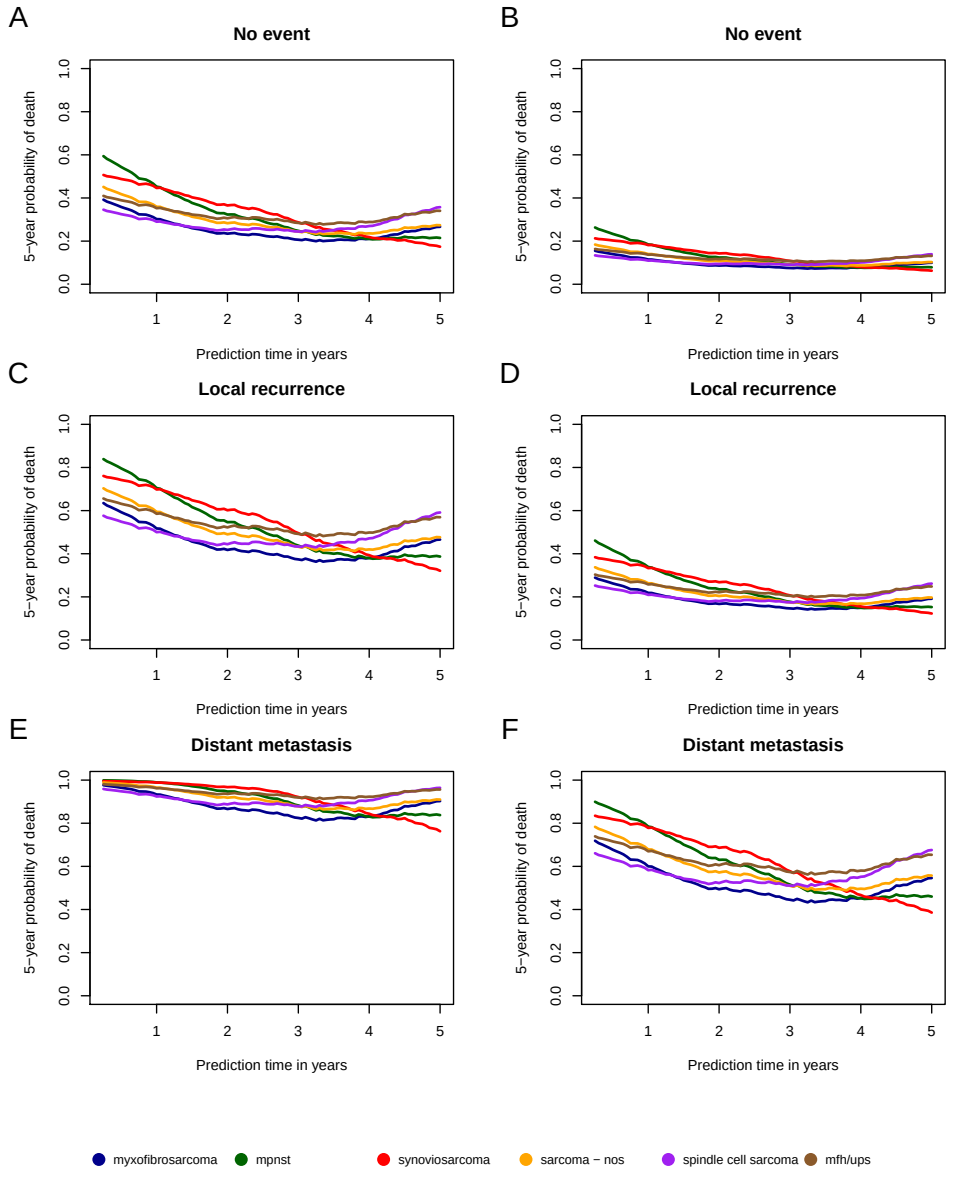

Figure 6.4: 5-year probability of death estimates for patients with different characteristics and at different states of disease progression. A, C, and E: 61 years, 9 cm deep tumor, with R0 margin and no radiotherapy. (A) Without LR or DM at time of prediction (t_p). (C) diagnosed with LR before time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

§6.4 Discussion

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

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

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

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

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