CHAPTER 7

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

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

Abstract

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

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

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

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

§7.1 Introduction

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

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

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

§7.2 Methods

§7.2.1 Study design

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

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

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

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

§7.2.2 Patients and Variables

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

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

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

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

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

§7.2.3 Statistical analysis

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

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

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

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

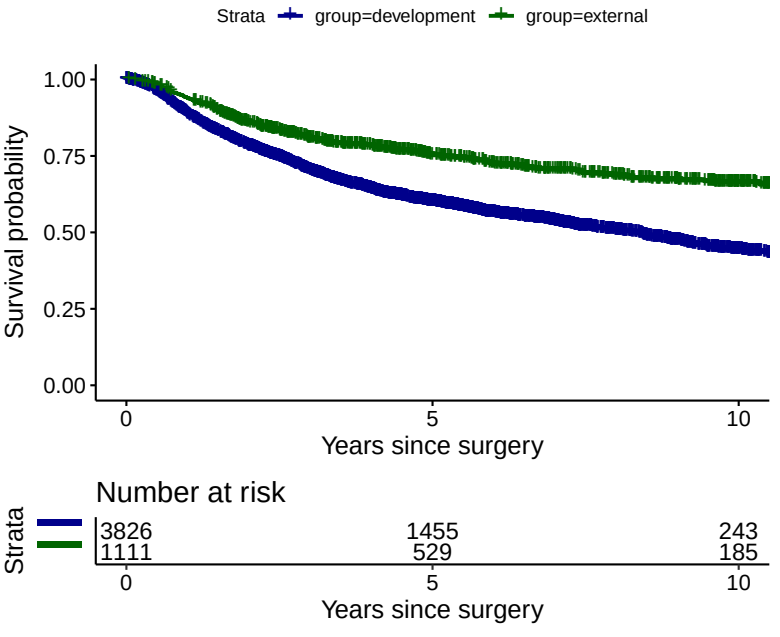

Figure 7.1: Kaplan-Meier curves for development and external cohort.

§7.3 Results

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

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

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

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

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

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

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

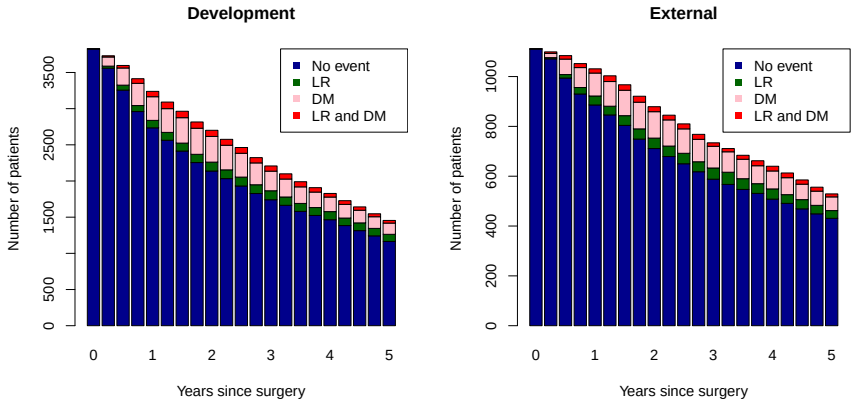

Figure 7.2: Number of patients at risk in development and external data set respectively. Red: patients with local recurrence and distant metastasis; pink: patients with distant metastasis; green: patients with local recurrence; blue: patients without local recurrence or distant metastasis.

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

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

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

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

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

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

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

Table 7.2: (continued)

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

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

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

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

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

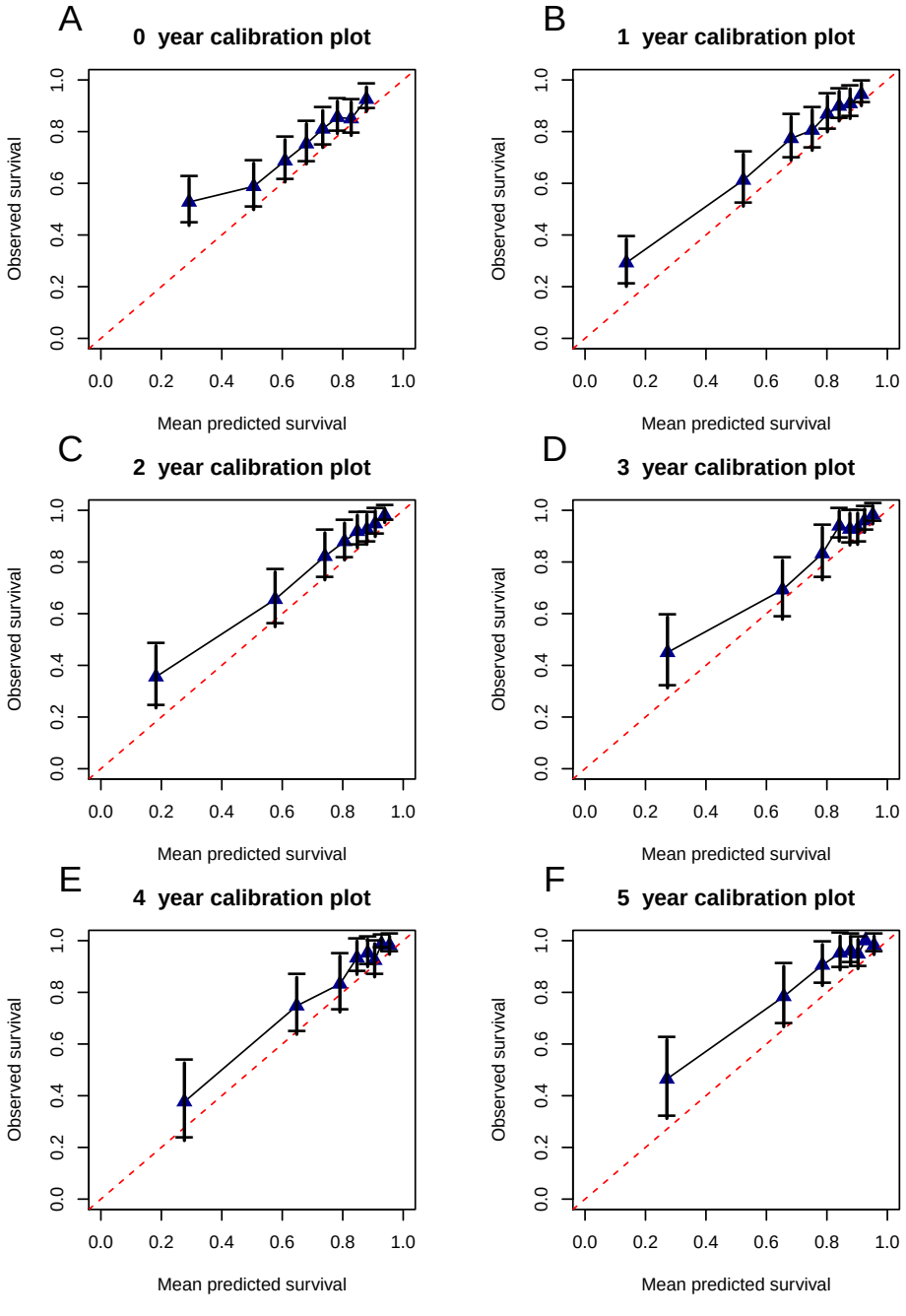

Figure 7.3: Calibration plots for predictions of 5-year DOS from 0-, 1-, 2-, 3-, 4- and 5-years post-surgery.

prognosis compared to a patient without DM.

§7.4 Discussion

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

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

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

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