Estimating the risk of locoregional recurrence and second primary breast cancer: the role of distant metastasis as a competing risk

Master thesis for
Health science; optimalization of health care processes
University of Twente

Date: 25th of September 2023
Author: Anne Fleur Boone

Boone, A.F. (Fleur, Student M-HS)
S2885743 | a.f.boone@student.utwente.nl

UNIVERSITY OF TWENTE. iKNL integraal kankercentrum Nederland EVIDENCIO MEDICAL PREDICTION MODELS

Supervision by

Prof. dr. Sabine Siesling | s.siesling@iknl.nl
Prof. Dr. Job van der Palen | J.vanderPalen@mst.nl
Dr. Tom Hueting | tom@evidencio.nl

Abstract

Purpose Follow-up after breast cancer treatment focuses on early detection of locoregional recurrences (LRR) or second primary breast cancer (SP) to improve patient outcomes. Estimating the patients individual 5-year recurrence risk can help healthcare providers and patients develop personalised risk-based follow-up pathways. As the diagnosis of distant metastasis (DM) likely affects the risk of detecting LRR or SP, it should be considered as a competing risk when developing models to predict LRR or SP. The objective of this study is to assess the role of DM as a competing risk when predicting the 5-year recurrence risk for LRR and SP.

Methods Data from 13,494 breast cancer patients were used. Two models were created for both outcomes LRR and SP, one where DM is considered a competing risk and one where DM is not considered a competing risk. The statistical approaches Cox regression analysis (COX) and random survival forest (RSF) were used to develop the prediction models. The predictive performance was assessed on model calibration and discrimination, absolute, mean, and relative predicted risks to assess the impact of including DM as a competing risk.

Results The RSF approach showed a 5-year AUC of 0.76 for predicting LRR and DM was considered a competing risk and also when DM was not considered a competing risk. The 5-year AUC for predicting SP is 0.71 when DM was considered a competing risk and 0.70 when DM was not considered a competing risk. The mean risk difference for the predicted 5-year risk with the RSF approach was 0.07% and 0.04% for models predicting LRR and SP, respectively. For the COX approach, this was 0.11% and 0.04% for models predicting LRR and SP, respectively. In both groups, the mean predicted risk for the unknown DM cohort was lower than the mean predicted risk for the known DM cohort. The models calibration and discrimination appear largely uniform in both situations.

Conclusions The mean predicted risks for all patients and model performances are almost similar in both models, DM has no substantial role in predicting the risk for LRR and SP.

Introduction

Breast cancer is the most common cancer among women in the Netherlands, with approximately 16,000 new cases diagnosed each year, accounting for over 25% of all cancer cases in women.[1,2] Although the incidence of breast cancer has increased over the past few decades, the mortality rate has decreased due to advancements in early detection and treatment options, this leads to increased women in the aftercare for breast cancer.[3,4] After successful treatment for breast cancer, a follow-up plan is designed for aftercare and surveillance. Aftercare aims to address the psychological, social, and physical effects of breast cancer (treatment) by providing guidance, support, and treatment. Surveillance aims to detect any possible recurrence of cancer in the treated breast or in the axilla (locoregional recurrence (LRR)), or the development of a new primary cancer in the opposite breast (second primary contralateral breast cancer (SP)).[5] The risk for developing LRR or SP can vary between patients. Factors that influence the risk of recurrence include; tumour size, age, multifocality, histological grade, hormone receptor status and treatment of the primary tumour.[6] Risk prediction models can contribute to create personalised surveillance strategies.[7]

A risk prediction model recently developed, updated and available for use to support personalized surveillance is the INFLUENCE model.[8] The current INFLUENCE 2.0 (available at https://www.evidencio.com/models/show/2238) model predicts time-dependent individual risks of LRR, SP, and distant metastases (DM) over 5-years after surgical treatment. Therefore, INFLUENCE can support in developing a personalised follow-up plan. The current INFLUENCE 2.0 model is based on a population of patients diagnosed in 2007, 2008 and the first quarter of 2012. In this population, patients who received neoadjuvant systemic therapy were excluded because the number of patients who received this therapy was too low. In the recent years breast cancer incidence, mortality and treatment changed, e.g. neoadjuvant systemic therapy is increasingly applied in the treatment of breast cancer.[9] And the early detection of breast cancer potentially reduces mortality and treatment related burden which could have decreased the risk of LRR or SP. [10] For that reason, an update of the INFLUENCE model is in progress.

Data used to develop and validate the INFLUENCE model was obtained through the Netherlands Cancer Registry (NCR). The NCR is a nationwide population-based cancer registry including all hospitals in the Netherlands (n=89).[11] After notification through the nationwide pathology archive (PALGA), information on each patient is collected directly from patient files by specially trained registration clerks. These data include patient demographics, tumour-, and treatment characteristics. In addition, vital status and date of death were regularly retrieved through linkage with the national municipality registry. Data on follow-up were previously collected using a cohort approach, meaning that data managers gathered information in retrospect from the patient files of patients diagnosed and treated 5 years ago. This approach ensured that data on follow-up is complete, but is very time-consuming to perform. Since the number of patients requiring follow-up care has been rising, other approaches were implemented/required to avoid capacity constraints. To collect data in an efficient way, the (NCR) received a notification when any new event occurred by the Dutch pathology labs through PALGA. And instead of manually checking the patient file, an algorithm based on the incidence dates and diagnostic codes, dictated by the pathologist, is made to identify patients with LRR and SP. The limitation of this way of more automated data collection is that data on DM is not completely registered. DM are usually not punctured, meaning that no pathological confirmation is available and the notification algorithm cannot identify the DM event. As a result, the clinical diagnosis will only be reported in the NCR when the patient files are manually registered by the data clerks. In this case, DM can statistically be seen as a competing risk. A competing risk is an event whose occurrence excludes the occurrence of the primary event of interest or has an influence on the risk of the primary event occurring.[12] Being incomplete on the possible occurrence of a DM will prolong the disease-free follow-up time for many patients in the data set because patients are assumed to be tumour-free in case of no LRR or SP and will not be censored, while some have had DM. The assumption is that the risk of LRR or SP breast cancer is different when DM cannot be regarded as a competing risk due to missing data, especially in

subgroups at high risk for DM. This is because a patient who ideally would be censored when the DM occurs, will not be censored and the follow-up time will proceed as event free for LRR and SP in the survival analyses. The aim of this study is to assess the role and effect of DM as a competing risk on the prediction of LRR and SP for breast cancer patients.

The research design and method(s)

Study population and variables

The NCR data used for the development of the INFLUENCE 2.0 which contains information about DM, was also used in this study. All women were selected with primary non-metastatic (pT1-3, any pN) invasive adenocarcinoma of the breast, diagnosed in 2007, 2008 or the first quarter of 2012. For this cohort, active follow-up for the first five years following the surgery (free margins) of the primary tumour was conducted in which information on recurrences (i.e. LRR, SP and DM) was collected. Patients were excluded in case of positive resection margins of the primary tumour, if neoadjuvant therapy was conducted, or if surgery took place later than 180 days after diagnosis.

Events of interest were LRR, defined as a reappearance of the tumour in the ipsilateral breast, chest wall or regional lymph nodes and SP, defined as a secondary primary tumour of the contralateral breast. DM was defined as pathologically or radiologically confirmed reappearance of tumour tissue at any location in the body other than LRR or SP.[13]

The cohort was used two times, once with the complete data where DM is considered a competing risk and a patient was censored if any event occurs (LRR, DM, SP), if it is not the event of interest (LRR or SP) referred to as the DM known cohort. And once where DM is not considered to be a competing risk. If the patient had DM this was ignored, and the follow-up time will continue as event free until the end of the follow-up or occurrence or any other event (LRR or SP) if it is not the event of interest (LRR or SP), referred to as the DM unknown cohort. The approach used is called a cause specific hazard model. The aim of this approach is to estimate the absolute risk of the event of interest.[14]

The same predictor variables as the INFLUENCE 2.0 were used for the model development; age, pT-stage, pN-stage, multifocality, grading, hormone receptor status (estrogen receptor (ER)- and progesterone receptor (PR)-status), antihormonal therapy, human epidermal growth factor receptor 2 (HER2-status), type of surgery, adjuvant chemotherapy, adjuvant radiation therapy and antibody therapy. Missing data was assumed to be missing completely at random, therefore a complete case analysis was done. These variables were selected based on previous studies and clinical expertise. Out of the 17,014 patients diagnosed with invasive breast adenocarcinoma in 2007, 2008 or during the first quarter of 2012 identified from the NCR for the development of the INFLUENCE 2.0, 13,494 individuals met all eligibility criteria and had complete data records.

Model development

The individual time-dependent risks for LRR and SP were estimated. Two statistical approaches were used to develop the models on the data: a Cox proportional hazards model and a random survival forest (RSF) model:

- The Cox proportional hazards method[15] is classified as a semi-parametric model because it does not make any assumptions about the baseline survival distribution. However, it assumes that the predictors have a consistent effect on the underlying hazard function.
- The Random Survival Forest[16] is a variation of the traditional Random Forest algorithm for binary outcomes[17] that can handle right-censored time-to-event data. It builds a forest of survival trees by employing a log-rank splitting rule to determine the best predictor variables. Survival estimates are then generated using a Kaplan-Meier estimator[18] within each terminal node at every time point.

Eight models were created to assess the impact of DM as a competing risk. The two statistical approaches are applied to cohort where DM is considered a competing risk, this gives a COX and an RSF prediction model for the risk prediction for LRR and SP. Secondly, this was done for the cohort where DM is not considered a competing risk.

Model performance

Models were developed and compared on their predictive performance in terms of discrimination and calibration. The model performance was measured using apparent and adjusted values. The apparent performance results reflect the performance of the models in the same data used to train them. The adjusted values reflect the performance measured using 200 bootstrap samples. Bootstrapping is a statistical procedure that resamples a single dataset to create many simulated samples. Every sample reflects the performance of an approach trained in a bootstrap sample applied to the entire dataset.[19] Calibration and discrimination were compared for both models as performance measures.[20] Calibration refers to the agreement between observed and predicted events. In order to provide measurable summaries of model calibration, the following metrics were calculated at 1, 2, 3, 4 and 5 years: the Integrated Calibration Index (ICI) is a weighted difference between observed and predicted probabilities, E50 is the median difference between observed and predicted values and E90 is the difference at the 90th percentile. These metrics quantify the absolute differences between predicted and observed probabilities.[21] To assess the adequacy of calibration, the ICI values were compared with the observed absolute event rates, with an adequate ICI below 0.01. The area under the receiver operating characteristic curve (AUC) was used to measure discrimination. This metric measures the likelihood that a random set of individuals experiencing an event will have a higher predicted risk than a random set of individuals not experiencing the event. An AUC value of 1.0 indicates perfect discrimination, while 0.5 indicates chance-level performance. The AUC was calculated on a yearly basis throughout the five-year prediction period in the DM known cohort and DM unknown cohort to assess changes in the AUC over time.[22]

The calibration and discrimination for every predicted risk in all five years of follow-up were evaluated. For every patient in the data set, the five-year probability was calculated with every model, and the absolute and relative risk differences between the models where DM is considered a competing risk and where DM is not considered a competing risk.

Triple-negative patients have a higher risk of DM[23] and were assessed in a subgroup analysis to evaluate the degree of underestimation of LRR or SP. The primary objective of this study is to assess the extent to which the risk of LRR or SP is underestimated within subgroups. To accomplish this, we will categorize predicted risks into three distinct groups: less than 5%, between 5% and 10%, and greater than 10%. This categorization will allow us to evaluate the disparities in predicted risk between scenarios with competing risk adjustments and those without. The distribution into <5%, 5%-10%, and >10% risk categories pertains to a 5-year risk prediction time span. Clinically significant risk differences, whether absolute or relative, are generally deemed clinically significant when they have the potential to influence treatment decisions, patient counselling, or the allocation of healthcare resources. Clinicians often rely on their expertise and available clinical guidelines to assess what constitutes a clinically meaningful difference in risk. There is no specific threshold that classifies patients as high risk for recurrence. Therefore, a wide range of illustrative thresholds have been shown. For each of those thresholds, it was evaluated how many patients would be classified as high risk and have a possible different follow-up based on either of the RSF models. High risk was defined as an individual predicted 5-year risk above the threshold.

Results

All relevant characteristics of the patient cohort are shown in Table 1. Among the patient population (n = 13,494), the majority had either a pT1 (64%) or pT2 (34%) tumour, with no involvement of lymph nodes (65%), and a low tumour grade (70% grade 1 or 2). Furthermore, 85% of patients had a unifocal tumour. Approximately 58.9% of cases underwent breast-conserving surgery. Adjuvant radiation therapy and adjuvant chemotherapy were administered to 67% and 38% of patients, respectively. Of the patients, 85% tested positive for ER and/or PR, and around 40% of them received antihormonal therapy. 12.5% of the patients were Her2-positive, and among them, 60% received antibody treatment. Within a span of five years since surgical treatment, 385 patients (2.8%) experienced an LRR, 411 patients (3.0%) had an SP, and 848 patients (6.3%) experienced a DM as their first event. The majority, 11,839 patients (87.7%), remained free of recurrence.

Table 1. Patient characteristics

Variable	*N (%) Total is 13,494*
Inclusion year	
2007	5508 (41.1%)
2008	5621 (41.7%)
2012	2365 (17.5%)
Age group	
50-59 years	3091 (22.9%)
60-69	3635 (26.9%)
70-79	3531 (26.2%)
≥80	3237 (24.0%)
Grading	
Grade 1	3409 (25.3%)
Grade 2	6047 (44.8%)
Grade 3	4038 (29.9%)
pT	
pT1	8692 (64.4%)
pT2	4514 (33.5%)
pT3	288 (2.1%)
pN	
pN0	8782 (65.1%)
pN1	3493 (25.9%)
pN2	790 (5.9%)
pN3	429 (3.2%)
Multifocality	
No	11,425 (84.7%)
Yes	2069 (15.3%)
Surgery	
Breast-conserving	7942 (58.9%)
Mastectomy	5552 (41.1%)
Chemotherapy	
No	8366 (62%)
Yes	5128 (38%)
Radiotherapy	
No	4403 (32.6%)
Yes	9091 (67.4%)
Hormonal therapy	
HR+ & no hormonal therapy	6560 (48.6%)
HR+ & hormonal therapy	4881 (36.2%)
HR-	2053 (15.2%)
Targeted therapy	
HER2+ & no targeted therapy	678 (5.0%)

HER2+ &targeted therapy	1015 (7.5%)
HER2-	11,801 (87.5%)
First event (5 year follow-up)	
LRR	385 (2.8%)
SP	411 (3.0%)
DM	848 (6.3%)
None	11,839 (87.7%)

The coefficients of the COX approach are compared in both situations. the biggest differences in coefficients for LRR and SP were 0.159 and 0.062, respectively for the variable pN3. See table 2 for all coefficients and their differences.

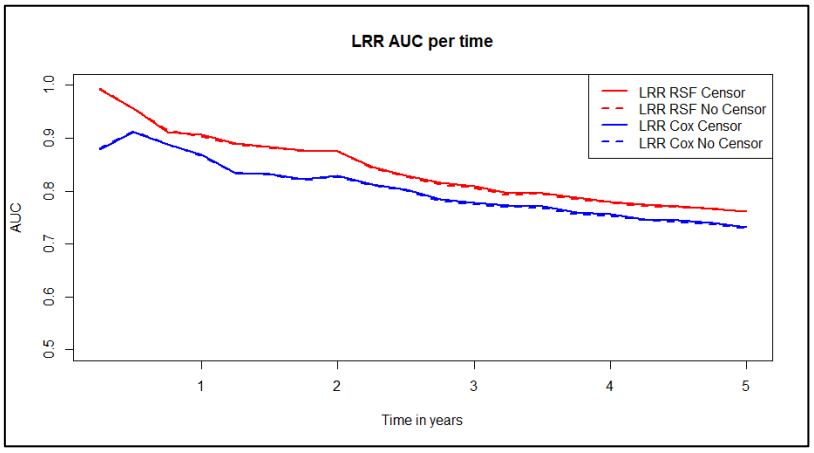
Table 2. Coefficients of the Cox regression model

	LRR (DM known cohort)	LRR (DM unknown cohort)	Difference	SP (DM known cohort)	SP (DM unknown cohort)	Difference
Age 50-59	Reference	Reference		Reference	Reference	
Age 60-69	-0.259	-0.249	**-0.010**	-0.101	-0.096	**-0.005**
Age 70-79	-0.414	-0.403	**-0.011**	-0.127	-0.122	**-0.005**
Age ≥80	-0.314	-0.315	**0.001**	-0.104	-0.102	**-0.002**
Grade I	Reference	Reference		Reference	Reference	
Grade II	0.496	0.488	**0.008**	0.122	0.119	**0.003**
Grade III	1.033	1.017	**0.016**	-0.202	-0.212	**0.01**
pT1	Reference	Reference		Reference	Reference	
pT2	0.711	0.686	**0.025**	0.085	0.074	**0.011**
pT3	1.231	1.214	**0.017**	0.594	0.579	**0.015**
pN0	Reference	Reference		Reference	Reference	
pN1	0.584	0.571	**0.013**	-0.186	-0.191	**0.005**
pN2	1.274	1.210	**0.064**	0.095	0.071	**0.024**
pN3	1.858	1.699	**0.159**	-0.697	-0.759	**0.062**
No multifocality	Reference	Reference		Reference	Reference	
Multifocality	0.099	0.091	**0.008**	0.176	0.174	**0.002**
Conserving surgery	Reference	Reference		Reference	Reference	
Mastectomy	-0.252	-0.224	**-0.028**	0.581	0.584	**-0.003**
No chemotherapy	Reference	Reference		Reference	Reference	
Chemotherapy	-0.852	-0.844	**-0.008**	-0.230	-0.222	**-0.008**
No radiotherapy	Reference	Reference		Reference	Reference	
Radiotherapy	-0.826	-0.792	**-0.034**	0.240	0.242	**-0.002**
HR - & treatment -	Reference	Reference		Reference	Reference	
HR + & treatment +	-1.459	-1.442	**-0.017**	-0.379	-0.372	**-0.007**
HR + & treatment -	-0.304	-0.320	**0.016**	0.177	0.180	**-0.003**
HER2 - & treatment -	Reference	Reference		Reference	Reference	
HER2 + & treatment +	-0.962	-0.975	**0.013**	-0.570	-0.576	**0.006**
HER2 - & treatment -	-0.161	-0.182	**0.021**	-0.120	-0.126	**0.006**
Baseline hazard year 1	*0.0068*	*0.0068*	***0.0000***	*0.0048*	*0.0052*	***-0.0004***
Baseline hazard year 2	*0.0202*	*0.0200*	***0.0002***	*0.0195*	*0.0243*	***-0.0048***
Baseline hazard year 3	*0.0359*	*0.0352*	***0.0006***	*0.0315*	*0.0359*	***-0.0044***
Baseline hazard year 4	*0.0498*	*0.0487*	***0.0011***	*0.0433*	*0.0479*	***-0.0045***
Baseline hazard year 5	*0.0603*	*0.0590*	***0.0014***	*0.0522*	*0.0543*	***-0.0020***

For the COX and RSF approach, the mean (table 3) and time dependent AUC (fig 1) are compared. It is visible that in the model where DM is considered a competing risk and in the model where DM is not considered a competing risk the AUC's are almost similar in both models. In table 3 the mean AUC's and the differences between the DM known cohort and the DM unknown cohort can be seen.

Fig. 1 a: Area under the Receiver operating characteristic curve (AUC) per year for the outcome locoregional recurrence (LRR). b: Area under the Receiver operating characteristic curve (AUC) per quarter year for the outcome of secondary primary breast cancer (SP). Cox proportional hazard model, and RSF (Random Survival Forest)

a.



b.

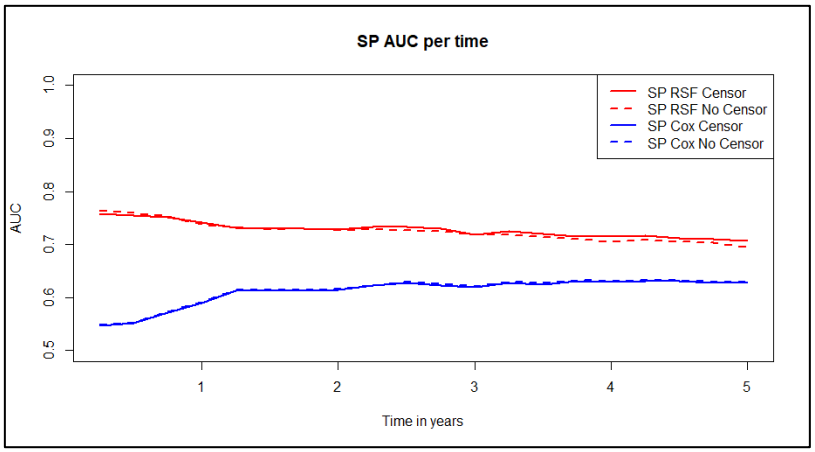


Table 3. Mean AUC

	LRR censor	LRR no censor	Difference	SP censor	SP no censor	Difference
Mean COX	0.731	0.725	-0.006	0.646	0.648	-0.002
Mean RSF	0.818	0.814	-0.004	0.778	0.773	0.005

The Mean AUC and the difference in AUC between the DM known cohort and DM unknown cohort.

In table 4 the calibrations of the models and their differences are shown in the DM known cohort and DM unknown cohort for both modelling approaches. It shows that both modelling approaches show adequate calibration at all tested time points, reflected by an ICI below 0.01. With the exception of the ICI in the SP DM known and unknown cohort for the RSF approach. Both the AUC and calibration show a better performance of the RSF approach in the DM known cohort and DM unknown cohort.

Table 4. Calibration results

	year	COX			RSF		
		E50	**E90**	**ICI**	**E50**	**E90**	**ICI**
LRR (DM known cohort)	**1**	0.0015	0.0035	0.0020	0.0016	0.0022	0.0031
	2	0.0027	0.0076	0.0036	0.0035	0.0043	0.0047
	3	0.0021	0.0050	0.0029	0.0053	0.0072	0.0069
	4	0.0017	0.0025	0.0042	0.0068	0.0083	0.0075
	5	0.0016	0.0074	0.0059	0.0076	0.0106	0.0087
LRR (DM unknown cohort)	**1**	0.0012	0.0029	0.0018	0.0016	0.0022	0.0032
	2	0.0027	0.0076	0.0037	0.0036	0.0043	0.0048
	3	0.0023	0.0052	0.0027	0.0052	0.0070	0.0072
	4	0.0016	0.0023	0.0035	0.0067	0.0080	0.0079
	5	0.0012	0.0064	0.0049	0.0075	0.0114	0.0092
Difference	***1***	*0.0003*	*0.0006*	*0.0002*	*0.0000*	*0.0000*	*-0.0001*
	2	*0.0000*	*-0.0000*	*-0.0001*	*-0.0001*	*0.0000*	*-0.0001*
	3	*-0.0002*	*-0.0002*	*0.0002*	*0.0001*	*0.0002*	*-0.0003*
	4	*0.0001*	*0.0002*	*0.0007*	*0.0001*	*0.0003*	*-0.0004*
	5	*0.0004*	*0.0010*	*0.0010*	*0.0001*	*-0.0008*	*-0.0005*
SP (DM known cohort)	**1**	0.0011	0.0025	0.0014	0.0025	0.0057	0.0043
	2	0.0011	0.0022	0.0013	0.0032	0.0081	0.0057
	3	0.0013	0.0020	0.0015	0.0088	0.0124	0.0083
	4	0.0013	0.0022	0.0014	0.0092	0.0138	0.0090
	5	0.0020	0.0025	0.0022	0.0122	0.0177	0.0111
SP (DM unknown cohort)	**1**	0.0011	0.0027	0.0014	0.0025	0.0059	0.0044
	2	0.0012	0.0024	0.0013	0.0032	0.0085	0.0057
	3	0.0012	0.0020	0.0015	0.0086	0.0129	0.0083
	4	0.0012	0.0021	0.0014	0.0089	0.0141	0.0091
	5	0.0017	0.0023	0.0020	0.0119	0.0182	0.0113
Difference	***1***	*0.0000*	*-0.0002*	*0.0000*	*0.0000*	*-0.0002*	*-0.0001*
	2	*-0.0001*	*-0.0002*	*0.0000*	*0.0000*	*-0.0004*	*0.0000*
	3	*0.0001*	*-0.0000*	*0.0000*	*0.0002*	*-0.0005*	*0.0000*
	4	*0.0001*	*0.0001*	*0.0000*	*0.0003*	*-0.0003*	*-0.0001*
	5	*0.0003*	*0.0002*	*0.0002*	*0.0003*	*-0.0005*	*-0.0002*

The ICI is the integrated calibration index, E50 is the median absolute difference between observed and expected, and E90 is the 90th percentile of the absolute difference. The numbers in red are all above 0.01 and can be considered inadequate.

See Table 5 for the mean risk predictions in the DM known cohort and DM unknown cohort. The mean predicted risk is lower in the DM unknown cohort. Except for the mean predicted risk for LRR in the group where the predicted risk was lower than 5% in the COX approach, and in the RSF approach between 5% and 10% and higher than 10%.

Table 6 shows the mean risk differences and mean relative risk differences for the risk in 5 years for both statistical approaches, and in the groups where the predicted risk was lower than 5%, the predicted risk was higher than 5% and lower than 10% and the predicted risk was higher than 10%. As can be seen in the higher risk predictions the higher the mean and mean relative risk differences. This could be explained because patients with a higher risk for LRR would also have a higher risk for DM. Apart from the mean risk differences, the maximum absolute risk difference in the prediction for LRR with the COX and RSF model is 7.77% and 2.03% respectively, for the prediction of SP with the COX and RSF model this is 0.58% and 0.64%, respectively. When the absolute risk difference was high the predicted risk was also high. In lower predicted risks the absolute risk difference was also lower.

Table 5. Mean risk predictions

		LRR (DM known cohort)	LRR (DM unknown cohort)	SP (DM known cohort)	SP (DM unknown cohort)
COX	Mean predicted risk	3.39% (2.14-4.97)	3.28% (2.06-4.81)	3.24% (2.11-4.66)	3.21% (2.09-4.61)
	Mean predicted risk (risk <5%)	2.06% (1.52-2.43)	2.06% (1.51-2.43)	2.90% (2.10-3.58)	2.88% (2.08-3.56)
	Mean predicted risk (risk 5% - 10%)	6.87% (6.76-6.80)	6.81% (6.68-6.73)	5.94% (5.46-6.51)	5.90% (5.44-6.48)
	Mean predicted risk (risk >10%)	19.48% (17.11-20.10)	18.5% (16.24-19.33)	11.62% (NA-12.90)	11.58% (NA-12.76)
RSF	Mean predicted risk	3.17% (2.13-4.53)	3.11% (2.10-4.43)	3.25% (2.41-4.23)	3.21% (2.38-4.18)
	Mean predicted risk (risk <5%)	2.56% (1.94-3.10)	2.54% (1.92-3.12)	3.14% (2.41-3.70)	3.11% (2.38-3.68)
	Mean predicted risk (risk 5% - 10%)	6.58% (6.57-6.71)	6.59% (6.62-6.77)	5.37% (NA-5.99)	5.33% (NA-6.02)
	Mean predicted risk (risk >10%)	14.86% (12.40-16.66)	14.92% (12.19-16.41)	11.27% (NA-12.91)	10.77% (NA-12.97)

The mean risk predictions in the DM known cohort and DM unknown cohort with their difference.

Table 6. Mean and mean relative risk differences

	COX		RSF	
	LRR	SP	LRR	SP
Mean risk difference	0.11%	0.04%	0.07%	0.04%
Mean relative risk difference	1.01	1.01	1.02	1.01
Mean risk difference (risk <5%)	0.00%	0.03%	0.02%	0.03%
Mean relative risk difference (risk <5%)	1.01	1.01	1.01	1.01
Mean risk difference (risk 5% - 10%)	0.06%	0.04%	0.01%	0.04%
Mean relative risk difference (risk 5% - 10%)	1.04	1.01	1.03	1.01
Mean risk difference (risk >10%)	1.02%	0.04%	0.06%	0.50%
Mean relative risk difference (risk >10%)	1.07	1.04	1.05	1.05
Mean risk difference triple negative patients	0.34%	0.06%	0.15%	0.06%
Mean relative risk difference, triple negative patients	1.04	1.02	1.02	1.02

The mean and mean relative risk differences between the DM known cohort and DM unknown cohort, for the predicted risk in year 5.

In table 7, all thresholds; range 1-20% for LRR and 1-14% for SP, the LRR range goes to 20% because above a risk prediction of 20%, the treatment will not be different and the range stops at 14% for SP because this is the highest risk prediction. More patients would be classified as high risk based on the DM known model, in comparison to the predictions from the DM unknown model. For example, applying a 10% threshold to predict 5-year LRR risk would result in 7 more false negative patients when DM is not modelled as a competing risk (n = 332 vs n = 325).

Table 7. The sensitivity, specificity, positive predicted value and negative predicted value for certain thresholds for the risk prediction of LRR and SP

		LRR	DM known			LRR	DM unknown	
Threshold	false negative	true negative	true positive	false positive	false negative	true negative	true positive	false positive
1%	0	7	385	13102	0	7	385	13102
2%	30	3672	355	9437	31	3766	354	9343
3%	101	8243	284	4866	103	8360	282	4749
4%	207	11189	178	1920	209	11242	176	1867
5%	234	11795	151	1314	238	11863	147	1246
6%	269	12332	116	777	275	12389	110	720
7%	289	12558	96	551	294	12613	91	496
8%	302	12693	83	416	303	12715	82	394
9%	309	12776	76	333	310	12790	75	319
10%	325	12880	60	229	332	12915	53	194
11%	337	12946	48	163	337	12960	48	149
12%	339	12971	46	138	341	12987	44	122
13%	344	12995	41	114	348	13006	37	103
14%	350	13014	35	95	353	13025	32	84
15%	353	13028	32	81	356	13042	29	67
16%	357	13046	28	63	358	13058	27	51
17%	358	13057	27	52	360	13066	25	43
18%	360	13067	25	42	366	13081	19	28
19%	365	13081	20	28	369	13084	16	25
20%	369	13084	16	25	370	13085	15	24
		SP	DM known			SP	DM unknown	
Threshold	false negative	true negative	true positive	false positive	false negative	true negative	true positive	false positive
1%	0	0	411	13083	0	0	411	13083
2%	1	644	410	12439	1	670	410	12413
3%	59	5268	352	7815	61	5396	350	7687
4%	271	11255	140	1828	273	11363	138	1720
5%	353	12538	58	545	357	12556	54	527
6%	401	13029	10	54	403	13042	8	41
7%	405	13061	6	22	405	13065	6	18
8%	406	13076	5	7	406	13077	5	6
9%	407	13083	4	0	408	13083	3	0
10%	409	13083	2	0	409	13083	2	0
11%	410	13083	1	0	410	13083	1	0
12%	410	13083	1	0	411	13083	0	0
13%	411	13083	0	0	411	13083	0	0
14%	411	13083	0	0	411	13083	0	0

Discussion

In this study, multiple comparisons have been performed to illustrate the effect on predicted estimates and model performance between models developed with, and without adjustments for DM as a competing risk for the prediction of LRR and SP. Discrimination was the best in the RSF approach, as in the INFLUENCE 2.0 development paper.[8] Calibration was good with an adequate ICI below 0.01, with an exception for the 5-year ICI for SP in the DM known and unknown cohort. Calibration is better in the INFLUENCE 2.0, possibly explained due to different statistical methods used. The DM known cohort and DM unknown cohort have both good performance with no difference or a difference near 0. As expected, the mean risk differences show that the predictions for the DM unknown cohort are lower than the predictions in the DM known cohort for the 5-year predicted risk, this would be explained due to the longer 'disease free' follow-up time. Because the patient would be censored when DM occurred, when the information about DM is missing the patient would remain in the follow-up as event free until the end of follow-up or another event (LRR or SP). Thus, resulting in an underestimation of the risks. As that may be, the mean predictions close to each other when DM is considered a competing risk and not considered a competing risk. During the data preparation, it came to light that with the development of the INFLUENCE 2.0, the data had been prepared in a way that did not take competing risks into account. This means that the INFLUENCE 2.0 was built as if information about other events was not known. Recommended is that for the development of the INFLUENCE 3.0 SP will be seen and treated as a competing risk for the risk prediction of LRR, and LRR will be seen and treated as a competing risk for the risk prediction of SP.

There are multiple options to consider competing risks. In this study the cause specific hazard model method is used, another option to handle competing risks is the Fine-Gray sub distribution hazard model. Both methods have their advantages and disadvantages, the Fine-Gray method is preferable when the interest is for a single outcome type, thus the all over risk for LRR and SP and not LRR and SP apart from each other. However, the Fine-Gray method can produce risk predictions that exceed 100% as described in the paper of Austin et al.[24] This means, that the Fine-Gray method may have undesirable effects when one wants to estimate the incidence of all of the different competing events. Therefore the cause specific hazard model is used in this study. With this method, the specific risk for LRR and SP can be predicted. The result interpretation for predicted risk with the cause specific hazard model is that the predicted risk is the risk when no other event has occurred. For the predicted risk with the Fine-Gray model, it is the risk regardless of another event occurring. In general, the greater the incidence of competing events, the greater the risk of bias in considering competing events as censoring events. Absolute prevalence of competing events above 10% merits serious consideration, demanding careful attention to the scientific objectives of the analysis and the appropriate choice of endpoint and method of analysis.[14] In our data, the rate of women with DM was below 10% (6.3%).

The data used to perform the analysis does not include patients treated with neoadjuvant therapy, which influences the risk of LRR, SP, and DM[25]. So, the impact of missing information on DM in the patient group who received neoadjuvant therapy could not be assessed. In the available data, only the time to the first event was known, if a patient had LRR first and also had DM only the time to LRR was known in the data. In the data, it was not described for what reason a patient is

Other models have been developed for the risk prediction of breast cancer recurrence. In the study of Giardiello et al.[26] a model is developed that predicts the risk for SP, in this study only DM and death were considered as a competing risk. The Fine-Gray method was used for the model development to take the competing risks into account. Another prediction model is that of Corso et al.[27] this model predicts the risk of LRR recurrence. The paper states that competing risks have been taken into account but it is not said what was seen as competing risks, they mentioned that a patient who died was censored.

The predicted risks and model performances are very similar to each other, it could be said that there is no clinically meaningful difference in the models. There is no established threshold in clinical practice where a patient will be classified as low or high risk. This makes it difficult to say whether there is a real clinically relevant difference or not, for illustrative purposes a wide range of possible thresholds has been made with the corresponding sensitivity, specificity, positive predicted value and negative predicted value. As expected, the DM unknown made an underestimation for the risk of an LRR and SP, it shows that more patients would be classified as low risk and possibly have a different personalised follow-up. This could result in missing more breast cancer recurrences, assuming these recurrences were diagnosed when a patient returned for their follow-up appointment. However, in most cases the recurrence is not detected during routine follow-up appointments, but in interval appointments on demand of the patient.[28] This does not imply that risk predictions are not of use. Risk predictions can have an uncertainty and it is therefore suited that there is no strict threshold. In this case, a range of predicted risks could be given instead of an absolute risk. This could make the risk interpretation more understandable for the clinician and patient.[29]

To conclude, the model predictions and performance difference are minimal, this implies DM does not have a significant role as a competing risk for the prediction of LRR and SP. However, there is no clinical threshold for low or high risk that would say if both prediction models have the same impact in clinical practice. This study assures that data on DM is not crucial for developing models to predict trustworthy estimates for the risk of LRR of SP.

Literature

1. Chhikara, B. S. & Parang, K. *Chemical Biology LETTERS Global Cancer Statistics 2022: the trends projection analysis*. *Chemical Biology Letters Chem. Biol. Lett* vol. 2023 https://pubs.thesciencein.org/cbl.

2. IKNL. Integraal kankercentrum Nederland Cijfers over kanker: Incidentie, Aantal Borstkanker. *https://nkr-cijfers.iknl.nl/#/viewer/facf895a-ef76-4968-8449-d762c0ee3891*.

3. van der Meer, D. J. *et al.* Comprehensive trends in incidence, treatment, survival and mortality of first primary invasive breast cancer stratified by age, stage and receptor subtype in the Netherlands between 1989 and 2017. *Int J Cancer* **148**, 2289–2303 (2021).

4. IKNL. Integraal kankercentrum Nederland Cijfers over kanker: Overleving borst. *https://applicatie.nkr-cijfers.nl*.

5. Eijkelboom, A. H. *et al.* Routine and interval detection of locoregional breast cancer recurrences and risk of subsequent distant metastasis. *Breast Cancer Res Treat* **197**, 123–135 (2023).

6. Witteveen, A. *et al.* Personalisation of breast cancer follow-up: a time-dependent prognostic nomogram for the estimation of annual risk of locoregional recurrence in early breast cancer patients. *Breast Cancer Res Treat* **152**, 627–636 (2015).

7. Ankersmid, J. W. *et al.* Supporting Shared Decision-making About Surveillance After Breast Cancer With Personalized Recurrence Risk Calculations: Development of a Patient Decision Aid Using the International Patient Decision AIDS Standards Development Process in Combination With a Mixed Methods Design. *JMIR Cancer* **8**, (2022).

8. Völkel, V. *et al.* Improved risk estimation of locoregional recurrence, secondary contralateral tumours and distant metastases in early breast cancer: the INFLUENCE 2.0 model. *Breast Cancer Res Treat* **189**, 817–826 (2021).

9. Chiba, A. *et al.* Trends in Neoadjuvant Endocrine Therapy Use and Impact on Rates of Breast Conservation in Hormone Receptor-Positive Breast Cancer: A National Cancer Data Base Study. *Ann Surg Oncol* **24**, 418–424 (2017).

10. de Munck, L. *et al.* Impact of mammographic screening and advanced cancer definition on the percentage of advanced-stage cancers in a steady-state breast screening programme in the Netherlands. *Br J Cancer* **123**, 1191–1197 (2020).

11. Van Steenbeek, C. D. *et al.* Facilitating validation of prediction models: A comparison of manual and semi-automated validation using registry-based data of breast cancer patients in the Netherlands. *BMC Med Res Methodol* **19**, (2019).

12. Putter, H., Fiocco, M. & Gekus, R. B. Tutorial in biostatistics: Competing risk and multi-state models. *Stat Med* **26**, 2389–2430 (2007).

13. Moossdorff, M. *et al.* Maastricht Delphi consensus on event definitions for classification of recurrence in breast cancer research. *J Natl Cancer Inst* **106**, (2014).

14. Austin, P. C., Lee, D. S. & Fine, J. P. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* **133**, 601–609 (2016).

15. George, B., Seals, S. & Aban, I. Survival analysis and regression models. *Journal of Nuclear Cardiology* vol. 21 686–694 Preprint at https://doi.org/10.1007/s12350-014-9908-2 (2014).

16. Ishwaran, H. Variable importance in binary regression trees and forests. *Electron J Stat* **1**, 519–537 (2007).

17. Breiman, L. *Random Forests*. vol. 45 (2001).

18. Schemper, M. & Smith, T. L. *A Note on Quantifying Follow-up in Studies of Failure Time*. (1996) doi:https:// doi. org/ 10. 1016/ 0197- 2456(96) 00075-x.

19. Steyerberg, E. W. *Statistics for Biology and Health Clinical Prediction Models A Practical Approach to Development, Validation, and Updating Second Edition*. http://www.springer.com/series/2848 (2009).

20. Moons, K. G. M. *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* vol. 98 691–698 Preprint at https://doi.org/10.1136/heartjnl-2011-301247 (2012).

21. Austin, P. C. & Steyerberg, E. W. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med* **38**, 4051–4065 (2019).

22. Kamarudin, A. N., Cox, T. & Kolamunnage-Dona, R. Time-dependent ROC curve analysis in medical research: Current methods and applications. *BMC Med Res Methodol* **17**, (2017).

23. Anders, C. K. & Carey, L. A. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin Breast Cancer* **9**, (2009).

24. Austin, P. C., Putter, H., Lee, D. S. & Steyerberg, E. W. Estimation of the Absolute Risk of Cardiovascular Disease and Other Events: Issues with the Use of Multiple Fine-Gray Subdistribution Hazard Models. *Circulation: Cardiovascular Quality and Outcomes* vol. 15 E008368 Preprint at https://doi.org/10.1161/CIRCOUTCOMES.121.008368 (2022).

25. Liu, H., Lv, L., Gao, H. & Cheng, M. Pathologic Complete Response and Its Impact on Breast Cancer Recurrence and Patient's Survival after Neoadjuvant Therapy: A Comprehensive Meta-Analysis. *Comput Math Methods Med* **2021**, (2021).

26. Giardiello, D. *et al.* PredictCBC-2.0: a contralateral breast cancer risk prediction model developed and validated in ~200,000 patients. doi:10.21203/rs.3.rs-1767532/v1.

27. Corso, G. *et al.* Validation of a Novel Nomogram for Prediction of Local Relapse after Surgery for Invasive Breast Carcinoma. *Ann Surg Oncol* **27**, 1864–1874 (2020).

28. Beltran-Bless, A. A. *et al.* Regularly scheduled physical examinations and the detection of breast cancer recurrences. *Breast* **69**, 274–280 (2023).

29. Ankersmid, J. W., Drossaert, C. H. C., van Riet, Y. E. A., Strobbe, L. J. A. & Siesling, S. Needs and preferences of breast cancer survivors regarding outcome-based shared decision-making about personalised post-treatment surveillance. *Journal of Cancer Survivorship* (2022) doi:10.1007/s11764-022-01178-z.