Dynamic prediction in event history analysis

Grand, M.K.

1

Dynamic prediction of survival probabilities

Predictive models are an integral part of current clinical practice and help determine optimal treatment strategies for individual patients. A drawback is that covariates are assumed to have constant effects on overall survival (OS), when in fact, these effects may change during follow-up (FU). Furthermore, breast cancer (BC) patients may experience events that alter their prognosis from that time onwards. We investigated the 'dynamic' effects of different covariates on OS and developed a nomogram to calculate 5-year dynamic OS (DOS) probability at different prediction time points (t_p) during FU.

1.1 Introduction

Breast cancer (BC) comprises a heterogeneous disease with diverse features that can interact with outcomes, making it difficult to obtain estimations of individual prognoses. The overwhelming popularity of tools such as Adjuvant! or the Nottingham Prognostic Index (NPI) illustrates the importance of prediction models for physicians and patients, providing guidance for adjuvant treatment decisions[2,74]. Most prediction models, however,cannot be used for cancer patients at specific time points during the follow-up (FU) period, as these models have been designed for use immediately after diagnosis. Apart from the caveats associated with available 'static' prediction models, there are some important reasons why these models may give misleading results when used during FU. First, the fact that patients have already survived a number of years after diagnosis may change a patient's prognosis. For instance, BC recurrence rates peak at 12 years after diagnosis and decline thereafter, resulting in an improved prognosis[77,81,103]. Second, in the time between diagnosis and the moment of prediction, important events may have taken place, such as locoregional recurrence (LRR) and/or distant recurrences (DR) or premature discontinuation of treatment, which may alter a patient's prognosis. Third, some variables included in current models may exhibit time-varying effects on outcome, resulting in a change in mortality risk as time progresses. Consequently, too much emphasis may be placed on variables with a strong impact on outcome early in the FU period, whereas this effect might be much smaller later on. Available static models are based on probabilities of survival at the time of diagnosis and may not accurately portray a patient's survival probability later on in the FU period. The concept of updating survival probabilities by both incorporating time-varying covariates and allowing for time-varying effects is called dynamic prediction. By design, these variables are not included in the static risk prediction models, and these considerations illustrate a need for better prediction models for cancer patients. To investigate the clinical applicability of dynamic prediction, we utilized a dataset from a large randomized clinical trial of postmenopausal hormone receptor-positive (HR+) early BC patients treated with

endocrine treatment (ET) in the Netherlands and Belgium. The aim of the current analysis was to develop a clinically applicable nomogram to facilitate the prediction of an individual patient's probability of surviving an additional 5 years at any prediction timepoint (t_p) up to 3 years after starting adjuvant ET. This concept of continually updating 5-year overall survival (OS) from a certain t_p is referred to as 5-year dynamic overall survival (DOS). We designed a dynamic predictive model, taking into account various patient- and tumor-specific covariates with time-varying and time-constant effects during FU.

1.2 Method

The Tamoxifen Exemestane Adjuvant Multinational (TEAM) trial is a randomized, phase III, multinational, open-label study conducted in postmenopausal women with HR+ BC, who were eligible for adjuvant ET and randomized to either 5 years of exemestane (25 mg) or 2.53 years of tamoxifen (20 mg) followed by exemestane (25 mg) for 2.52 years[90]. The TEAM trial protocol was approved by regulatory and ethics authorities of all participating centers in all participating countries. The trial was registered in the Netherlands and Belgium with the Netherlands Trial register, NTR 267. All patients provided written informed consent. Details of the study and data collection have been published previously[90]. In the Netherlands and Belgium, 3168 postmenopausal, early BC patients were enrolled in the TEAM trial. Patients who did not start randomized treatment ($n = 19$) or had missing end point data ($n = 4$), metastatic disease before the start of ET ($n = 7$), and patients with missing data regarding covariates used in the model ($n = 528$) were excluded (Figure 1.1). Patients with estrogen receptor (ER) and progesterone receptor (PR)-negative disease ($n = 8$) were excluded. Due to the unavailability of regular FU data by countries other than the Netherlands and Belgium beyond the initially planned 5 years of FU, the dynamic prediction model does not include data from all participating TEAM trial countries (Table 1.1). The primary outcome of the present investigation was OS, which was the time from randomization

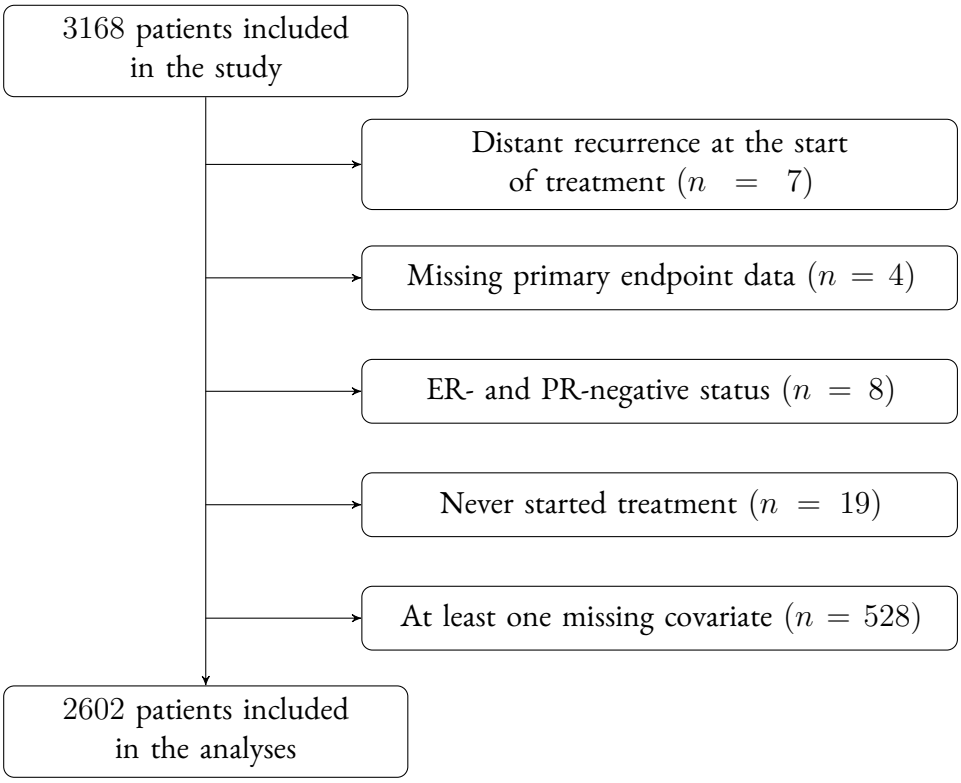

Figure 1.1: CONSORT diagram of patients included in the analyses.

to the date of death or last recorded FU. LRR was defined as any BC recurrence in the ipsilateral breast and/or lymph nodes as well as in supraclavicular lymph nodes. LRR did not include ductal carcinoma in situ relapses. DR comprised all other accounts of BC recurrence.

1.2.1 Statistical analysis

Statistical analyses were carried out using the programs SPSS (version 20) and R (version 2.15.1). We used the proportional baselines landmark super model[92,95] to obtain

Table 1.1: Comparison of the patients in the Dynamic Prediction study population with all other TEAM trial patients. BR, Bloom and Richardson.

	Non-study population ($n = 7165$)		Study population ($n = 2602$)		
Characteristics	n	(%)	n	(%)	p-value
Age at diagnosis (years) (mean, SD)	64.2	(8.9)	64.8	(9.19)	
Tumor stage					< 0.01
T1 (< 2cm)	4556	(64)	1135	(44)	
T2 ($2- < 5$cm)	2316	(32)	1276	(49)	
T3 / T4 (5cm)	266	(4)	191	(7)	
Unknown	27	(0)	0	(0)	
Nodal stage					< 0.01
No	4290	(60)	821	(32)	
N1	2514	(35)	1344	(52)	
N2/N3	295	(4)	437	(17)	
Unknown	66	(1)	0	(0)	
Histological grade (BR)					< 0.01
BR I	1295	(18)	382	(15)	
BR II	3596	(50)	1202	(46)	
BR III	1420	(20)	1018	(39)	
Unknown	854	(12)	0	(0)	
Estrogen receptor status					0.09
Negative	119	(2)	57	(2)	
Positive	7042	(98)	2545	(98)	
Unknown	4	(0)	0	(0)	
Progesterone receptor status					< 0.01
Negative	1146	(16)	579	(22)	
Positive	5278	(74)	2023	(78)	
Unknown	741	(10)	0	(0)	
HER2 status					< 0.01
Negative	3169	(44)	1898	(73)	
Positive	826	(12)	257	(10)	
Unknown	3170	(44)	447	(17)	
Most extensive surgery					< 0.01
Mastectomy	2911	(41)	1422	(55)	
Breast conserving surgery	4244	(59)	1180	(45)	
Unknown	10	(0)	0	(0)	
Radiotherapy					< 0.01
Yes	4981	(70)	1718	(66)	
No	2091	(29)	884	(34)	
Unknown	93	(1)	0	(0)	
Chemotherapy					< 0.01
Yes	2679	(37)	843	(32)	
No	4481	(63)	1759	(68)	
Unknown	5	(0)	0	(0)	

dynamic predictions of the 5-year DOS probability. The model requires a number of landmark time points (t_{LM}); in the current model t_{LM} was established at every third month between 0 and 3 years after the start of ET. A prediction model for 5-year DOS at a specific t_{LM} is constructed by selecting the individuals at risk at that t_{LM} and incorporating the values of any time-dependent covariates at that respective t_{LM} in a Cox proportional hazards model[11]. The landmark prediction models at different t_{LM}s may be combined into a single super model (Appendix 1.5). Using this analysis in the clinical setting, we can obtain DOS predictions at any prediction time point, t_{LM} between 0 and 3 years after starting adjuvant ET. For this specific model, the prediction window was set to 5 years after the established t_{LM}. Baseline patient- and tumor-specific factors included in the model comprised age at diagnosis (continuous, linear, and quadratic terms), Bloom & Richardson (BR) histological grade (I, II, III), tumor stage $(1, 2, 3/4)$, nodal stage (No, N1, N2/N3), ER and PR status (positive, negative), HER2 status (positive, negative, missing), most extensive surgery (mastectomy, breast-conserving surgery), and radiotherapy (yes, no), chemotherapy (yes, no). ER and PR status were considered positive if at least 10% of tumor cells stained positively following immunohistochemical staining, as defined by the Dutch BC treatment guidelines[63]. The model also included three dynamic variables whose values may change during ET, namely current ET status (on versus off ET), LRR (yes,no), and DR (yes, no). To assess whether a patient had stopped treatment, we used the last treatment date, as reported on the case-report forms. If no last treatment date was available, the patient was assumed to be on-treatment. According to the TEAM trial protocol, patients with LRR or DR discontinued or switched ET. In order to test for time-varying covariate effects, interactions between covariates and t_{LM} (both linear and quadratic) were included in the model. A backward selection procedure was then carried out in two steps. In the first step, all quadratic t_{LM} interactions with the covariates were tested. Nonsignificant quadratic interactions were removed, and those covariates which did not have significant interactions in the first step were then tested in the second step for linear t_{LM} interactions. Again, only significant interactions were retained. Wald tests, based on robust standard errors, were

used and a p-value of 0.05 was considered statistically significant (Appendix 1.5). Main effects of the covariates and of t_{LM} and t_{LM}^2 were included, irrespective of statistical significance. The model was then validated by internal calibration using the heuristic shrinkage factor by van Houwelingen et al.[93]. The model's ability to correctly discriminate between patients was evaluated using the dynamic cross-validated c-index. A c-index of 1 resembles a model that can perfectly discriminate between patients, while with a c-index of 0.5, the prediction is as good as chance[95].

1.2.2 Nomogram

The nomogram is a user-friendly tool for calculating survival probabilities based on a prediction model, and graphically computes 5-year DOS based on an individual patient's unique characteristics. For each prognostic factor, a number of risk points are assigned to each corresponding covariate, which can be read off the nomogram. The sum of the risk points represents a total risk point score, from which the corresponding 5-year DOS probability can be assessed at any t_{LM} (between 0 and 3 years) after the start of ET. A web-based dynamic prediction tool based on the nomogram has been created to facilitate the calculation of 5-year dynamic overall survival rates and aid in the decision-making process in clinical practice.

1.3 Results

In total, 2602 TEAM trial patients with a median age of 64.8 years (range $38-92$ years), were included in the analyses (Figure 1.1). Baseline characteristics of included patients are depicted in the second column in Table 1.1. The majority of patients included in this trial had adjuvant radiotherapy (66%) and did not receive adjuvant chemotherapy (68%). Figure 1.2 provides an overview of the total number of patients in the landmark datasets at successive t_{LM}s in relation to treatment compliance and disease recurrence status. Table 1.2 depicts the regression coefficients and hazard ratios (HR) with 95% confidence intervals (95% CI) of the covariates included in the model. Covariates with

Table 1.2: The dynamic prediction model with time-constant and time-varying covariates. CI, confidence interval; t_p , prediction time point, time elapsed (years) since the start of treatment.

Covariates with time-constant effects	Coefficient	Hazard ratio	(95% CI)	p-value
Age at diagnosis (ref: 65 years, per 10 years)				< 0.001
Age	0.365	1.440	$(1.254 - 1.653)$	
Age^2	0.154	1.166	$(1.067 - 1.275)$	
Tumor size [ref: T1 (< 2 cm)]				< 0.001
T2 (25 cm)	0.256	1.291	$(1.052 - 1.5850)$	
T3/T4 (5 cm)	0.306	1.357	$(0.956 - 1.928)$	
Histological grade (BR) (ref: BR I)				0.001
BR II	-0.018	0.982	$(0.729 - 1.3230)$	
BR III	0.346	1.413	$(1.038 - 1.923)$	
Estrogen receptor status (ref: positive)				0.073
Negative	0.566	1.761	$(0.948 - 3.271)$	
Progesterone receptor status (ref: positive)				< 0.001
Negative	0.456	1.577	$(1.301 - 1.913)$	
Most extensive surgery (ref: mastectomy)				0.683
Breast-conserving surgery	0.055	1.057	$(0.811 - 1.377)$	
Radiotherapy (ref: yes)				0.157
No	0.195	1.216	$(0.928 - 1.592)$	
Chemotherapy (ref: yes)				0.384
No	0.127	1.136	$(0.853 - 1.512)$	
Treatment status (ref: on-treatment)				0.224
Off-treatment	0.234	1.263	$(0.867 - 1.841)$	
Distant recurrence (ref: no)				< 0.001
Yes	2.709	15.018	$(9.934 - 22.705)$	
Covariates with time-varying effects				
Prediction time (ref: start of treatment)				0.057
t_p	0.017	1.017	$(0.920 - 1.125)$	
t_p^2	-0.034	0.967	$(0.945 - 0.989)$	
Nodal stage (ref: No)				
Constant				< 0.001
N1	0.303	1.354	$(1.021 - 1.795)$	
N2/N3	1.287	3.621	$(2.596 - 5.052)$	
Time-varying effect				0.026
N1 (t_p)	-0.047	0.954	$(0.869 - 1.048)$	
N2/N3 (t_p)	-0.204	0.816	$(0.722 - 0.922)$	
HER2 status (ref: HER2 negative)				
Constant				0.214
Positive	0.211	1.235	$(0.885 - 1.724)$	
Time-varying effect				0.015
Positive (t_p)	-0.162	0.851	$(0.747 - 0.969)$	
Locoregional recurrence (ref: no LRR)				
Constant				< 0.001
LRR	2.131	8.427	$(2.885 - 24.617)$	
Time-varying effect				0.013
LRR (t_p)	-0.540	0.583	$(0.380 - 0.893)$	

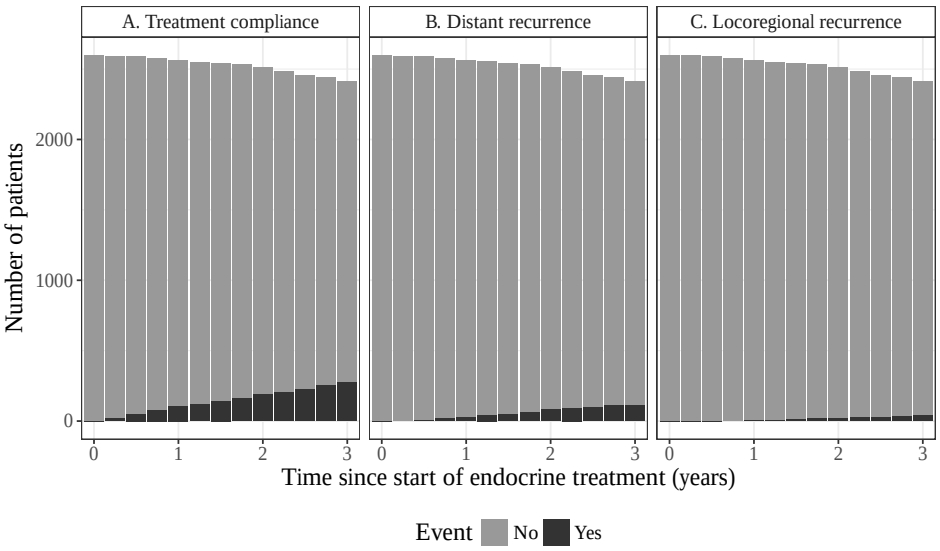

Figure 1.2: Number of patients at risk in relation to follow-up time since the start of endocrine treatment. Number of patients in the landmark datasets (i.e. at risk) over time (t_{LM}) since the start of adjuvant endocrine treatment in relation to (A) treatment compliance status (on-treatment/off-treatment)(B) distant recurrence status (yes, no) and (C) locoregional recurrence status (yes, no).

time-constant effects and covariates with time-varying effects on 5-year DOS are shown. Age at diagnosis demonstrated a time-constant effect, with 5-year DOS being a quadratic function of age (Figure 1.3).

Interestingly, high-risk nodal stage (N2/N3), compared with N0, demonstrated a significant time-varying effect on 5-year DOS with each successive t_{LM}, while nodal stage N1 did not (Figure 1.4.B). To illustrate, the HR of a patient with nodal stage N2/N3 immediately after primary treatment compared with a patient with nodal stage No (reference) is 3.621, calculated by the following formula (Table 1.2):

$$\text{HR} = (\text{constant} \cdot \text{time-varying effect})^{t_p} = 3.621 \cdot 0.816^0 \ ,$$

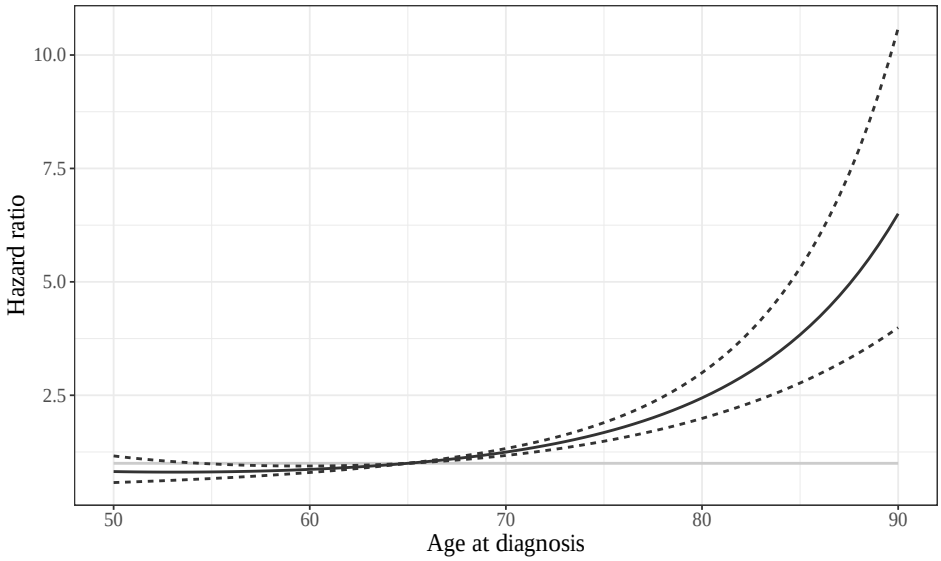

Figure 1.3: Hazard ratio for age at diagnosis depicted with a 95% confidence interval. The hazard ratio increases with increasing age.

but decreases to 2.401 (HR $= 3.621 \cdot 0.816^2$) at 2 years after the start of ET. HER2-positive status also demonstrated a significant time-varying effect on 5-year DOS (Table 1.2, Figure 1.4.A). Next, covariates whose status have the potential to change over time (i.e. treatment compliance status and disease recurrence) were investigated for their influence on 5-year mortality risk. Patients who went off-treatment during the FU period had a higher residual mortality risk compared with patients who remained compliant, although this was not statistically significant. The effect of treatment discontinuation was constant over time (Table 1.2). Simultaneously, LRR had a time-varying influence on 5-year DOS, revealing a subsiding mortality risk with each successive t_{LM} (Figure 1.4.C). Compared with no LRR, having a LRR at $1, 2$, and 3 years after the start of ET increased 5-year mortality risk with HR $= 4.913(2.444 - 9.877)$, HR $= 2.864(1.851 - 4.431)$, and HR $= 1.670(1.005 - 2.773)$, respectively (Table 1.2). In contrast, developing distant metastases (versus no distant metastases) was associated with an increased 5-year mortality risk, with a constant effect over time [HR $= 15.018(9.934 - 22.705)$].

Figure 1.5 illustrates differences in the 5-year DOS in the event of a LRR in a patient who presents with the most commonly occurring baseline characteristics (average patient) found in this cohort, as well as in a high-risk patient. In the absence of a LRR, 5-year mortality probabilities are 3% and 10%, respectively, at all t_ps. However, in case of a LRR, 5-year mortality probabilities in both the average patient and the high-risk patient are initially high, and decrease with time.

1.3.1 Internal model validation

The heuristic shrinkage factor was 0.995, indicating good calibration of the model. Furthermore, the model's discriminatory accuracy had a dynamic cross-validated c-index of $0.70, 0.72, 0.76$, and 0.79 at $0, 1, 2$, and 3 years respectively.

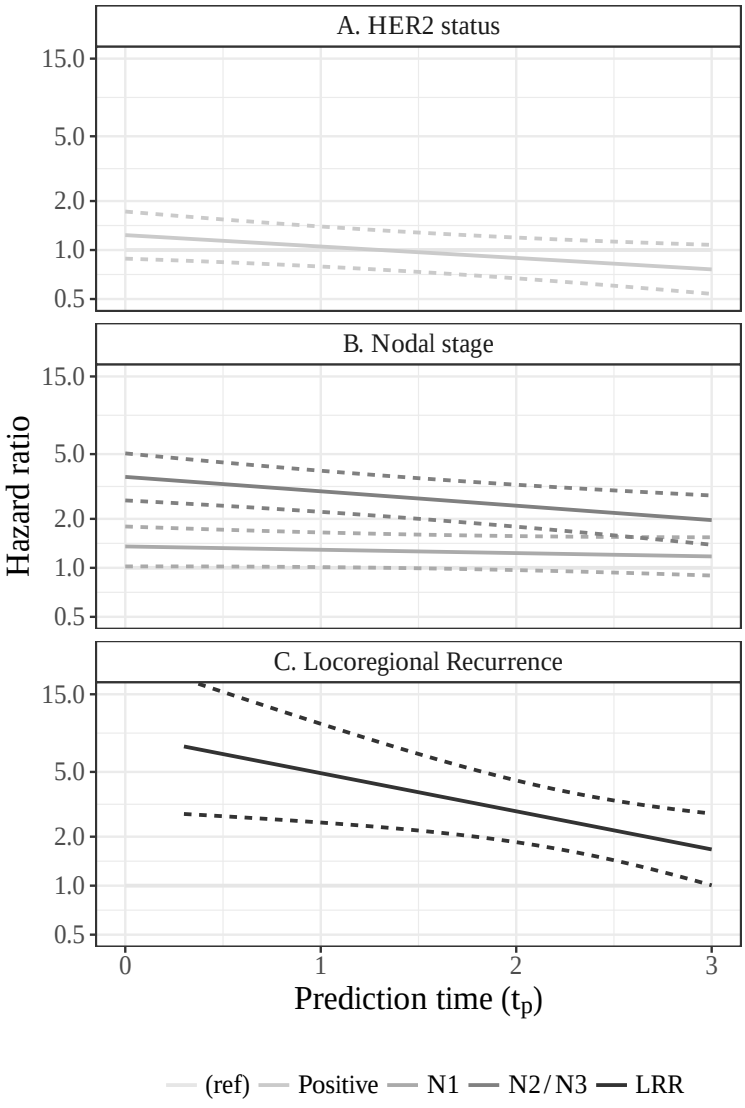

Figure 1.4: Time-varying hazard ratios for nodal stage, HER2 status, and locoregional recurrence status. t_p , prediction timepoint; LRR, locoregional recurrence. Hazard ratios for nodal stage, HER2 status, and locoregional recurrence status as time since the start of endocrine treatment (t_p) increases (depicted as a hazard ratio with 95% confidence interval).

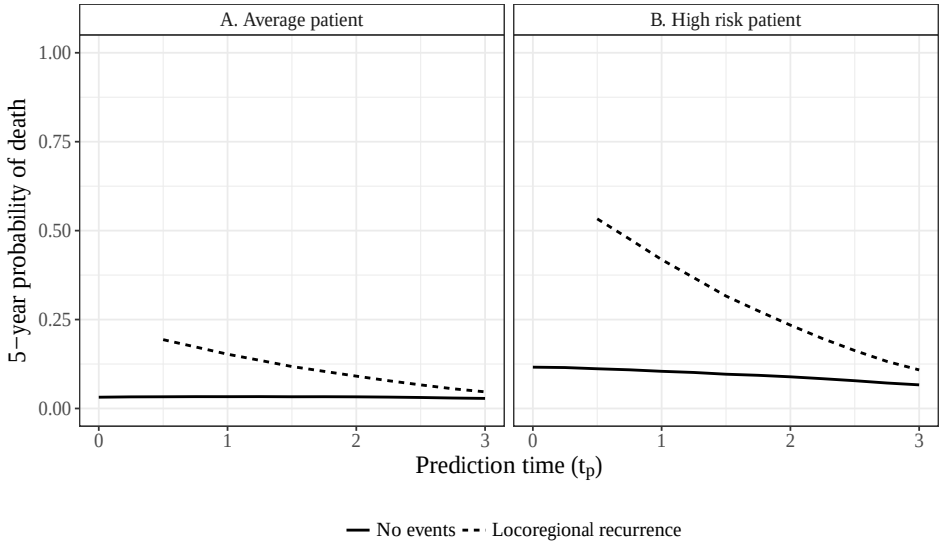

Figure 1.5: Change in 5-year dynamic probabilities of death based on the occurrence of a locoregional recurrence in two example patients. t_p , prediction time point; LRR, locoregional recurrence; ER, estrogen receptor; PR, progesterone receptor. This figure illustrates how 5-year dynamic probabilities of death changes if a patient who is on-treatment throughout the entire follow-up period develops a LRR during follow-up. Two example patients are depicted in (A) and (B). (A) Average patient with the following characteristics: age at diagnosis $= 65$ years, tumor stage T2, nodal stage N1, histological grade II (Bloom and Richardson), HER2 negative, ER and PR positive, treated with breast-conserving surgery, adjuvant radiotherapy and adjuvant chemotherapy. (B) High-risk patient with the following characteristics: age at diagnosis $= 65$ years, tumor stage T3, nodal stage N2, histological grade III (Bloom and Richardson), HER2 negative, ER and PR positive, treated with mastectomy, adjuvant radiotherapy and adjuvant chemotherapy.

1.3.2 Using the nomogram

The nomogram (Figure 1.6) provides estimates for 5-year DOS probabilities at different t_ps from the start of ET and onwards, provided that adequate surgery has been carried out. The probabilities can be calculated by adding the risk points for each covariate corresponding to the patient's individual characteristics. For each characteristic, the number of associated risk points can be determined by drawing a vertical line straight up from the covariate's corresponding value to the axis with risk points $(0 - 80)$. While the majority of covariates are considered 'static' and defined at the start of ET, some covariates are 'dynamic', and can alter during the course of FU, such as treatment compliance status and the occurrence of LRR or distant metastases during FU. The covariates marked with '(t_p)' (prediction time point) include nodal stage (N2/3), HER2 status (positive), and LRR (yes), and have time-varying effects on 5-year DOS. This means that the effect of having characteristics that pertain to one these specific covariates varies as the time since starting treatment progresses and that the time since the start of ET needs to be taken into account when making a 5-year DOS prediction. The sum of the risk points is equal to the total risk point score, which is depicted on the axis of the nomogram entitled 'Total Points'. From here, a vertical line can be drawn toward the axis labeled '5-year survival probability', which is the corresponding 5-year DOS at that specific t_p. To illustrate, we consider a 69-year-old postmenopausal woman (14 points) who has been using ET for two years ($t_p = 2$; 191 points). She had a grade III tumor (13 points) with a diameter of 1.5 cm (0 points), ER-positive (0 points), PR-positive (0 points) and HER2-negative (10 points), and 5 tumor-positive lymph nodes(at $t_p = 2$; 32 points). The patient has undergone breast-conserving surgery (2 points) with adjuvant radiotherapy (0 points) and adjuvant chemotherapy (0 points). She is still on-treatment (0 points) and disease-free (0 points) (no locoregional or DR). To calculate her 5-year DOS probability, we take her total risk point score (90 points) and draw a vertical line down to the '5-year survival probability' axis. For this patient, the 5-year DOS is 75%. If our patient had developed a LRR in the 2-year period since ET, one must add an additional 38 points (total $= 128$ points) to her total risk prediction score,

resulting in a 5-year DOS of 42%.

1.4 Discussion

To our knowledge, this is the first dynamic prediction model in clinical oncology, designed to optimize the prediction of the 5-year DOS at specific time points after the start of adjuvant ET in postmenopausal, endocrine-sensitive early BC patients. The key advantage of this model is that it takes into account dynamic factors that can influence a patient's prognosis after some time has passed since starting ET, including treatment compliance and the occurrence of LRR or distant metastases. Moreover, covariates with time-varying effects are also accounted for in the model, including high-risk nodal stage (N2/3) and HER2-positive status.

Current nomograms are suboptimal for cancer patients, because their reference point is commonly the time of diagnosis or the start of adjuvant ET. Aiming at further personalized BC treatment, continuous re-evaluation of the residual risk of BC recurrence and mortality during FU is crucial. Patients may develop disease recurrences or discontinue ET before the predesignated end-date, which may alter a patient's prognosis from that time point onward. Additionally, the effect of a covariate on 5-year survival probabilities may not be constant over time. These changes are more prominent than current statistical models account for, which could lead to the risk of developing less effective treatment guidelines. Therefore, survival prediction models need to be adapted for long-term outcome prediction in individual patients. Specifically, dynamic prediction models can be used to determine whether a patient will benefit from further adjuvant systemic therapy or, conversely, whether ET can be discontinued at a certain time point during FU.

The current nomogram can be applied to postmenopausal, HR+ BC patients undergoing adjuvant ET and have had an axillary lymph node dissection in case of macrometastases. For patients who have had breast-conserving surgery, the model assumes that the breast was irradiated. The current nomogram also assumes that disease relapse

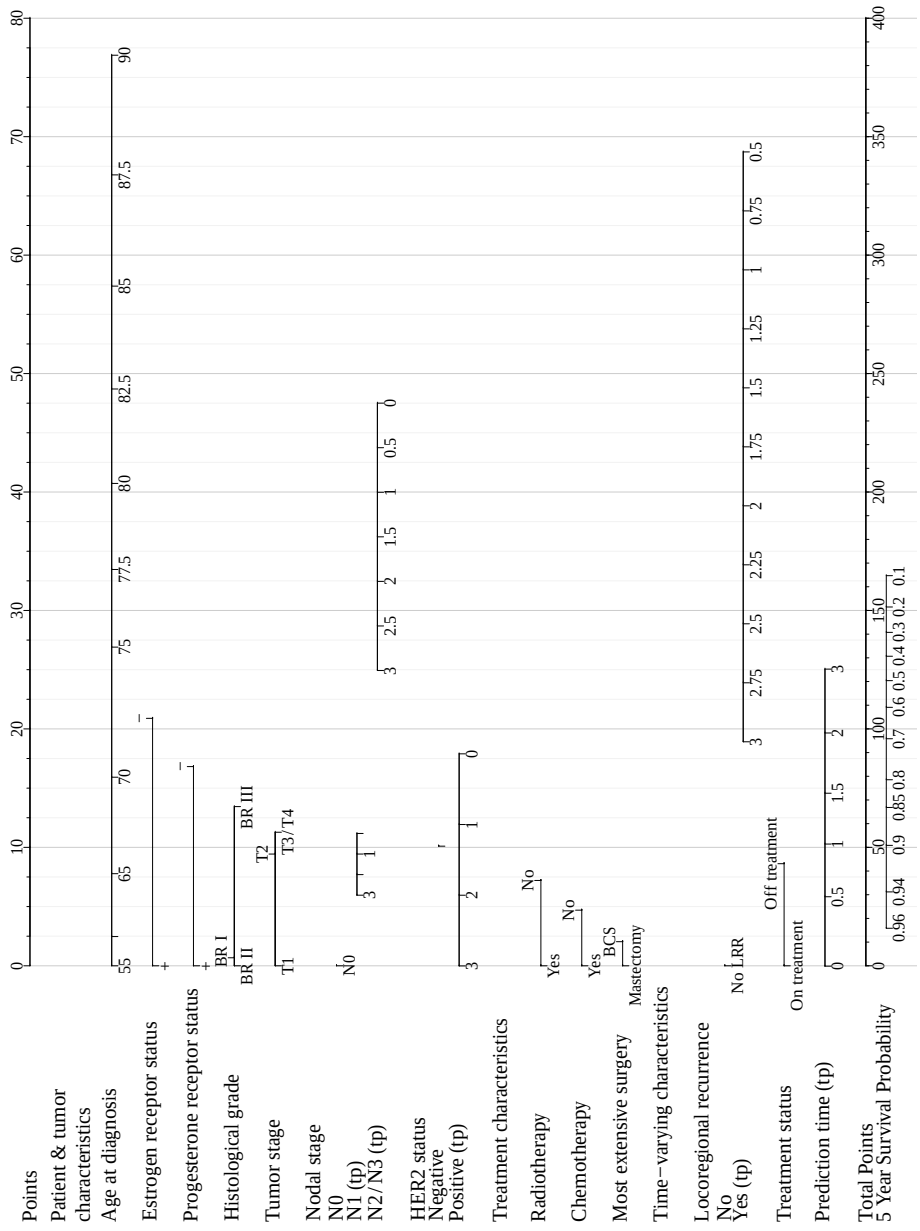

Figure 1.6: Nomogram for dynamic prediction of the 5-year survival probability. BR, Bloom and Richardson; BCS, breast-conserving surgery; t_p , prediction time point.

implies discontinuation of ET from that moment onward. In case of disease recurrence, data on subsequent treatment were not available for all patients; hence, our ability to draw conclusions for this subgroup is limited. LRR is considered a 'dynamic' covariate, as patients can develop a LRR at any moment during FU. LRR also had 'time-varying' properties, as the event of a LRR revealed a changing impact on 5-year DOS at different time points after starting ET. Our findings parallel those of several other studies, which have shown that early LRRs are predictive of a worse prognosis than late LRRs [21,32,97,64,91]. It can therefore be of major clinical importance to include this factor in dynamic survival prediction. Moreover, this model could potentially help evaluate the need for additional adjuvant chemotherapy in case of LRR. Data on the benefit of additional chemotherapy are still relatively lacking, although the nomogram could be useful in this setting. The current model also revealed a time-varying relationship between high nodal stage (N2/3) and 5-year DOS probability. A similar time-varying effect was shown with regard to 5-year DOS in HER2-positive patients, although no patients received anti-HER2 treatment. To our knowledge, no prior reports have investigated the time-varying effects of these two prognostic factors, hence warranting further investigation. Our dynamic prediction model also accounts for the effect of early treatment discontinuation for reasons other than BC relapse. Although the effect of treatment discontinuation did not reach statistical significance, possibly due to the low number of patients who discontinued treatment within three years (Figure 1.2.A), we retained this data in our model, as an earlier review revealed the importance of treatment compliance on survival outcomes[1]. The number and site(s) of DR are known to be prognostic for subsequent survival[16,18,85]. The dynamic prediction model incorporates the occurrence of distant metastases, but does not include this in the nomogram due to insufficient data concerning first site of DR and subsequent treatment. For this reason, it is not advised to use the dynamic prediction model for patients with distant metastases as first site of disease recurrence. Internal validation demonstrated that the model had a good ability to discriminate between patients. To elucidate, internal validation of Adjuvant! showed a c-index of 0.71 for discriminatory accuracy (the ability

for the model to distinguish patients who will versus those who will not die of BC) and a predictive accuracy of 0.73 at diagnosis, which is similar to that of our prediction model[61]. The predictive accuracy of Adjuvant! 'after diagnosis' has not been studied; in contrast, our dynamic prediction model showed a cross-validated c-index that improved from 0.70 to 0.79 3 years after the start of adjuvant ET. Due to the unavailability of regular FU data for the entire TEAM trial population, our dynamic prediction model includes Dutch and Belgian TEAM trial patients only. As shown in Table 1.1, characteristics of the Dutch trial population differed slightly in comparison to the rest of the TEAM trial population. These differences depict that patients in current cohort have a slightly higher disease stage and subsequent variations in treatment. The dynamic prediction model is a multivariate model that corrects for each of these variables. Therefore, inclusion of the entire TEAM trial population in the model could alter individual predictions. Importantly, however, this is not expected to affect the 'correctness' of the model, which would only be affected in case of lack of model fit. Of note, one must also consider that any trial population is not representative of the general BC population as a whole. For this reason, further external validation of the prediction model is required in greater (non-trial) cohorts to allow for full applicability in the clinical setting. An independent population with adequate FU data for performing an external validation of the dynamic prediction model was not available at the time of conducting this study. In summary, the importance of using dynamic prediction models for clinical guidance, not only at the start of treatment, but also during FU, permits continuous revision of a patient's residual mortality risk and can help motivate a patient to continue treatment, improve compliance, and ultimately improve survival. This proof-of-principle study demonstrates a novel technique for performing dynamic prediction of BC survival probabilities over time, enabling a more individualized prediction of the 5-year DOS in individual patients at various time points during adjuvant ET. The most important advantage of this model is that it takes into account factors that can influence an individual patient's prognosis after some time has passed since starting adjuvant ET. Notwithstanding the feasibility of our dynamic prediction model, further external val-

idation with longer FU is necessary to enable implementation in clinical practice.

1.5 Appendix

This appendix provides a more detailed description of the statistical method applied in this paper. The method builds upon the concept of landmarking, which was introduced by Anderson et al.[11] as a way to deal with time-dependent covariates in survival analysis in order to avoid immortal time bias. Later van Houwelingen[92], van Houwelingen & Putter[94,95] proposed to use landmarking for dynamic prediction of the survival probability with time-dependent covariates.

1.5.1 Dynamic prediction using landmarking

The idea behind landmarking is to select a point in time s known as a landmark. By only selecting subjects at risk at s a landmark data set is constructed, which can be seen as imposing artificial left-truncation at time s. In addition, we can also select a prediction window ω and impose artificial right-censoring at time $s + \omega$ (Figure 1.7). For a time-dependent covariate $Z(t)$, such as distant recurrence, the current value $Z(s)$ at s is used. Here distant recurrence was included as a indicator function for whether or not distant metastases had been detected. The resulting landmark data can be analysed using standard methods such as Kaplan-Meier or Cox regression using $Z(s)$ as a time-constant covariate.

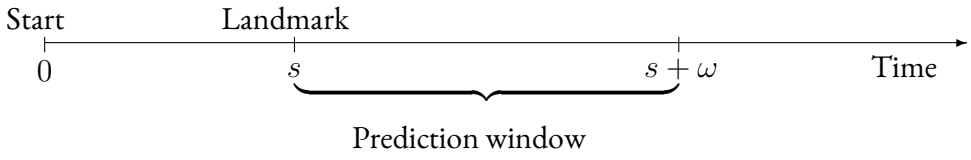

Figure 1.7: Time line illustrating the idea behind dynamic prediction using landmarking.

Using a Cox proportional hazards model implies that the hazard, i.e. the instantaneous risk of dying, is given by

$$\lambda(t|\boldsymbol{Z}(s), s) \quad = \quad \lambda_0(t|s) \exp\left(\beta(s)^\top \boldsymbol{Z}(s)\right) \ ,$$

where $\lambda_0(t|s)$ is the baseline hazard given survival up until time s. The proportionality factor

$$\exp\left(\beta(s)^\top \boldsymbol{Z}(s)\right)$$

depends on the covariates fixed at their current value at the landmark time $\boldsymbol{Z}(s)$ and their effect $\beta(s)$. The model can be fitted with standard software to obtain estimates of $\lambda_0(t|s)$ and $\beta(s)$. We can then use the estimates to predict the conditional probability of surviving ω more years after time s, for a new subject with covariate values $\boldsymbol{Z}^*(s)$. To this end, we can use the relation between the survival- and hazard function

$$S(t|\boldsymbol{Z}^*(s), s) \quad = \quad \exp\left(-\int_s^{s+\omega} \lambda(u|\boldsymbol{Z}^*(s), s)du\right) \ . \tag{1.1}$$

The prediction are obtained by plugging in the estimated hazard function

$$\hat{\lambda}(t|\boldsymbol{Z}^*(s), s) \quad = \quad \hat{\lambda}_0(t|s) \exp\left(\hat{\beta}(s)^\top \boldsymbol{Z}^*(s)\right) \ .$$

This method can be applied for one or for several landmark times. However, in order to be able to predict survival for any time between start and up to some natural limit (we cannot predict beyond end of follow-up), we can use a landmark super model instead.

1.5.2 Landmark super models

The general idea of the landmark super model is to select not just one, but several landmark time points $\{s_1, \ldots, s_K\}$. For each of these landmark times a landmark data set is created, as described above, by imposing left-truncation and right-censoring. The K data sets are then stacked into a super landmark data set. This is similar to longitudinal

survival data, where a subject can contribute with several observations.

For the landmark super model for the TEAM data we used a Cox proportional hazards model

$$\lambda(t|\boldsymbol{Z}(s), s) \quad = \quad \lambda_0(t) \exp\left(\theta(s)\beta(s)^\top \boldsymbol{Z}(s)\right) \ ,$$

where $\theta(s)$ is a function of s, which describes how the baseline changes over the landmark time. Similarly, $\beta(s)$ is a vector of functions that describes changes in the covariates' effect. We chose to use smooth parametric functions such as

$$\theta(s) = \theta_0 s + \theta_1 s^2 \ ,$$

where θ_0 and θ_1 are parameters to be estimated. With this choice of function the baseline hazard is allowed to vary non-linearly across landmark time. The interpretation of the parameter functions $\beta(s)$ in the super model is comparable to the interpretation in the traditional Cox model; the effect of the covariates works multiplicatively on the baseline hazard.

In order to find a suitable model for the TEAM data we went through a model building process. The first step of the process was to select covariates that were known to be predictors of overall survival. In the second step we investigated whether any of these covariates had (landmark) time-varying effects by allowing them to be non-linear. A model including all covariates was therefore fitted, where all covariate effects were of the form

$$\beta(s) = \beta_0 + \beta_1 s + \beta_2 s^2 \ ,$$

where β_0, β_1 and β_2 are parameters. A model selection procedure was then carried out in two steps: In each one a backward selection was used to decide in which order to remove terms. In the first step the quadratic time interactions were tested, i.e. for each covariate the hypothesis $\beta_2 = 0$ was tested. Nonsignificant quadratic interactions, at the 5% level, were removed. In the second step all linear interactions, $\beta_1 = 0$, were

tested, but only if the quadratic time interaction was removed in the first step. The resulting model is then what is presented in the paper. It was decided to also retain covariates for which the effect was not significant, because they are known to be important predictors.

Once the model had been finalized, we obtained predictions of the survival probability for any time s, between s_1 and s_K, using the same formula as before (1.1). The stacked landmark data set contain repeated observations on the same subjects and to account for this one can use the robust sandwich estimator[59] to estimate the variance. In summary, the difference between the landmark super model and having separate models for each selected landmark is that we can predict at any time between s_1 and s_K, and not just at those exact times. This is due to the fact that the super model assumes a structure for how the hazard and the covariate effects change with s.