Validation of the online prediction tool PREDICT v. 2.0 in the Dutch breast cancer population

M.C. van Maaren[1,2], C.D. van Steenbeek[1,2], P.D.P. Pharoah[3], A. Witteveen[2], G.S. Sonke[4], L.J.A. Strobbe[5], P.M.P. Poortmans[6], S. Siesling[1,2]

[1] Department of Research, Netherlands Comprehensive Cancer Organisation, Utrecht, the Netherlands

[2] Department of Health Technology & Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, the Netherlands

[3] Department of Oncology, University of Cambridge, Cambridge, United Kingdom

[4] Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, the Netherlands

[5] Department of Surgical Oncology, Canisius Wilhelmina Hospital, Nijmegen, the Netherlands

[6] Department of Radiation Oncology, Institut Curie, Paris, France

Corresponding address:

M.C. van Maaren

Netherlands Comprehensive Cancer Organisation

P.O. Box 19079

3501 DB Utrecht

The Netherlands

m.vanmaaren@iknl.nl

Tel. +31 88-2346000

Abstract

Background

PREDICT version 2.0 is increasingly used to estimate prognosis in breast cancer. This study aimed to validate this tool in specific prognostic subgroups in the Netherlands.

Methods

All operated women with non-metastatic primary invasive breast cancer, diagnosed in 2005, were selected from the nationwide Netherlands Cancer Registry. Predicted and observed 5- and 10-year overall survival (OS) were compared for the overall cohort, separated by oestrogen receptor (ER) status, and predefined subgroups. A >5% difference was considered as clinically relevant. Discriminatory accuracy and goodness-of-fit were determined using the area under the receiver operating characteristic curve (AUC) and the Chi2-test.

Results

We included 8,834 patients. Discriminatory accuracy for 5-year OS was good (AUC 0.80). For ER+ and ER- patients, AUCs were 0.79 and 0.75, respectively. Predicted 5-year OS differed from observed by -1.4% in the entire cohort, -0.7% in ER+ and -4.9% in ER- patients. Five-year OS was accurately predicted in all subgroups. Discriminatory accuracy for 10-year OS was good (AUC 0.78). For ER+ and ER- patients AUCs were 0.78 and 0.76, respectively. Predicted 10-year OS differed from observed by -1.0% in the entire cohort, -0.1% in ER+ and -5.3 in ER- patients. Ten-year OS was overestimated (6.3%) in patients ≥75 years and underestimated (-13.%) in T3 tumours and patients treated with both endocrine therapy and chemotherapy (-6.6%).

Conclusions

PREDICT predicts OS reliably in most Dutch breast cancer patients, although results for both 5-year and 10-year OS should be interpreted carefully in ER- patients. Furthermore, 10-year OS should be interpreted cautiously in patients ≥75 years, T3 tumours and in patients considering endocrine therapy and chemotherapy.

Keywords: PREDICT; prediction model; breast cancer; validation; overall survival; population-based study

Word counts

Abstract: 264

Body of text: 2733

Number of tables: 1

Number of figures: 4

Introduction

Adjuvant systemic therapy is shown to reduce recurrence rates in breast cancer patients[1, 2]. In the current era of personalised cancer medicine, limiting under- and overtreatment is increasingly important to optimize the therapeutic benefit while minimizing short- and long-term side effects of treatment[3]. To personalize breast cancer treatment, several prediction tools have been developed including the online tools Adjuvant! Online[4] and PREDICT[5]. Adjuvant! Online is developed using the Surveillance, Epidemiology and End-Results (SEER) registry and predicts 10-year risks for recurrence, breast cancer specific-mortality and mortality due to other causes, including the expected benefit of adjuvant systemic treatments based on patient- and tumour-related factors. In the Netherlands, the online prediction tool Adjuvant! Online[4] has been widely used in clinical practice[6]. However, Adjuvant! Online has been described to overestimate survival outcomes in several breast cancer populations[7-11]. PREDICT was developed using cancer registry data from the United Kingdom (UK) and predicts 5-year and 10-year overall survival (OS) for individual breast cancer patients, based on several patient- and tumour-related characteristics[5]. It also provides the expected benefits of chemotherapy, endocrine therapy and trastuzumab. PREDICT version 1 was released in 2011 and the use of the tool increased steadily until 2016, whereafter its use increased substantially following disabling of Adjuvant! Online[12]. PREDICT version 1 has been validated on multiple independent case-cohorts from several countries including the UK, Canada, Malaysia and the Netherlands[13-18].

Yet PREDICT has not been validated on the entire Dutch breast cancer population. Moreover, a new version of PREDICT, version 2.0, has become available recently[19]. In version 2.0, the model was improved with the addition of the options of using the exact tumour size in millimetres, the exact number of positive lymph nodes, and the presence of micrometastases. In addition, follow-up was extended.

This study aimed to validate the online prediction tool PREDICT version 2.0 in a large population-based cohort in the Netherlands. Separate analyses were performed to study its validity in specific prognostic subgroups.

Methods

Design

In this large historic population-based cohort study, data on patient-, tumour- and treatment-related characteristics were obtained from the Netherlands Cancer Registry (NCR). The NCR is hosted by the Netherlands Comprehensive Cancer Organisation (IKNL), employing trained registrars to derive data of all patients newly diagnosed with cancer directly from patient records. Tumour topography and morphology were coded according to the International Classification of Diseases for Oncology, 3rd edition[20]. Staging was coded according to the tumour, node and metastasis (TNM) classification system, 6th edition[21]. Additional data on vital status and date of death were derived from the Municipal Personal Records database, which was complete until February 2016.

Patients

All women diagnosed with non-metastatic primary invasive breast cancer in 2005 in the Netherlands, who received surgery as part of their treatment, were included. Patients who received primary systemic therapy or had no pathologically established tumour were excluded. In addition, patients with unknown tumour size, number of positive lymph nodes, differentiation grade or oestrogen receptor (ER) status were excluded, since PREDICT does not allow missing values for these variables.

Statistical analysis

The outcomes of interest were 5-year and 10-year OS. The original script of PREDICT was used to calculate the expected OS and this was compared with observed OS obtained from the NCR. Comparisons were performed for the overall cohort, separated by ER status, and for subgroups based on age, stage, presence of micrometastases, grade, HER2 status, type of surgery, use and type of adjuvant systemic therapy and generation chemotherapy. A 95% confidence interval (CI) was calculated for the observed numbers. The predicted proportions are no real proportions, but reflect the sum of all predictions for each individual. Since uncertainties around these predictions were not build into PREDICT, confidence intervals around the prediction estimates and differences could not be calculated. To assess goodness-of-fit of the model in each subgroup, observed and average predicted events were compared using a Chi2-test. Mode of detection and Ki67 status were set to unknown for each patient since these variables were not registered in the NCR. Consequently, PREDICT uses the weighted mean coefficient of the unknown variable for these patients. For 2,278 patients who received chemotherapy, it was unknown which generation chemotherapy was administered. For this reason, every analysis was performed four times. Patients were classified as second generation, third generation, a combination of second and third generation (meaning that for these patients the mean coefficient of second generation and third generation was used in the predictions) or they were excluded. The analyses were compared, and no significant differences in calibration and discrimination between the four methods were observed, both for 5-year and 10-year OS (Supplementary Figure 1 to 4). Based on these results, it was decided to include all patients with an unknown chemotherapy generation and classify them as a combination of second and third generation for all further analyses.

Since PREDICT has been generated on ER+ and ER- patients separately, overall results were stratified by ER status, and graphical discrimination and calibration were determined. Discrimination was graphically shown in a receiver operating characteristic (ROC) curve. Here, the sensitivity (the proportion of patients who survived and were predicted correctly) was plotted against the 1-specificity (the proportion of patients who did not survive but were predicted as they would have survived). The discriminatory accuracy was quantified by the area under the ROC curve (AUC). An AUC of 0.5 indicates that the model performs as good as flipping a coin, whereas an AUC of 1 indicates perfect discrimination. In addition, model calibration was determined by plotting the averages of the

observed against the predicted outcomes [with 95% confidence interval (CI)], grouped by quintiles based on the predicted estimates. The estimates were subsequently compared with the perfect prediction line (y=x). An a priori assumption was that PREDICT accurately predicted OS whenever the differences between predicted and observed outcomes were within a range of 5%, since differences outside of this range were considered as clinically relevant. A p-value<0.05 was considered as statistically significant. All statistical analyses were performed in STATA version 14.1.

Results

In total, 10,338 patients with operated, non-metastatic primary invasive breast cancer, diagnosed in 2005 were identified. Patients receiving primary systemic therapy (n=529), patients without a pathologically established tumour stage (n=2) or unknown tumour size in millimetres (n=332), missing number of positive lymph nodes (n=118), unknown differentiation grade (n=418) or ER status (n=105) were excluded, leading to a final study population of 8,834 patients (85% of the total population). Most patients presented with T1 (60.4%), N0 stage (62.2%), grade II (45.1%), ER positive (82.2%) and HER2 negative disease (69.8%). The majority of the patients did not receive adjuvant systemic therapy (61.7%) (Table 1). The median age was 58 years (interquartile range: 49-69 years). Median follow-up of this population was 10.4 years (interquartile range 9.2-10.7).

Discriminatory accuracy and calibration for 5-year OS

In the entire Dutch validation population, discriminatory accuracy for 5-year OS was good with an AUC of 0.80. Within ER+ and ER- patients separately, the AUCs were 0.79 and 0.75, respectively (Figure 1). The predicted number of survivors after 5 years in the entire cohort was 7595.2 (86.0%) compared to 7723 (87.4%) observed survivors (Table 1). The difference was -1.4%, which was not significant (p=0.14). In ER+ patients, the difference between predicted and observed events was -0.7% (p=0.53). In ER- patients, the difference between predicted and observed events was -4.9%, which was statistically significant (p=0.02), but just within the range of 5% (Table 1). Figure 2 shows the predicted and observed 5-year OS by quintiles of the predicted survival. For the entire cohort and for ER+ patients, the predicted and observed 5-year OS do not differ significantly. However, for ER- patients, the predicted 5-year OS was significantly lower than the observed 5-year OS, with the largest deviations seen in the lowest and highest quintiles (Figure 2). For patients treated with breast-conserving surgery, 5-year OS was slightly underestimated with 2.9% (p=0.03), but this was within the range of 5%. For all other predefined subgroups no statistically significant differences between predicted and observed events were observed (Table 1).

Discriminatory accuracy and calibration for 10-year OS

In the entire Dutch validation population, discriminatory accuracy for 10-year OS was good with an AUC of 0.78. Within ER+ and ER- patients separately, the AUCs were 0.78 and 0.76, respectively (Figure 3). The predicted number of survivors after 10 years in the entire cohort was 6404 (72.5%) compared to 6493 (73.5%) observed events. The difference was -1.0%, which was not significant (p=0.27). In ER+ patients, the difference between predicted and observed events was -0.1% (p=0.92). In ER- patients, the difference between predicted and observed events was -5.3%, which was statistically significant (p=0.01). (Table 1). Figure 3 shows the predicted and observed 10-year OS by quintiles of the predicted survival. For the entire cohort and for ER+ patients, the predicted 5-year OS did not differ from the observed 5-year OS. However, for ER- patients a significant underestimation was seen (p=0.01), with the most pronounced deviations in the two highest quintiles. In the predefined subgroups, a significant overestimation (6.3%, p<0.01) of 10-year OS was observed in patients ≥75 years. Ten-year OS was significantly underestimated by PREDICT in T3 tumours (-13%, p<0.01), grade III (-3.2%, p=0.03), patients treated with breast-conserving surgery (-3.0%, p=0.02), patients treated with chemotherapy only (-4.8%, p=0.01, patients treated with both endocrine therapy and chemotherapy (-6.6%), p=0.03) and in patients with an unknown generation of chemotherapy (-4.4%, p=0.02). However, the only differences outside the range of 5%, were in patients ≥75 years (overestimation), T3 tumours and patients treated with both endocrine therapy and chemotherapy (underestimation).

Discussion

PREDICT version 2.0 accurately predicts 5-year OS in the entire Dutch validation population and in all predefined subgroups. Ten-year survival was predicted quite well, although underestimation was observed in ER- patients, T3 tumours and patients treated with both endocrine therapy and chemotherapy. In addition, 10-year OS was overestimated in patients ≥75 years. Of note, 5-year OS for ER- patients was underestimated (4.9%). Although this difference was within the range of 5%, it was statistically significant and became larger when estimating 10-year OS.

The finding that 10-year OS was underestimated in ER- patients, but was accurately predicted in ER+ patients is in contrast to a previous validation study of PREDICT version 2.0 where slightly better predictions were reported for ER- patients than for ER+ patients[19]. A possible explanation for this discrepancy may be the inclusion of different populations (UK versus Dutch population). Two previous studies in the Netherlands have validated PREDICT version 1.2. The first, only including patients ≥65 years, showed that PREDICT version 1.2 largely overestimated 5-year OS in patients ≥85 years, and that 10-year OS was highly overestimated in patients ≥75 years[17]. The second, only including patients <50 years, showed that PREDICT version 1.2 accurately predicted 10-year OS in patients <50 years, but that it was underestimated for patients <40 years[18]. Another validation study of the 1.2 version in the UK in patients <40 years showed that 10-year OS was accurately predicted, but that 5-year OS was highly overestimated with 25%[15]. The updated version 2.0 is described to provide more

accurate predictions in all age groups in three independent validation cohorts, in contrast to version 1.2[14], suggesting that the newest version may perform better in all subgroups based on age[19]. Our study confirms this for 5-year OS, which was accurately predicted in all subgroups based on age. However, for patients ≥75 years, we still found an overestimation for 10-year OS. Of note, this was to a much lesser extent than the previous validation study in the Netherlands[17]. The still existing overestimation of 10-year OS in patients ≥75 years may partially be explained by the prevalence of comorbidities in older patients. PREDICT gives survival estimates for individual patients based on the average comorbidity for women with breast cancer of a similar age[12]. However, any increased prevalence of comorbidities or an overrepresentation of older patients in our cohort may have led to a lower OS.

Our findings that PREDICT version 2.0 underestimates 10-year OS in T3 tumours and patients treated with both endocrine therapy and chemotherapy may partly be explained by the fact that the PREDICT model has been generated on a population diagnosed from 1999 to 2003 in the UK. Our validation population consisted of patients diagnosed in 2005, in which differences in OS may partly reflect increased survival over time. Another explanation may be the differences in health care provided in the Netherlands and the UK. The large underestimations in patients with T3 stage is most likely to be explained by an underrepresentation of this group in the UK development population[5], but it may also be caused by differences in treatment strategies or other prognostic characteristics differing between our validation population and the development population. Of course, patients with T3 stage have a poorer prognosis compared to patients with T1 or T2 stage and have a higher likelihood of being treated with adjuvant systemic therapy. Since PREDICT is designed to assist in taking treatment decisions, the underestimation of 10-year OS may be therefore of less relevance in this specific patient group as far as decision-making concerns. It remains of course a shortage for the prediction of outcome to inform the patients. Similarly, for the underestimation of 10-year OS in patients treated with both endocrine therapy and chemotherapy, it should be noted that treatment effects cannot be extracted from predictions of the model, since the patients included in the validation already received their treatment, independent of the predictions in the model.

The accuracy of a prediction model can be interpreted in different ways. Several studies consider a difference between predicted and observed outcomes of less than 2% as accurate[8, 22], while other studies base their interpretation on the 95%CI and corresponding p-value[16, 18]. According to the Dutch national guidelines, adjuvant systemic therapy is recommended in case of an absolute risk of 10-year breast cancer-related mortality of 15% or more. With the 15-year relative mortality reductions described by the EBCTCG ranging between 20 to 57%[1], the absolute mortality reduction for most patients will be at least 4-5%[23]. A significant difference between predicted and observed survival of 5% may therefore systematically be accepted as the threshold to alter treatment decisions and was selected as a measure for whether or not differences between predicted and observed outcomes are of clinical relevance.

Importantly, we should be aware that any inaccurate prediction for a specific subgroup as a whole does not mean that this is applicable to every single patient in this subgroup.

Strengths and limitations

To our knowledge, this is the first population-based study in the Netherlands covering the entire PREDICT target population that validates PREDICT version 2.0 in specific prognostic subgroups. The population-based setting and the large number of included patients increase the generalisability and reliability of the results.

A limitation of this study is the absence of knowledge on cause-specific mortality, preventing us from determining whether discrepancies are due to breast cancer-specific mortality or other causes of death. Another limitation of this study is lacking data on mode of detection and Ki67, which were set to unknown for all patients accordingly. Symptomatic cancers are more likely to present with unfavourable tumour characteristics compared to screen-detected cancers[24]. Thereby, this limitation may partly be neutralised by inclusion of the most important prognostic tumour characteristics in the PREDICT model. However, even after correction for these characteristics, mode of detection remains associated with survival outcomes[24] and may therefore have affected our results. Furthermore, Ki67 has been described to play an important role in breast cancer prognosis[25], and not taking Ki67 status into account may have affected the results. However, substantial heterogeneity is observed in the methods of Ki67 assessment[26], which may limit its usefulness in determining a patient's prognosis and may explain why it is not routinely used in the Netherlands.

Conclusions

PREDICT accurately predicts 5-year OS in the overall Dutch validation population, and in all predefined subgroups. Although within the range of 5%, 5-year OS for ER- patients should be interpreted with care. Furthermore, 10-year OS was overestimated for patients ≥75 years, and underestimated for ER- disease, T3 tumours and for patients receiving both endocrine therapy and chemotherapy. Given PREDICT's intentions to guide treatment decision-making, PREDICT may serve as a reliable prediction tool for the Dutch breast cancer population. However, 5-year OS should be interpreted cautiously in ER- disease, and 10-year OS should be interpreted with care in patients ≥75 years, ER- disease, T3 tumours and in patients considering endocrine therapy and chemotherapy.

Acknowledgements

We thank the Netherlands Cancer Registry for providing the data, as well as the registration clerks for their effort in gathering the data in the Netherlands Cancer Registry.

Conflicts of interest statement

None declared

Role of the funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365(9472):1687-717.

[2] Perez EA, Romond EH, Suman VJ, Jeong JH, Sledge G, Geyer CE, Jr., et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. J Clin Oncol 2014;32(33):3744-52.

[3] Katz SJ, Morrow M. Addressing overtreatment in breast cancer: The doctors' dilemma. Cancer 2013;119(20):3584-8.

[4] Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001;19(4):980-91.

[5] Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast cancer research : BCR. 2010;12(1):R1.

[6] Engelhardt EG, Pieterse AH, van Duijn-Bakker N, Kroep JR, de Haes HC, Smets EM, et al. Breast cancer specialists' views on and use of risk prediction models in clinical practice: a mixed methods approach. Acta oncol 2015;54(3):361-7.

[7] Bhoo-Pathy N, Yip CH, Hartman M, Saxena N, Taib NA, Ho GF, et al. Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. Eur J Cancr 2012;48(7):982-9.

[8] Campbell HE, Taylor MA, Harris AL, Gray AM. An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. Br J Cancer 2009;101(7):1074-84.

[9] de Glas NA, van de Water W, Engelhardt EG, Bastiaannet E, de Craen AJ, Kroep JR, et al. Validity of Adjuvant! Online program in older patients with breast cancer: a population-based study. Lancet Oncol 2014;15(7):722-9.

[10] Mook S, Schmidt MK, Rutgers EJ, van de Velde AO, Visser O, Rutgers SM, et al. Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study. Lancet Oncol 2009;10(11):1070-6.

[11] Yao-Lung K, Dar-Ren C, Tsai-Wang C. Accuracy validation of adjuvant! online in Taiwanese breast cancer patients--a 10-year analysis. BMC Med Inform Decis Mak 2012;12:108.

[12] Eastern Cancer Registry and Information Centre. PREDICT, http://www.predict.nhs.uk/; 2017 [accessed 13.04.17].

[13] Wishart GC, Bajdik CD, Azzato EM, Dicks E, Greenberg DC, Rashbass J, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. Eur J Surg Oncol 2011;37(5):411-7.

[14] Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. Br J Cancer 2012;107(5):800-7.

[15] Maishman T, Copson E, Stanton L, Gerty S, Dicks E, Durcan L, et al. An evaluation of the prognostic model PREDICT using the POSH cohort of women aged 40 years at breast cancer diagnosis. Br J Cancer 2015;112(6):983-91.

[16] Wong HS, Subramaniam S, Alias Z, Taib NA, Ho GF, Ng CH, et al. The predictive accuracy of PREDICT: a personalized decision-making tool for Southeast Asian women with breast cancer. Medicine (Baltimore) 2015;94(8):e593.

[17] de Glas NA, Bastiaannet E, Engels CC, de Craen AJ, Putter H, van de Velde CJ, et al. Validity of the online PREDICT tool in older patients with breast cancer: a population-based study. Br J Cancer 2016;114(4):395-400.

[18] Engelhardt EG, van den Broek AJ, Linn SC, Wishart GC, Rutgers EJT, van de Velde AO, et al. Accuracy of the online prognostication tools PREDICT and Adjuvant! for early-stage breast cancer patients younger than 50 years. Eur J Cancer 2017;78:37-44.

[19] Candido Dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. Breast Cancer Res 2017;19(1):58.

[20] Fritz A PC, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S. International Classification of Diseases for Oncology, 3rd edition. Geneva: World Health Organization. 2000.

[21] Sobin LH WC. International Union Against Cancer, TNM classification of malignant tumours, 6th edition. New York: Wiley. 2002.

[22] Olivotto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BD, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23(12):2716-25.

[23] NABON. Dutch Guideline Breast Cancer (Landelijke richtlijn mammacarcinoom), http://www.oncoline.nl/mammacarcinoom; 2012 [accessed 13.04.17].

[24] Kobayashi N, Hikichi M, Ushimado K, Sugioka A, Kiriyama Y, Kuroda M, et al. Differences in subtype distribution between screen-detected and symptomatic invasive breast cancer and their impact on survival. Clin Trans Oncol 2017; doi: 10.1007/s12094-017-1660-z

[25] Niikura N, Masuda S, Kumaki N, Xiaoyan T, Terada M, Terao M, et al. Prognostic significance of the Ki67 scoring categories in breast cancer subgroups. Clin Breast Cancer 2014;14(5):323-9 e3.

[26] Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 2011;103(22):1656-64.

Table 1. Observed and predicted 5- and 10-year overall survival by patient-, tumour- and treatment-related characteristics

		5-year OS				10-year OS			
	N (%)	**Predicted (%)**	**Observed (%, 95% CI)**	**Difference (%)**	**p-value***	**Predicted (%)**	**Observed (%, 95% CI)**	**Difference (%)**	**p-value***
Entire cohort	8,834 (100)	7595.2 (86.0)	7723 (87.4) (86.7 – 88.1)	-127.8 (-1.4)	0.14	6404.2 (72.5)	6493 (73.5) (72.6 – 74.4)	-88.8 (-1.0)	0.27
ER status									
Positive	7,263 (82.2)	6466.9 (89.0)	6517 (89.7) (89.0 – 90.4)	-50.1 (-0.7)	0.53	5454.2 (75.1)	5460 (75.2) (74.2 – 76.2)	-5.8 (-0.1)	0.92
Negative	1,571 (17.8)	1128.3 (71.8)	1206 (76.8) (74.6 – 78.8)	-77.5 (-4.9)	**0.02**	950.0 (60.5)	1033 (65.8) (63.3 – 68.1)	-83.0 (-5.3)	**0.01**
Age									
<40	495 (5.6)	437.5 (88.4)	451 (91.1) (88.3 – 93.5)	-13.5 (-2.7)	0.52	393.3 (79.5)	404 (81.6) (77.9 – 84.9)	-10.7 (-2.2)	0.59
40-49	1,735 (19.6)	1562.7 (90.1)	1602 (92.3) (91.0 – 93.5)	-39.3 (-2.2)	0.32	1413.7 (81.5)	1485 (84.0) (83.8 – 87.2)	-71.3 (-2.6)	0.06
50-64	3,529 (40.0)	3173.0 (90.0)	3247 (92.0) (91.1 – 92.9)	-74.0 (-2.1)	0.19	2826.8 (80.1)	2918 (82.7) (81.4 – 83.9)	-91.2 (-2.6)	0.09
65-74	1,869 (21.2)	1606.5 (6.0)	1644 (88.0) (86.4 – 89.4)	-37.5 (-2.0)	0.35	1302.1 (69.7)	1321 (70.7) (68.6 – 71.7)	-18.9 (-1.0)	0.60
≥75	1,206 (13.7)	815.5 (67.6)	779 (64.6) (61.8 – 67.3)	36.5 (3.0)	0.20	468.2 (38.8)	392 (32.5) (29.9 – 35.2)	76.2 (6.3)	**0.00**
T stage									
1	5,331 (60.4)	4822.1 (90.5)	4912 (92.1) (91.4 – 92.8)	-89.9 (-1.7)	0.19	4233.0 (79.4)	4280 (80.3) (79.2 – 81.4)	-47.0 (-0.9)	0.47
2	3,176 (36.0)	2563.9 (80.7)	2592 (81.6) (80.3 – 83.0)	-28.1 (-0.9)	0.58	2035.7 (64.1)	2058 (64.8) (63.1 – 66.5)	-22.3 (-0.7)	0.62
3	230 (2.6)	145.8 (63.4)	164 (71.3) (65.5 – 77.2)	-18.2 (-7.9)	0.13	95.2 (41.4)	125 (54.3) (47.7 – 60.9)	-29.8 (-13.0)	**0.00**
4	97 (1.1)	63.4 (65.4)	55 (56.7) (46.3 – 66.7)	8.4 (8.7)	0.29	40.2 (41.5)	30 (30.9) (21.7 – 40.1)	10.2 (10.6)	0.11
N stage									
0	5,496 (62.2)	4908.9 (89.3)	4994 (90.9) (90.1 – 91.6)	-85.1 (-1.5)	0.23	4257.0 (77.5)	4290 (78.1) (77.0 – 79.2)	-33.0 (-0.6)	0.61

1	2,408 (27.3)	2057.5 (85.4)	2078 (86.3) (84.9 – 87.7)	-20.5 (-0.8)	0.65	1715.6 (71.2)	1723 (71.6) (69.8 – 73.4)	-7.4 (-0.3)	0.86
2	604 (6.8)	439.2 (72.7)	455 (75.3) (71.9 – 78.8)	-15.8 (-2.6)	0.45	320.1 (53.0)	349 (57.8) (53.9 – 61.7)	-28.9 (-4.8)	0.11
3	326 (3.7)	189.6 (58.1)	196 (60.1) (54.8 – 65.4)	-6.4 (-2.0)	0.64	111.4 (34.2)	131 (40.2) (34.9 – 45.5)	-19.6 (-6.0)	0.06
Presence of micrometastases									
No	8,321 (94.2)	7130.6 (85.7)	7251 (87.1) (86.4 – 87.9)	-120.4 (-1.4)	0.15	5995.5 (72.1)	6090 (73.2) (72.2 – 71.1)	-94.5 (-1.1)	0.22
Yes	513 (5.8)	464.5 (90.6)	472 (92.0) (89.7 – 94.4)	-7.5 (-1.5)	0.73	408.7 (79.7)	403 (78.6) (75.0 – 82.1)	5.7 (1.1)	0.78
Grade									
I	1,992 (22.6)	1841.2 (92.4)	1856 (93.2) (92.1 – 94.3)	-14.8 (-0.7)	0.73	1632.5 (82.0)	1620 (81.3) (79.6 – 83.0)	12.5 (0.6)	0.75
II	3,983 (45.1)	3507.5 (88.1)	3546 (89.0) (88.1 – 90.0)	51.5 (-1.0)	0.38	2968.3 (74.5)	2979 (74.8) (73.4 – 76.1)	-10.7 (-0.3)	0.84
III	2,859 (32.4)	2246.5 (78.6)	2321 (81.2) (79.7 – 82.6)	-74.5 (-2.6)	0.12	1803.3 (63.1)	1894 (66.2) (64.5 – 68.0)	-90.7 (-3.2)	**0.03**
HER2 status									
Negative	6,169 (69.8)	5350.0 (86.7)	5433 (88.1) (87.3 – 88.9)	-83.0 (-1.3)	0.26	4524.0 (73.3)	4580 (74.2) (73.2 – 75.3)	-56.0 (-0.9)	0.41
Positive	1,149 (13.0)	950.3 (82.7)	1001 (87.1) (85.2 – 89.1)	-50.7 (-4.4)	0.10	804.5 (70.0)	853 (74.2) (71.7 – 76.8)	-48.5 (-4.2)	0.09
Unknown	1,516 (17.2)	1294.8 (85.4)	1289 (85.0) (83.2 – 86.8)	5.8 (0.4)	0.86	1075.6 (71.0)	1060 (69.9) (67.6 – 72.2)	15.6 (1.0)	0.63
Type of surgery									
Breast-conserving surgery	5,070 (57.4)	4563.0 (90.0)	4709 (92.9) (92.2 – 93.6)	-146.0 (-2.9)	**0.03**	4008.5 (79.1)	4162 (82.1) (81.0 – 83.1)	-153.5 (-3.0)	**0.02**
Mastectomy	3,764 (42.6)	3032.2 (80.1)	3014 (80.1) (78.8 – 81.4)	18.2 (0.5)	0.74	2395.7 (63.6)	2331 (61.9) (60.4 – 63.5)	64.7 (1.7)	0.19
Adjuvant systemic therapy									
No	5,454 (61.7)	4730.9 (86.7)	4748 (87.1)	-17.2 (-0.3)	0.81	3952.2 (72.5)	3888 (71.3)	64.2 (1.1)	0.31

			(86.2 – 87.9)				(70.1 – 72.5)		
Only endocrine therapy	214 (2.4)	172.4 (80.6)	165 (77.1) (71.5 – 82.7)	7.4 (3.5)	0.57	126.6 (59.2)	112 (52.3) (45.6 – 59.0)	14.6 (6.8)	0.19
Only chemotherapy	2,306 (26.1)	1904.4 (82.6)	1985 (86.1) (84.7 – 87.5)	-80.6 (-3.5)	0.06	1632.3 (70.8)	1743 (75.6) (73.8 – 77.3)	-110.7 (-4.8)	**0.01**
Both	860 (9.7)	787.5 (91.6)	825 (95.9) (94.6 – 97.3)	-37.5 (-4.4)	0.18	693.1 (80.6)	750 (87.2) (85.0 – 89.4)	-56.9 (-6.6)	**0.03**
Generation chemotherapy									
No chemotherapy	5,668 (64.2)	4903.3 (86.5)	4913 (86.7) (85.8 – 87.6)	-9.7 (-0.2)	0.89	4078.8 (72.0)	4000 (70.6) (69.4 – 71.8)	78.8 (1.4)	0.22
Generation 2	615 (7.0)	555.1 (90.3)	585 (95.1) (93.4 – 96.8)	-29.9 (-4.9)	0.20	491.3 (79.9)	532 (86.5) (83.8 – 89.2)	-40.7 (-6.6)	0.07
Generation 3	416 (4.7)	355.7 (85.5)	378 (90.9) (88.1 – 93.6)	-22.3 (-5.3)	0.24	311.0 (74.8)	344 (82.7) (79.1 – 86.3)	-33.0 (-7.9)	0.06
Generation unknown	2,135 (24.2)	1781.0 (83.4)	1847 (86.5) (85.1 – 88.0)	-66.0 (-3.1)	0.12	1523.1 (71.3)	1617 (75.7) (73.9 – 77.6)	-93.9 (-4.4)	**0.02**

Abbreviations: N = total number, SE = standard error, CI = confidence interval, ER = oestrogen receptor, HER2 = human epidermal growth factor receptor.

*The p-value was calculated by using a Chi2-test. P-values indicated in bold are considered as statistically significant ($p<0.05$)

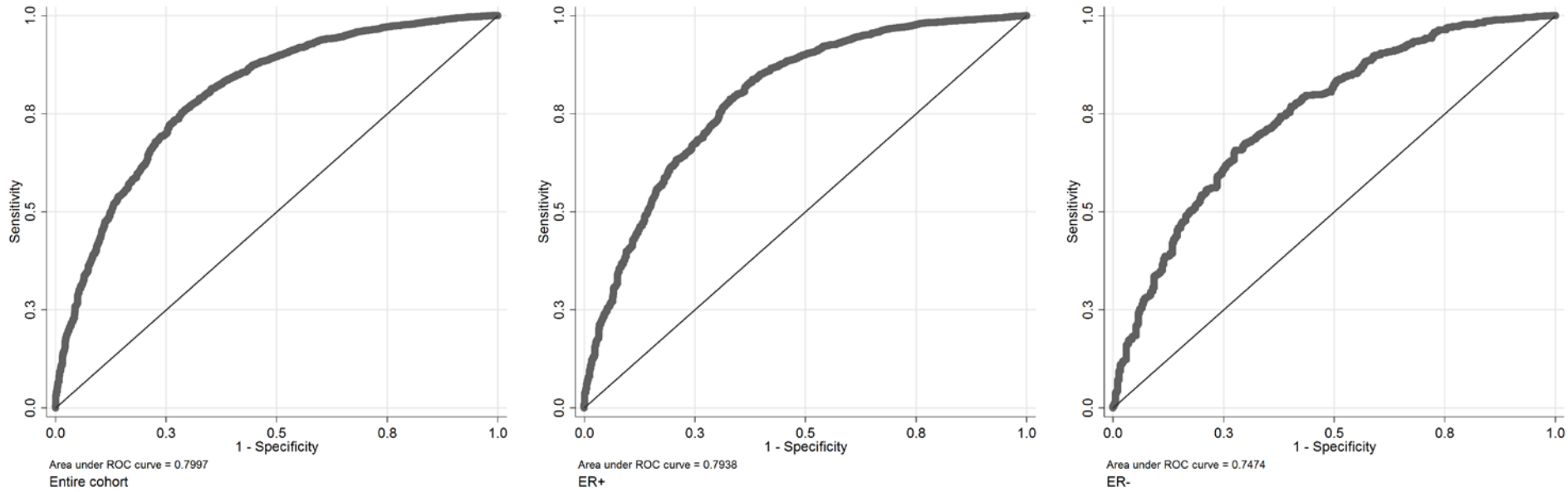

Figure 1. Discriminatory accuracy of 5-year overall survival for the entire cohort, ER+ patients and ER- patients. Abbreviations: ROC = receiver operating characteristic curve, ER = oestrogen receptor

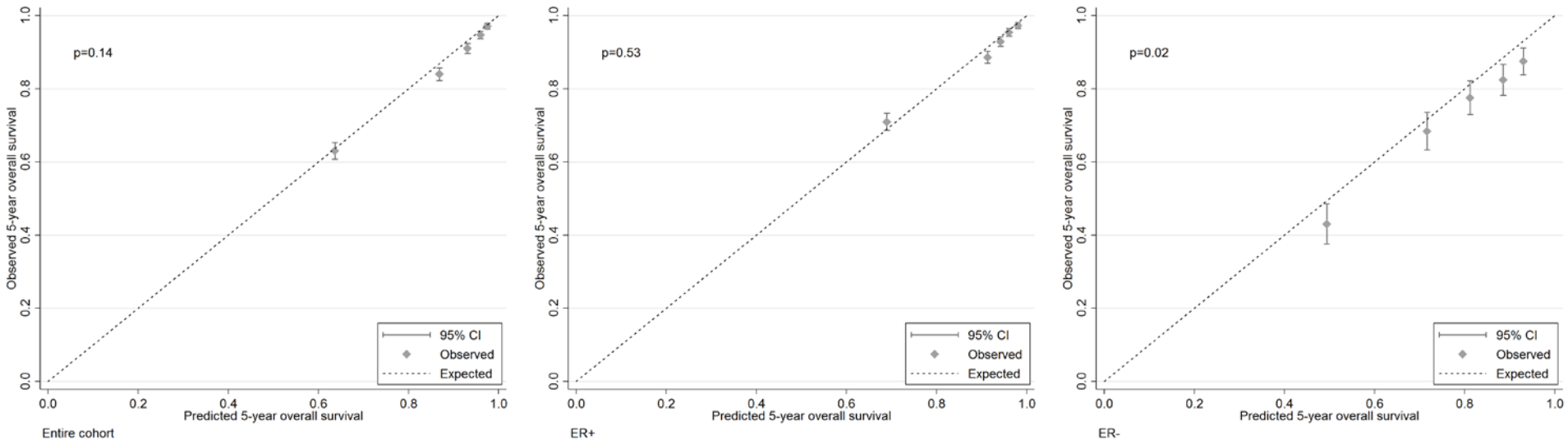

Figure 2. Observed and predicted 5-year overall survival for the entire cohort, ER+ patients and ER- patients. Abbreviations: CI = confidence interval, ER = oestrogen receptor

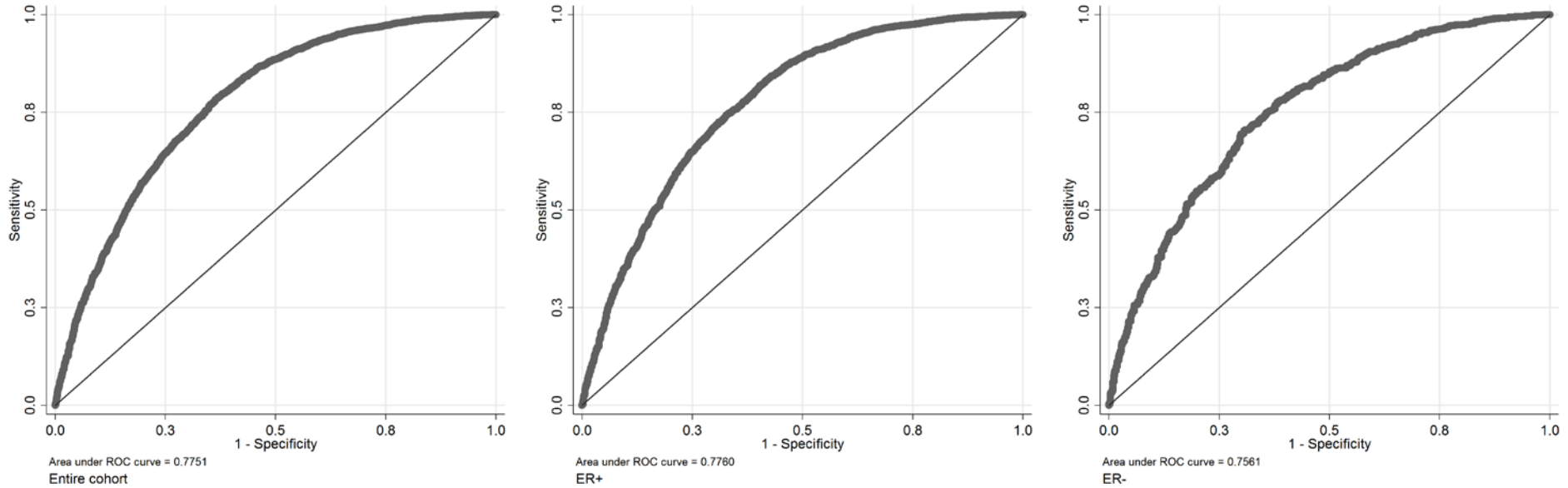

Figure 3. Discriminatory accuracy of 10-year overall survival for the entire cohort, ER+ patients and ER- patients. Abbreviations: ROC = receiver operating characteristic curve, ER = oestrogen receptor

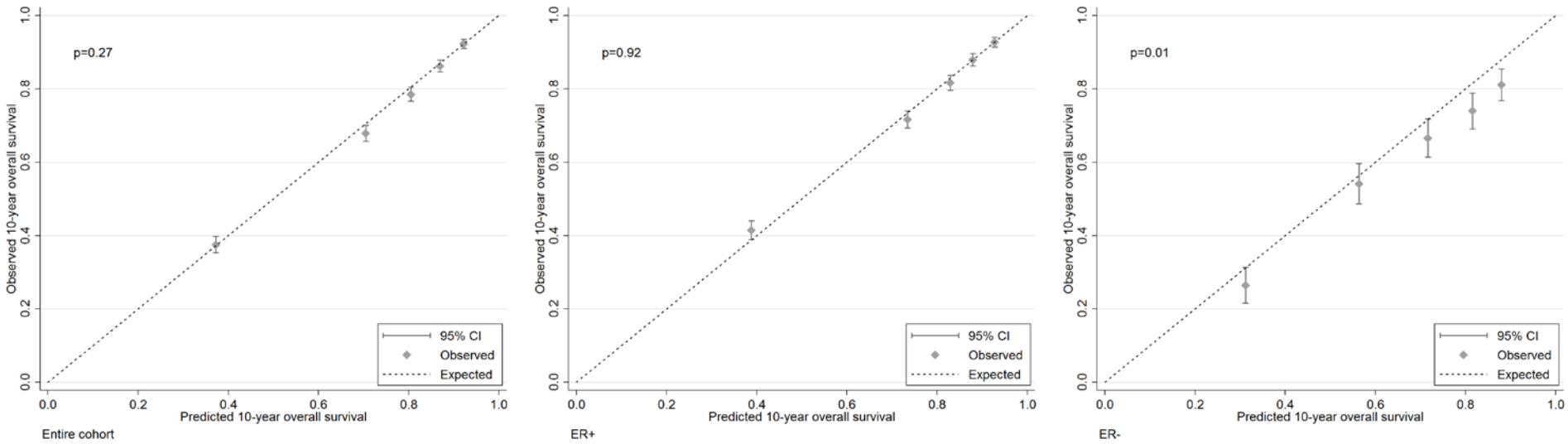

Figure 4. Observed and predicted 10-year overall survival for the entire cohort, ER+ patients and ER- patients. Abbreviations: CI = confidence interval, ER = oestrogen receptor