



Direct comparison of three different mathematical models in two independent datasets of EUSOMA certified centers to predict recurrence and survival in estrogen receptor-positive breast cancer: impact on clinical practice

Cristiana Iacuzzo¹ · Fabiola Giudici^{2,3} · Serena Scomersi⁴ · Rita Ceccherini⁴ · Fabrizio Zanconati⁴ · Daniele Generali^{5,6} · Marina Bortul⁴

Received: 29 December 2020 / Accepted: 8 February 2021 / Published online: 1 March 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Purpose Prediction algorithms estimating survival rates for breast cancer (BC) based upon risk factors and treatment could give a help to the clinicians during multidisciplinary meetings. The aim of this study was to evaluate accuracy and clinical utility of three different scores: the Clinical Treatment Score Post-5 Years (CTS5), the PREDICT Score, and the Nottingham Prognostic Index (NPI).

Methods This is a retrospective cohort analysis conducted on prospectively recorded databases of two EUSOMA certified centers (Breast Unit of Trieste Academic Hospital and of Cremona Hospital, Italy). We included patients with Luminal BC undergone to breast surgery between 2010 and 2015, and subsequent endocrine therapy for 5 years for curative intent.

Results A total of 473 patients were enrolled in this study. ROC analysis showed fair accuracy for NPI, good accuracy for PREDICT, and optimal accuracy for CTS5 (AUC 0.7, 0.76, and 0.83, respectively). The three scores seemed strongly correlated in Spearman's rank correlation coefficient analysis. PREDICT partially overestimated OS in patients undergone to mastectomy, and in pT3-4, G3 tumors. Considering DRFS as a surrogate of OS, CTS5 showed women in intermediate and high risk class had shorter OS too (respectively $p=0.001$ and $p<0.001$). Combining scores does not improve prognostication power.

Conclusion Mathematical models may help clinicians in decision making (adjuvant therapies, CDK4/6i, genomic test's gray zones). CTS5 has the higher prognostic accuracy in predicting recurrence, while score predicting OS did not show substantial advances, proving that pN, G, and pT are still the most important variables in predicting OS.

Keywords Breast cancer · Mathematical models · Scores · Recurrence · Overall survival

✉ Cristiana Iacuzzo
cricsi@hotmail.it

¹ Department of General Surgery, Academic Hospital of Trieste, Trieste University Hospital, Strada di Fiume 447, 34149 Trieste, Italy

² Department of Medicine, Surgery and Health Sciences, University of Trieste, Biostatistics Unit, Trieste, Italy

³ Epidemiology and Public Health Department, University of Padua, Biostatistics Unit, Padua, Italy

⁴ Breast Unit, Academic Hospital of Trieste, Trieste, Italy

⁵ Breast Cancer Unit, Azienda Socio Sanitaria Territoriale di Cremona, Cremona, Italy

⁶ Department of Medical, Surgery and Health Sciences, University of Trieste, Trieste, Italy

Introduction

Over the past 10–15 years, treatment concepts of breast cancer (BC) have evolved to take the wide heterogeneity of this disease and its development into account. Emphasis has been especially placed on biologically tailored therapies and treatment de-escalation, in order to reduce adverse effects of pharmacologic therapies, radiotherapy, and invasiveness of surgical procedures.

Despite the inherent molecular heterogeneity, which is a driving principle of modern-day treatments, some features, such as the impact of loco-regional tumor burden or metastatic patterns, are not yet completely understood and subsequently influence therapy effectiveness and survival.

Actually the two major pillars of BC management are loco-regional treatment and systemic therapy, and the histological and molecular characteristics of BC largely influence treatment decisions.

All patients with estrogen receptor-positive (ER+) BC, independently of HER2 status, should receive endocrine therapy to block ER activity. The main questions in ER+/HER2-negative early BCs are which patients need chemotherapy in addition to endocrine therapy and which patients may benefit from extending endocrine therapy beyond 5 years [1]. Extended adjuvant endocrine therapy for up to 10 years, or even 15 years, impacts favorably on patient outcomes [2]. Nevertheless, the decision for such an approach needs to take relapse risk and tolerability into account, considering that extended adjuvant endocrine therapy is particularly beneficial for patients at high risk for relapse [1].

Despite therapeutic improvement, almost 30% of stage I and II BCs experience recurrence during follow-up. Decision-making process following surgery for early BC should be essentially built on an accurate prediction of survival. These decisions for the most part are still largely based on well-known pathological prognostic factors that are known to be independent on multivariate analysis—i.e., tumor size, grading, and lymph node status [3–5]

In the last decades, several scores and mathematical models have been created by incorporation of several variables and the final construction of a continuous function for calculating the probability of disease-specific death. By now, readily online tools, which rely on standard parameters gathered routinely in clinical practice and may be able to predict which women with ER+ BC have a low risk or high risk of a late distant recurrence (LDR), are available and they may help the clinicians in their therapeutic decision making [6, 7].

The aim of our study was to evaluate, with a direct comparison, the accuracy along their impact and the clinical utility of three different scores as the Clinical Treatment Score Post-5 Years (CTS5), the PREDICT, and the Nottingham Prognostic Index (NPI).

Materials and methods

This is a retrospective cohort analysis conducted on two prospectively collected independent databases of two EUSOMA certified centers: the Breast Unit of Trieste Academic Hospital and the Breast Unit of Cremona Hospital, Italy. We included in our analysis only female patients with ER+HER2-negative BCs undergone to breast surgery from January 2010 to December 2015, and subsequently submitted to adjuvant endocrine therapy for 5 years for curative intent. We excluded from the analysis all the patients undergone to neoadjuvant therapy, to adjuvant chemotherapy,

and patients with prior history of BC (invasive and non-invasive) at diagnosis. We followed up all the patients for at least 5 years, or until death.

Statistical analysis

Preoperative factors were recorded, including age, pre- or post-menopausal state, and method of BC detection (screening program versus symptoms). Intraoperative details included type of breast surgery (conservative surgery versus mastectomy), type of axillary surgery (sentinel lymph node biopsy only, contemporary axillary dissection, or delayed axillary dissection), and number of excised lymph nodes. Pathological details of the BC were recorded: pT, pN, number of positive lymph nodes, histology, grading, presence of DCIS, receptor profile, and Ki67. Postoperative management was recorded, in particular the type of endocrine adjuvant therapy administered (Tamoxifen versus aromatase inhibitors), and any postoperative radiotherapy performed on mammary bed and/or nodal stations. For what concerns follow-up data, we recorded the status of the patient at follow-up (alive and free of disease, alive but metastatic, and dead), any local, loco-regional or distant recurrence occurred during the follow-up period, and timing of any recurrence.

Categorical variables were expressed by frequencies and percentage, while continuous variables were expressed by mean (standard deviation, SD) or median [range (min–max)], as appropriate. Women were stratified into different risk groups based on their calculated CTS5, NPI, and PREDICT scores, which predict their risk of LDR for CTS5 and of death for NPI and PREDICT. Although the CTS5 algorithm was developed to predict the risk of LDR, in our study we purposely used this tool to classify the risk of early metastases (from diagnosis) as a surrogate indicator of mortality, in order to compare this score with NPI and PREDICT that are used to determine 5-year overall survival (OS) after BC surgery.

The demographical and clinical-pathological variables were compared among different risk groups by Kruskal–Wallis test or one-way Anova for continuous data (according to data distribution tested with Shapiro–Wilk normality test) and independent Chi-square test for categorical data. Patients OS was defined as the time (years) from date of surgery to either death or last observation, while distant recurrence-free survival (DRFS) as the time from surgery to either metastases or last follow-up. OS and DRFS were described using the Kaplan–Meier approach. The observed survival differences between the prognosis groups of CTS5 and NPI were analyzed using log-rank test. The univariate Cox regression was used to estimate prognostic scores associated with OS or DRFS, after the assumption of the proportional hazard was verified. The proportional hazard assumption was tested using the Schoenfeld residual

test. Hazard ratios (HRs) were reported with 95% confidence intervals (CIs).

Discrimination of the scores (CTS5, NPI, and PREDICT) as prognostic indices was assessed by calculating the area under the receiver operator curve (ROC) (AUC) with 95% confidence interval (95%CI) for 5-year distant recurrence and 5-year mortality. The area under the ROC curve (AUC) gives an indication of the discriminatory performance of the model, whereby it can be interpreted as the proportion of patients who are correctly predicted to be alive or dead (distant recurrence free or not) at 5 years. An AUC of 0.5 indicates no discriminative performance, whereas an AUC of 1.0 indicates perfect discrimination. To assess the calibration of the PREDICT model, the observed and predicted 5-year OS rates were compared. For evaluating the correlation between the values predicted by the three scores, Spearman's correlation coefficient and Lin's concordance correlation coefficient were calculated. All the statistical tests were two sided, and statistical significance was defined as p value < 0.05 .

Results

Two independent datasets with a total of 473 patients treated in the Breast Unit of Trieste and in the Breast Unit of Cremona have been considered in this study. Demographics, intraoperative details, histologic characteristics, and postoperative strategy are described in Table 1.

Median follow-up was 5.7(4.55–8.54) years. As displayed in Table 2, only 2.7% of the cohort experienced loco or loco-regional recurrence, 5.9% distant metastases. At the end of the period of analysis, 14.6% of the patients were dead after BC or other causes. 5-year distant recurrence-free survival (DRFS) and overall survival (OS) were, respectively, 95.2% (92.8–96.8%) and 88% (84.7–90.6%). Twenty-two patients experienced distant recurrence during the first 5 years of follow-up, in particular during the first 2 years ($N = 10$), only 6 patients after the 5th year, but we must consider that a limited number of subjects in our cohort had a longer follow-up. Among the 68 deaths, 55 (79.7%) happened during the 5-year postoperative period, 14 after the 5th year, and this number neither is negligible.

The Nottingham prognostic index

We stratified patients into three classes of risk—a good prognosis group ($N = 296$, 62.58%), a moderate prognosis group ($N = 138$, 29.18%), and a poor prognosis group ($N = 39$, 8.25%)—and into five classes of risk—very good prognosis ($N = 53$, 11.21%), good prognosis group ($N = 243$, 51.37%), moderate prognosis I ($N = 98$, 20.72%), moderate

Table 1 Patients' demographics, intraoperative details, histologic characteristics, and postoperative strategy

| Variables | <i>N</i> | % |
|----------------------------------|----------|---------|
| Age | | |
| Mean (SD) | 68.5 | (11.6) |
| Median (range) | 70 | (33–92) |
| Menopausal state | | |
| Premenopausal | 54 | 11.4% |
| Postmenopausal | 419 | 88.6% |
| Method of detection | | |
| Screening | 164 | 34.7% |
| Symptoms | 309 | 65.3% |
| Breast surgery | | |
| Conservative surgery | 314 | 66.4% |
| Mastectomy | 159 | 33.6% |
| Axillary surgery | | |
| SLN biopsy only | 350 | 74.0% |
| SNL biopsy + axillary dissection | 71 | 15.0% |
| Axillary dissection ab initio | 52 | 11.0% |
| pT | | |
| pT1a-b | 155 | 32.8% |
| pT1c | 192 | 40.6% |
| pT2 | 102 | 21.6% |
| pT3-4 | 22 | 4.7% |
| pN | | |
| pN0 | 357 | 75.5% |
| pN1mi | 21 | 4.4% |
| pN1 | 71 | 15.0% |
| pN2 | 11 | 2.3% |
| pN3 | 13 | 2.7% |
| Positive lymph nodes | | |
| 0 | 356 | 75.3% |
| 1 | 68 | 14.4% |
| 2–3 | 24 | 5.1% |
| 4–9 | 12 | 2.5% |
| > = 10 | 13 | 2.7% |
| Grading | | |
| G1 | 67 | 14.2% |
| G2 | 356 | 75.3% |
| G3 | 50 | 10.6% |
| Ki-67 | | |
| < 20% | 336 | 71.0% |
| > = 20% | 137 | 29.0% |
| Radiotherapy | | |
| Yes | 319 | 67.4% |
| No | 154 | 32.6% |
| Endocrine therapy | | |
| AI | 370 | 78.2% |
| Tamoxifen | 81 | 17.1% |
| Tamoxifen + AI | 22 | 4.7% |

Table 2 Follow-up details

| Variables | N | % |
|----------------------------------|------|-------------|
| Follow-up (years) | | |
| Median (range) | 5.70 | (4.55–8.54) |
| Loco and locoregional recurrence | | |
| No | 460 | 97.3% |
| Yes | 13 | 2.7% |
| Distant recurrence | | |
| No | 445 | 94.1% |
| Yes | 28 | 5.9% |
| Death | | |
| No | 404 | 85.4% |
| Yes | 69 | 14.6% |

prognosis II ($N=40$, 8.46%), and poor prognosis ($N=39$, 8.25%)—to evaluate if accuracy improved.

As expected, mortality rate is higher in poor prognosis group, especially mortality after the 5th year of follow-up (Fig. 1). Patients of poor and moderate prognosis groups showed a significantly higher risk of death compared to women of the good prognosis group: 5-year OS is, respectively, 66.6% (49.5–79.0%), 83.2% (75.8–88.5%), and 93.1% (89.5–95.5%); the hazard ratio (95%CI) for the moderate prognosis group is 2.64 (1.54–4.50, $p < 0.001$) and for the poor prognosis group is 5.22 (2.75–9.91, $p < 0.001$). However, our analysis showed that expected 5-year OS overestimates observed OS in good prognosis group and in

particular in moderate prognosis group (6% difference). ROC analysis showed that this model has a fair accuracy in predicting OS, AUC = 0.70, 95% CI 0.63–76 (Fig. 2).

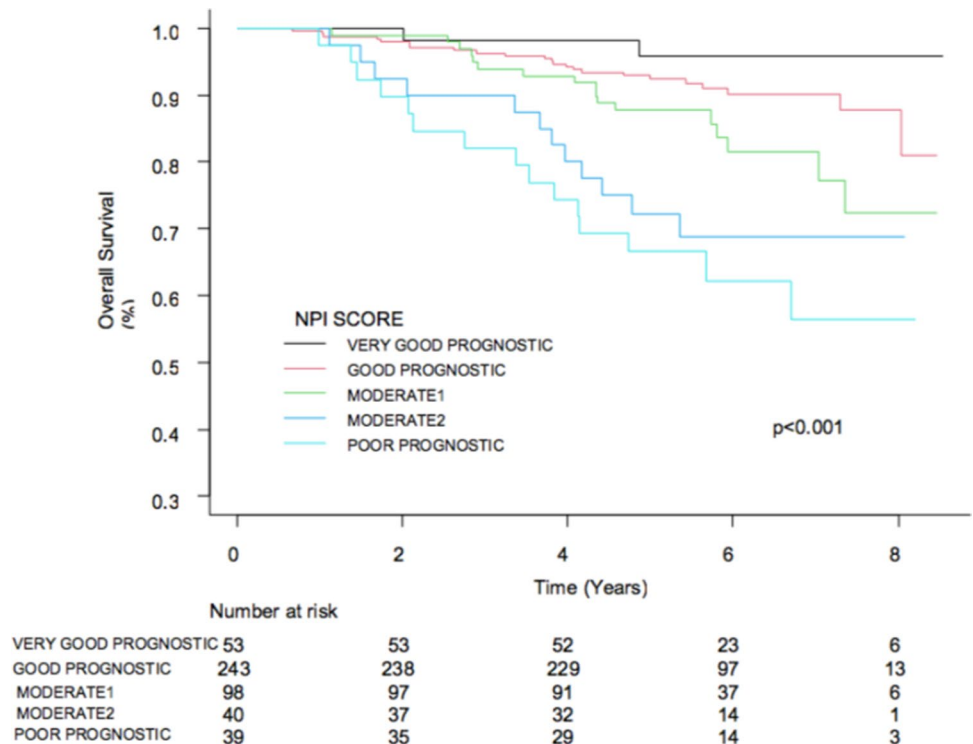
The PREDICT score

We stratified patients into 3 classes of risk: a low risk group characterized by OS < 79% ($N=222$ patients of our cohort, 46.9%), an intermediate risk group with OS 79–90% ($N=142$, 30%), and a high risk group with OS > 90% ($N=109$, 23%). Figure 3 shows expected OS for each patient. Globally, median 5-year OS predicted by the score results similar to the observed OS of our cohort (90% versus 88%). We compared expected OS with observed OS of our dataset, and we found out that PREDICT tends to overestimate OS in patients undergone to mastectomy, pT3-4 tumors, G3 tumors, and patients not submitted to radiotherapy, but to underestimate it in pN2-3 tumors (Table 3). ROC analysis showed that PREDICT discriminates alive patients of our cohort from dead patients with fair accuracy, AUC = 0.76, 95%CI 0.70–81 (Fig. 4).

The CTS5 score

According to CTS5, patients were stratified into 3 classes of risk: a low risk group (CTS5 < 3.13) with a < 5% possibility of developing distant recurrence ($N=269$, 56.9% of our cohort), an intermediate risk group (CTS5 3.13–3.86) with a 5–10% risk of developing a distant recurrence ($N=103$,

Fig. 1 OS per NPI 5 categories of risk



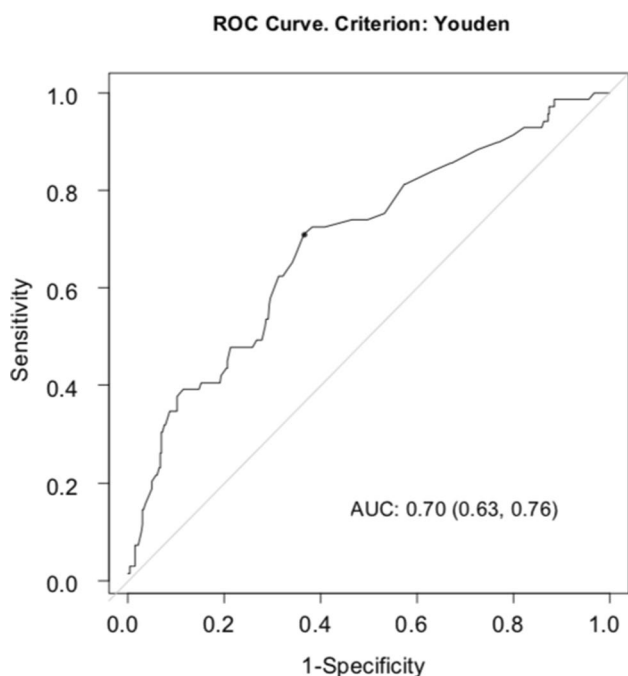
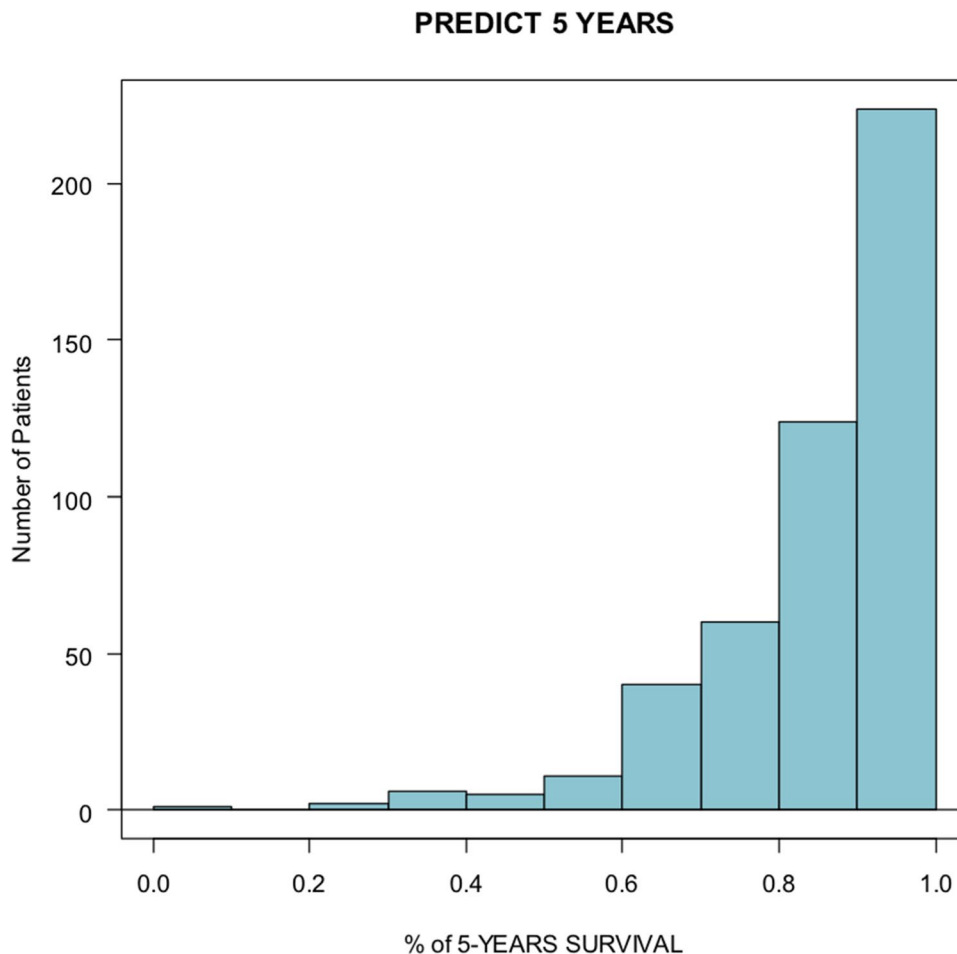


Fig. 2 NPI ROC curve

Fig. 3 PREDICT expected 5-year OS



21.8% of our cohort), and a high risk class (CTS5 > 3.86) with a > 10% risk of distant recurrence during follow-up (N = 101, 21.3% of our cohort). Median CTS5 was 3.01 (range: 1.47–6.05). As expected distant recurrence rate was higher in high risk class, and CTS5 correctly predicted the recurrence within the 5th year of follow-up in 86.4% of the women who experienced a relapse, classifying them as high risk (Fig. 5). Low risk and intermediate risk group did not show significant differences in DRFS ($p = 0.35$), unlike high risk group, $p < 0.001$. ROC analysis showed that this model discriminates with good accuracy distant relapsing patients from not-relapsing ones (AUC = 0.81, 95% CI 0.71–0.91, Fig. 6). Since DRFS can be considered a surrogate of OS, or at least the two are strictly related, we evaluated OS per CTS5 classes of risk. The patients included in intermediate and high risk classes showed a significantly shorter OS (respectively, $p = 0.001$ and $p < 0.001$).

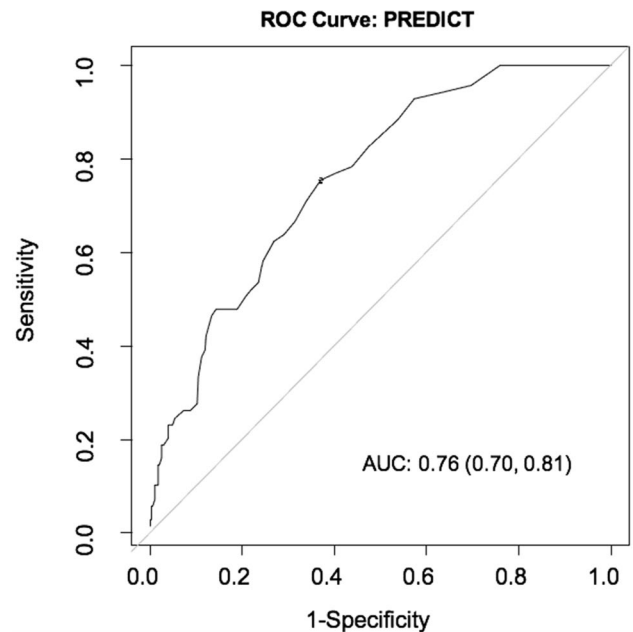
Correlations between the NPI, CTS5, and PREDICT scores

Spearman’s rank correlation coefficient showed a strong correlation between NPI and CTS5 scores (Fig. 7). The

Table 3 PREDICT expected OS versus observed OS

| Variables | Observed OS (95%CI) | PREDICT Median OS | Difference |
|----------------------|---------------------|-------------------|------------|
| Total | 88.0 (0.847–0.906) | 90.0 | 2.0 |
| Age | | | |
| < 60 | 99.0 (0.931–0.999) | 97.0 | 2.0 |
| > =60 | 85.1 (0.810–0.883) | 85.0 | –0.1 |
| Method of detection | | | |
| Screening | 95.7 (0.911–0.979) | 93.0 | –2.7 |
| Symptomatic | 84.0 (0.794–0.876) | 85.0 | 1.0 |
| Breast Surgery | | | |
| Conservative surgery | 90.7 (0.868–0.934) | 90.5 | –0.2 |
| Mastectomy | 82.7 (0.758–0.878) | 88.0 | 5.3 |
| pT | | | |
| pT1a-b | 96.1 (0.915–0.982) | 94.0 | –2.1 |
| pT1c | 91.0 (0.859–0.943) | 91.0 | 0.0 |
| pT2 | 78.1 (0.687–0.850) | 81.0 | 2.9 |
| pT3-4 | 50.0 (0.282–0.684) | 61.0 | 9.0 |
| pN | | | |
| N0-N1mi | 90.6 (0.871–0.932) | 91.0 | 0.4 |
| N1 | 78.5 (0.668–0.864) | 83.0 | 4.5 |
| N2-3 | 75.0 (0.526–0.879) | 61.0 | –14.0 |
| Ki-67 | | | |
| < 10% | 91.1 (0.864–0.942) | 91.0 | –0.1 |
| > = 10% | 85.3 (0.803–0.892) | 87.0 | 1.7 |
| Grading | | | |
| G1 | 92.3 (0.824–0.967) | 93.0 | 0.7 |
| G2 | 90.0 (0.864–0.928) | 90.0 | 0.0 |
| G3 | 67.9 (0.530–0.789) | 74.5 | 6.6 |
| Endocrine therapy | | | |
| AI | 86.6 (0.827–0.897) | 88.0 | 1.4 |
| Tamoxifen | 92.4 (0.838–0.965) | 97.0 | 4.6 |
| Tamoxifen + AI | 95.5 (0.719–0.993) | 92.0 | –3.5 |
| Radiotherapy | | | |
| Yes | 89.9 (0.860–0.927) | 95.0 | 5.1 |
| No | 84.1 (0.772–0.891) | 93.4 | 9.3 |

two scores perfectly agree in identifying the high risk class of patients ($N=39$). Even if NPI seems slightly more optimistic in the prediction of OS than CTS5 is in the prediction of recurrence ($N=44$ patients of the good prognosis NPI class are at intermediate risk for recurrence), there is not a complete disagreement in any of the patients, meaning that none of the patients are considered at high risk in one score but low risk in the other. Only 17 patients considered at low risk in CTS5 fell in the moderate prognosis group.

**Fig. 4** PREDICT ROC curve

Spearman's rank correlation coefficient showed a strong correlation between the two scores evaluating OS – NPI and PREDICT (Fig. 8). There is a significant difference in PREDICT prognosticated OS of the five classes from NPI ($p < 0.001$).

Spearman's rank correlation coefficient showed a good correlation even between PREDICT and CTS5 scores (Fig. 9). OS showed a significant difference among the three classes from CTS5 ($p < 0.001$).

The combination of CTS5 and NPI score

We finally used CTS5 and NPI score to identify a very high risk class, defined as patients with poor prognosis for NPI score and high risk for CTS5 score. We decided not to use the PREDICT score since it does not work with absolute numbers but rates. Instead, this new score (NPI+CTS5) could be easily obtained by the addition of the points of the two scores. This group of patients includes only 39 individuals of our cohort. As expected the score of the patients of the very high risk group is significantly higher than the rest of the cohort ($p < 0.001$). As expected OS and DRFS showed a significant difference among the very high risk and the low–intermediate risk groups ($p < 0.01$), but the addition of the two scores does not significantly increase the accuracy of the scores as separate

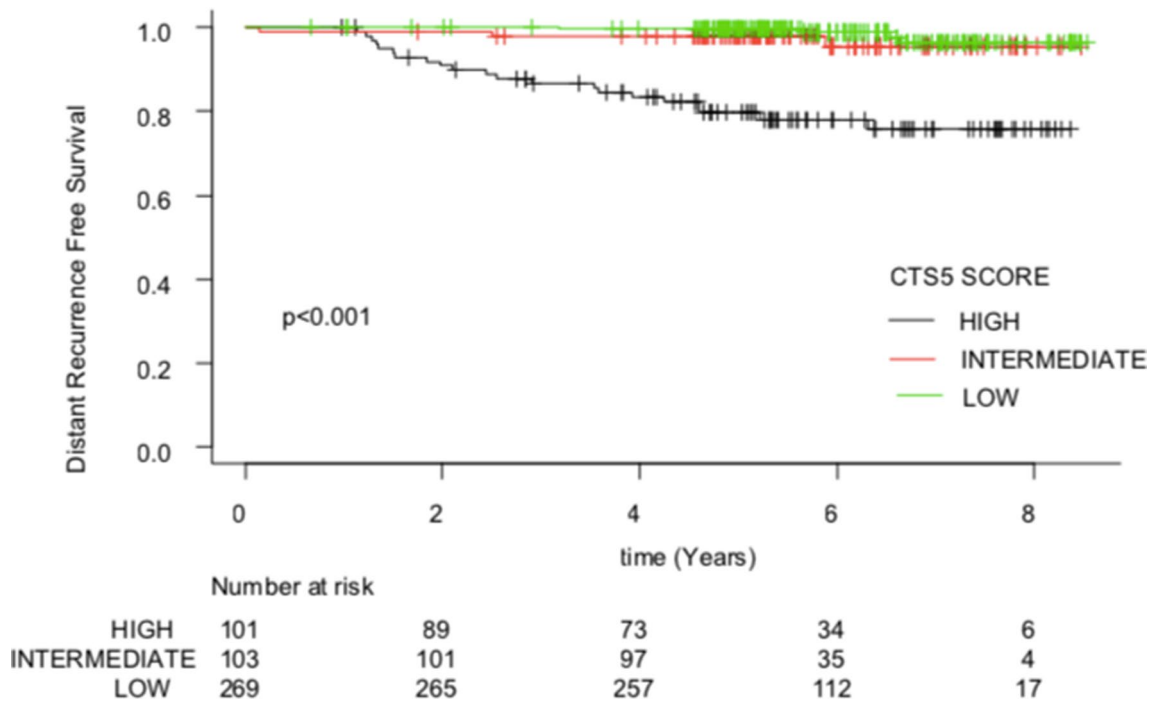


Fig. 5 DRFS per CTS5 classes of risk

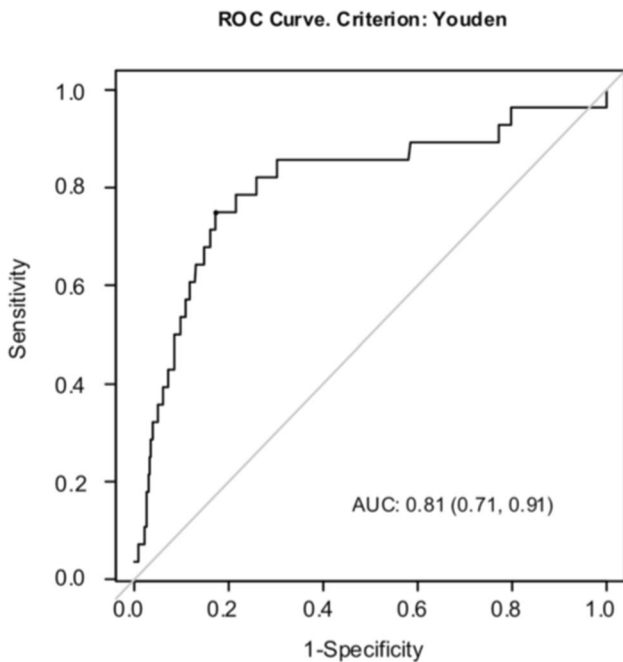


Fig. 6 CTS5 ROC curve

entities, as shown in ROC analysis in Fig. 10 (AUC 0.72 versus AUC 0.7 of NPI only and 0.81 of CTS5 only).

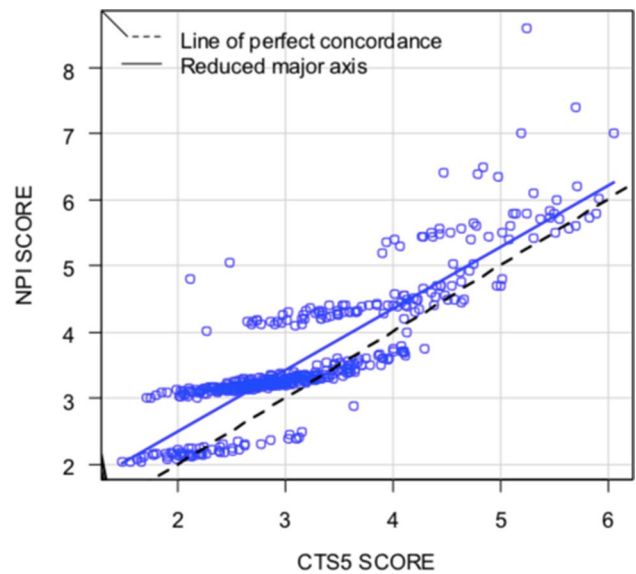
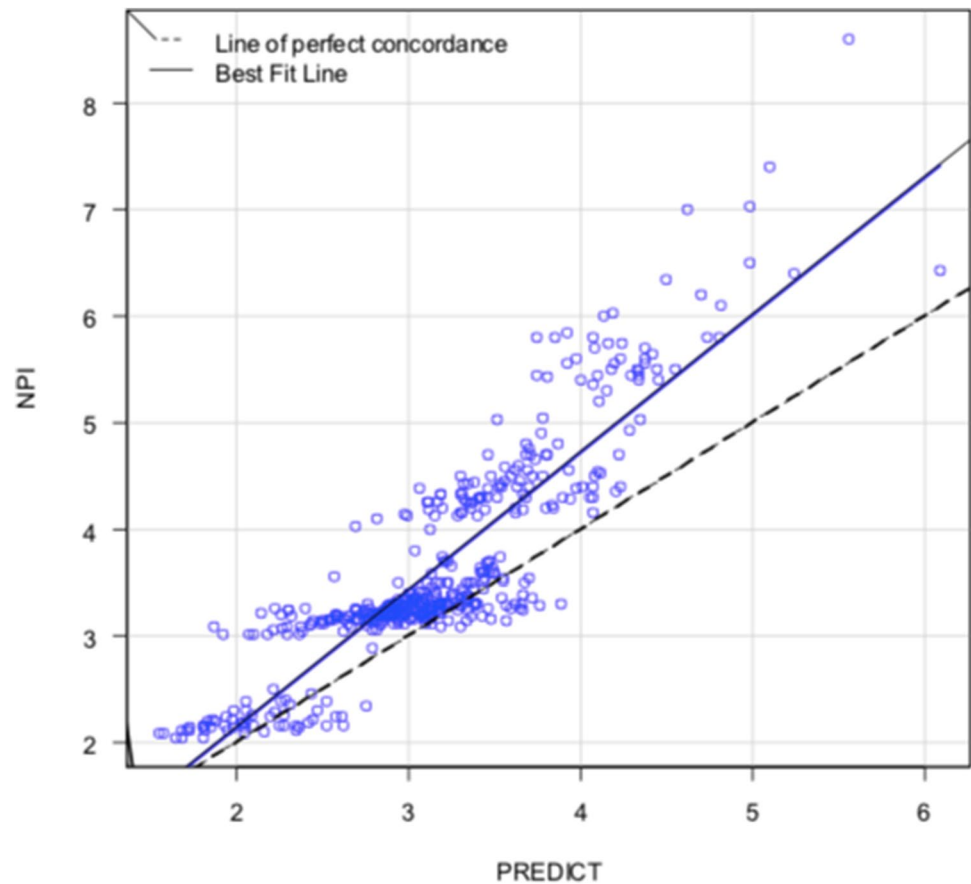


Fig. 7 Correlation between NPI and CTS5

Discussion

Before the era of genomic profile tests, the importance of validated prognostic models based on standard factors was stressed, in order to help the creation of standardized analyses and avoid claims concerning the relevance of new

Fig. 8 Correlation between NPI and PREDICT



potential prognostic factors and to homogenize clinicians' approach to adjuvant therapy [8]. Despite the unstoppable progress in pathological BC evaluation, nowadays there is still controversial discussion concerning biological variables, such as Ki-67 along the widely and routine use of genomic platform, so the necessity of a solid prognostic model based on standard factors has increased, to support clinical practice.

Nodal status is a powerful prognostic marker especially for late distant recurrence, whereas tumor size and grading are shown to be less prognostic after 5 years. Recently, an analysis of 60,000 women with ER + BCs, who were scheduled to receive a 5-year endocrine therapy regimen and remained disease free until the end of the treatment, reported a subsequent risk of late distant recurrence. Even in patients with T1N0 disease, the estimated risk between years 5 and 20 is 10% for those with low histologic grade, 13% for those with intermediate, and 17% for those with high. Although endocrine treatment markedly reduces mortality (by approximately 30% with 5 years of Tamoxifen and approximately 40% with an aromatase inhibitors), recurrences continue to occur after the 5-year treatment ends. The observation that these events can be decreased by continued treatment means that decisions about whether

to continue with therapy at year 5 are at the forefront of patient management at that time [9–14].

It is blatantly clear by now that NPI simple variables single handed are not sufficient to predict and explain the development of a large portion of early BCs [15]

The PREDICT is a prognostication model for early BC created basing on data collected from a large number of patients within a single UK cancer registry (Eastern Cancer Registration and Information Centre, ECRIC), tested, and validated on an external dataset of more than 5000 patients (West Midlands Cancer Intelligence Unit, WMCIU). Information obtained regarded age at diagnosis, number of positive sampled lymph nodes, tumor size, grading, ER status, modality of detection, information on local therapy, and adjuvant therapy. BC-specific mortality and competing mortality (mortality from other causes) were modeled separately and adjusted for the age at the diagnosis. Several other factors were gradually added to this score, in order to refine its prognostication power – i.e., DICS or LCIS only, menopausal state, HER2 status, and Ki-67. The final results are deemed to predict OS at 5, 10, and 15 years after surgery, and the additional benefit of endocrine therapy, chemotherapy, bisphosphonates, and anti-HER2 treatment [16]

Fig. 9 Correlation between PREDICT and CTS5

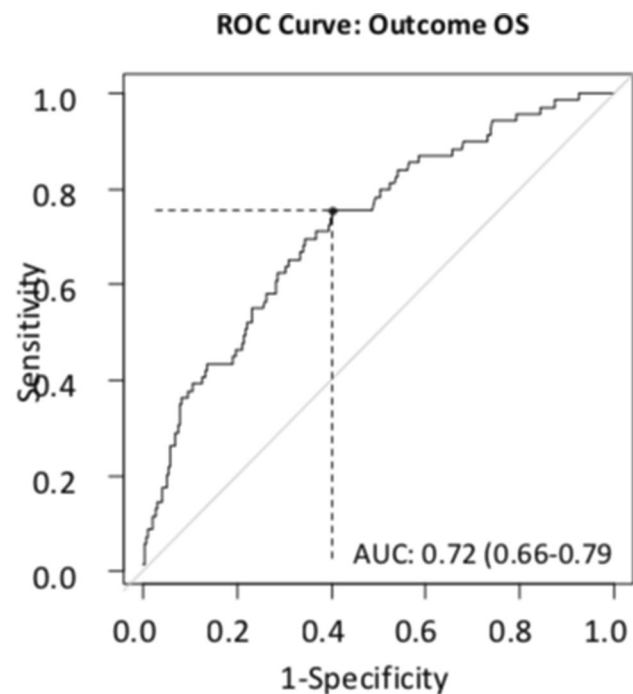
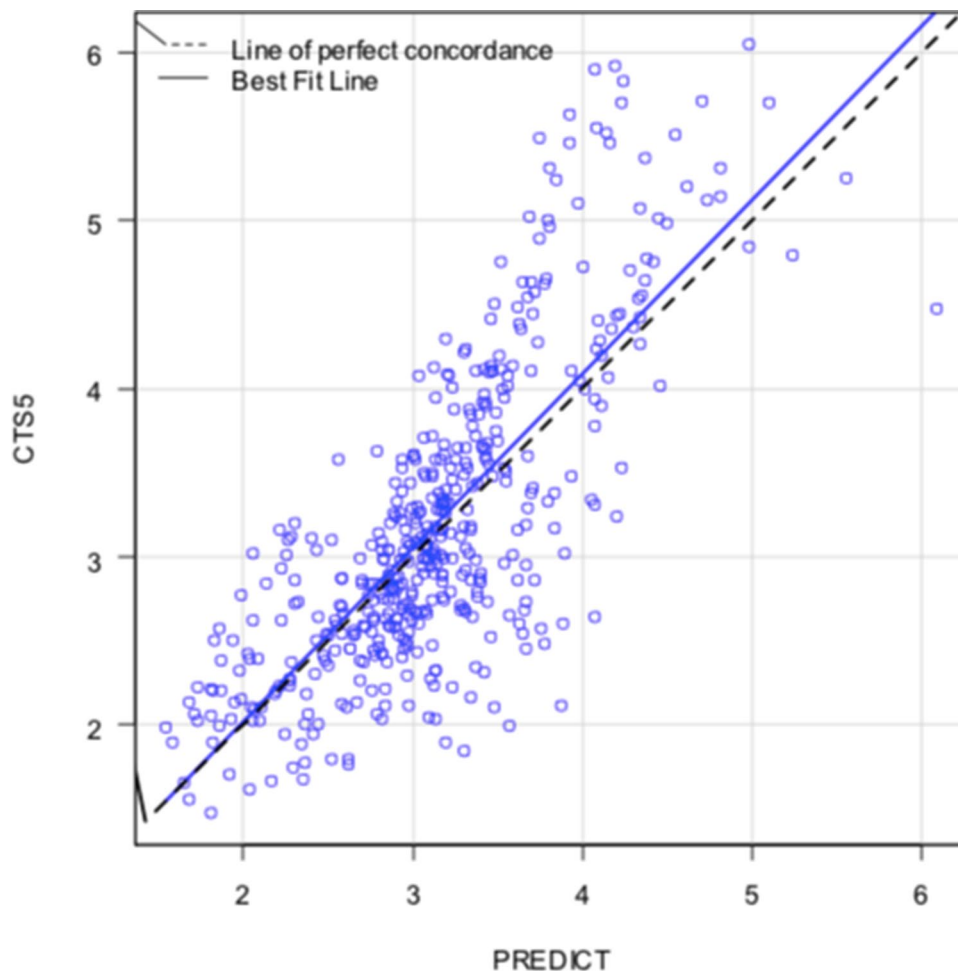


Fig. 10 NPI+CTS5 ROC curve

The CTS5 is a score aiming to estimate the risk of LDR (after 5 years) basing on clinical-pathological parameters potentially measurable in all patients with BC at diagnosis. The variables initially analyzed by univariable Cox regression were as follows: nodal status, tumor size, grading, age at the start of endocrine treatment, and type of endocrine therapy. CTS5 was deemed to define 3 different risk populations – low, intermediate, and high – between years 5 and 10 after surgery [9, 17]. Distant recurrence can be considered as a transient intermediate state from initial diagnosis of BC to death. The occurrence of distant recurrence indicates progression and consequently an increased risk of mortality; we purposely used the CTS5 score to predict the risk of metastasis from diagnosis as a surrogate indicator of mortality. We used this score to predict the risk of metastasis from diagnosis, and the head to head comparison between the considered scores has been performed to estimate the risk of developing early metastases with regard to the risk of death.

Our analysis showed the prognostic models we considered – NPI, PREDICT, and CTS5 – have fair accuracy in predicting clinical course of ER + /HER2- BC undergone to surgery for curative intent. Our results mostly agree with

what reported on the literature about both NPI and PREDICT [18]. In our cohort CTS5 seems to be the score that better identifies the group of patients at high risk of distant recurrence, and, if the DRFS is accepted as a surrogate marker of OS, with evident repercussions on survival.

Analyzing the subgroup of patients who were recurrence free 5 years after diagnosis ($n=451$), 6 of them developed LDR: Three of them were classified at high risk and three at low risk according to CTS5 score. Interestingly, among all the studied variables, only high ki-67 values were associated with LDR for patients with low CTS5 score ($p=0.03$), and this association was not found in intermediate–high groups. Although this result is probably influenced by the small sample size of this subgroup of patients, it is aligned to what shown in a recent report suggesting extended endocrine therapy also for patients with high Ki-67 ($>20\%$) in low CTS5 group [19].

If we specifically consider the two models predicting OS – NPI and PREDICT – it seems clear that adding to the model many variables does not significantly improve the accuracy of the score (AUC improves from 0.7 of NPI to 0.76 of PREDICT). It is conceivable that at the current state-of-the-art nodal status, tumor dimension and tumor grading still represent the three heaviest factors in BC history prognostication.

Even if our correlation analyses demonstrated that all three models agree in the identification of high risk classes of patients, the identification of a very high risk class, created by the addition of NPI and CTS5 scores points, does not significantly improve prognostic accuracy.

It is worth mentioning that the considered population had a mean age of 68.5 years and was mostly post-menopausal (88.6%), data that are in accordance with the national demographics. Thus, considering the global population is aged 60 years or a little bit over, the findings could be generalizable. We also performed a sensitivity analysis excluding all pre-menopausal women from the entire analyzed population and we observed very similar results in terms of prognostic scores accuracy between the pre- and post-menopausal setting (AUC, 95%CI for NPI, PREDICT, and CTS5 were, respectively, 0.68 (0.66–0.75), 0.70 (0.63–0.77), and 0.80 (0.71–0.92)). Therefore, the sensitivity analysis revealed a consistent result when excluding the pre-menopausal subgroup showing NPI, PREDICT, and CTS5 are useful prognostic scores among post-menopausal women.

All this considered, the CTS5 and the PREDICT may be daily used in clinical practice to give the clinician a general perspective on the risk of the patient. According to our data patients may be classified in three classes of risk: (1) a low risk class: patients in CTS5 low risk class and PREDICT low risk class; (2) a high risk class: patients at high risk in at least one of the two scores; and (3) a moderate risk class: patients in between.

This classification allows the clinician to stratify patients and, considering also age and comorbidities, to evaluate the use of adjuvant chemotherapy and/or 10-year endocrine therapy regimen in high risk class, 10-year endocrine therapy regimen in moderate risk, and 5-year endocrine therapy regimen in low risk class. These mathematical models – especially CTS5 that showed high accuracy in prognostication – may be used also in genomic tests gray zones (i.e., Oncotype DX RS of 26 to 30), to eventually tip the scale to a more aggressive or a more conservative management of adjuvant therapy.

In the upcoming future, the clinical use of endocrine therapy and CDK4/6i would be of routine practice in high-risk BC. Adding Abemaciclib to adjuvant endocrine therapy significantly reduces the risk of recurrence in patients with high risk, ER+, HER2 negative, early BC, as shown in phase-three trial findings [20]. In the absence of possibility to access to the genomic profile for the identification of individual BC patients' risk, the use of score models would be helpful in routine.

New studies are needed to identify other variables to create more accurate scores to help clinical decision making. The new frontier in prognostic model creation may be represented by machine learning, a branch of artificial intelligence enabling computer algorithms to learn from experience, without the need of being explicitly programmed. These techniques are cheaper than Multigene Signature Panels, and they are based on data that are already available in clinical routine [21].

Conclusion

Mathematical models may help clinicians in decision making in ER+ HER2-negative BC after surgery. However, our analysis showed the CTS5 seems to have the higher prognostic accuracy in predicting recurrence, while OS prediction did not show substantial advances of any of the three scores compared, proving that lymph node status, grading, and tumor burden are still the more important variables in predicting OS. New studies are needed to identify other variables to create more accurate scores.

Author contributions All the authors gave their own contribution to the design, statistical analysis, and writing of the manuscript.

Funding The authors certify that they have no affiliations or involvement in any organization or entity with financial interest (honoraria, educational grants, participation in speakers' bureaus, membership, employment, consultancies, stock ownership, or other equity interest, and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in

this manuscript. The manuscript has not and will not be a podium or poster meeting presentation.

Data availability The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors do not have interests to disclose.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. For this type of study, formal consent by the institutional research committee is not required in Italy.

Informed consent Informed consent was obtained from all the participants included.

References

- Harbeck N, Penault-Llorca F, Cortes J et al (2019) Breast cancer *Nat Rev Dis Primers* 5:66
- Gray R et al. Increasing the dose density of adjuvant chemotherapy by shortening intervals between courses or by sequential drug administration significantly reduces both disease recurrence and breast cancer mortality: an EBCTCG meta-analysis of 21,000 women in 16 randomised trials [abstract]. *SABCS GSI-GS01* (2018).
- Saphner T, Tormey DC, Gray R (1996) Annual hazard rates of recurrence for breast cancer after primary therapy. *J Clin Oncol* 14:2738–2746
- Goldhirsch A, Gelber RD, Price KN et al (1994) Effect of systemic adjuvant treatment on first sites of breast cancer relapse. *Lancet* 343:377–381
- Pagani O, Price KN, Gelber RD et al (2009) International Breast Cancer Study Group (IBCSG): Patterns of recurrence of early breast cancer according to estrogen receptor status: A therapeutic target for a quarter of a century. *Breast Cancer Res Treat* 117:319–324
- Kattan MW, Giri D, Panageas KS, Hummer A, Cranor M, Van Zee KJ, Hudis CA, Norton L, Borgen PI, Tan LK (2004) A tool for predicting breast carcinoma mortality in women who do not receive adjuvant therapy. *Cancer* 101:2509–2515
- D'Eredita G, Giardina C, Martellotta M et al (2001) Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with long-term follow-up that were treated in a single institution. *Eur J Cancer* 37:591–596
- Sauerbrei W, Royston P, Bojar H, Schmoor C, Schumacher M (1999) the German Breast Cancer Study Group (GBSG) Modeling the effects of standard prognostic factors in node positive breast cancer. *Brit J Cancer*. 79:1752–1760
- Dowsett M, Sestak I, Regan MM et al (2018) Integration of clinical variables for the prediction of late distant recurrence in patients with estrogen receptor-positive breast cancer treated with 5 years of endocrine therapy CTS5. *J Clin Oncol* 36(19):1941–1948. <https://doi.org/10.1200/JCO.2017.76.4258>
- Goss PE, Ingle JN, Pritchard KI et al (2016) Extending aromatase-inhibitor adjuvant therapy to 10 years. *N Engl J Med* 375:209–219
- Davies C, Pan H, Godwin J et al (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* 381:805–816
- Sestak I, Dowsett M, Zabaglo L et al (2013) Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 105:1504–1511
- Mittempergher L, Saghatchian M, Wolf DM et al (2013) A gene signature for late distant metastasis in breast cancer identifies a potential mechanism of late recurrences. *Mol Oncol* 7:987–999
- Pan H, Gray R, Braybrooke JP et al (2017) 20-year recurrence risks for patients with estrogen receptor-positive breast cancer after adjuvant endocrine treatment for only 5 years. *N Engl J Med* 377:1836–1846
- Wishart GC, Azzato EM, Greenberg DC et al (2010) PREDICT a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res.* 12(1):401. <https://doi.org/10.1186/bcr2464>
- Candido Dos Reis FJ et al (2017) An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res.* 19:58
- Regan MM, Neven P, Giobbie-Hurder A et al (2011) Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: The BIG 1–98 randomised clinical trial at 8.1 years median follow-up. *Lancet Oncol* 12:1101–1108
- Winzer KJ, Buchholz A, Schumacher M, Sauerbrei W (2016) Improving the Prognostic Ability through Better Use of Standard Clinical Data - The Nottingham Prognostic Index as an Example. *PLoS ONE* 11(3):e0149977. <https://doi.org/10.1371/journal.pone.0149977>
- Lee J, Cha C, Ahn SG et al (2020) Validation of Clinical Treatment Score post-5 years (CTS5) risk stratification in premenopausal breast cancer patients and Ki-67 labelling index. *Sci Rep* 10:16850
- Johnston SRD, Harbeck N, Hegg R, Toi M, Martin M, Shao ZM, Zhang QY, Martinez Rodriguez JL, Campone M, Hamilton E, Sohn J, Guarneri V, Okada M, Boyle F, Neven P, Cortés J, Huober J, Wardley A, Tolane SM, Cicin I, Smith IC, Frenzel M, Headley D, Wei R, San Antonio B, Hulstijn M, Cox J, O'Shaughnessy J, Rastogi (2020) P monarchE Committee Members and Investigators. Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-, Node-Positive, High-Risk, Early Breast Cancer. *J Clin Oncol* Doi: <https://doi.org/10.1200/JCO.20.02514>.
- Boeri C, Chiappa C, Galli F et al (2020) Machine Learning techniques in breast cancer prognosis prediction: A primary evaluation. *Cancer Med* 9(9):3234–3243. <https://doi.org/10.1002/cam4.2811>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.