Overestimation of late distant recurrences in high-risk patients with ER-positive breast cancer: validity and accuracy of the CTS5 risk score in the TEAM and IDEAL trials

Noordhoek, I.; Blok, E.J.; Kranenbarg, E.M.K.; Putter, H.; Duijm-de Carpentier, M.; Rutgers, E.J.T.; ... ; Liefers, G.J.

Citation

Overestimation of Late Distant Recurrences in High-Risk Patients With ER-Positive Breast Cancer: Validity and Accuracy of the CTS5 Risk Score in the TEAM and IDEAL Trials

Iris Noordhoek, BSc[1,2]; Erik J. Blok, MD, PhD[3]; Elma Meershoek-Klein Kranenbarg, MSc[1]; Hein Putter, PhD[4];
Marjolijn Duijm-de Carpentier[1]; Emiel J. T. Rutgers, MD, PhD[5]; Caroline Seynaeve, MD, PhD[6]; John M. S. Bartlett, PhD[7,8];
Jean-Michel Vannetzel, MD[9]; Daniel W. Rea, MBBS, BSc, PhD[10]; Annette Hasenburg, DrMed[11]; Robert Paridaens, MD, PhD[12];
Christos J. Markopoulos, PhD[13]; Yasuo Hozumi, MD, PhD[14]; Johanneke E. A. Portielje, MD, PhD[2]; Judith R. Kroep, MD, PhD[2];
Cornelis J. H. van de Velde, PhD[1]; and Gerrit-Jan Liefers, MD, PhD[1]

abstract

PURPOSE Most distant recurrences (DRs) in women with hormone receptor–positive breast cancer occur after 5 years from diagnosis. The Clinical Treatment Score post-5 years (CTS5) estimates DRs after 5 years of adjuvant endocrine therapy (AET). The aim of this study was to externally validate the CTS5 as a prognostic/predictive tool.

METHODS The CTS5 categorizes patients who have been disease free for 5 years into low, intermediate, and high risk and calculates an absolute risk for developing DRs between 5 and 10 years. Discrimination and calibration were assessed using data from the TEAM and IDEAL trials. The predictive value of the CTS5 was tested with data from the IDEAL trial.

RESULTS A total of 5,895 patients from the TEAM trial and 1,591 patients from the IDEAL trial were included. When assessing the CTS5 discrimination, significantly more DRs were found at 10 years after diagnosis in the CTS5 high- and intermediate-risk groups than in the low-risk group (hazard ratio, 5.7 [95% CI, 3.6 to 8.8] and 2.8 [95% CI, 1.7 to 4.4], respectively). In low- and intermediate-risk patients, the CTS5-predicted DR rates were higher, although not statistically significantly so, than observed rates. However, in high-risk patients, the CTS5-predicted DR rates were significantly higher than observed rates (29% v 19%, respectively; $P < .001$). The CTS5 was not predictive for extended AET duration.

CONCLUSION The CTS5 score as applied to patients treated in the TEAM and IDEAL cohorts discriminates between risk categories but overestimates the risk of late DRs in high-risk patients. Therefore, the numerical risk assessment from the CTS5 calculator in its current form should be interpreted with caution when used in daily clinical practice, particularly in high-risk patients.

J Clin Oncol 38:3273-3281. © 2020 by American Society of Clinical Oncology

INTRODUCTION

The disease course of estrogen receptor (ER)–positive breast cancer can be influenced significantly by targeting ER with adjuvant endocrine therapy (AET).[1] Currently, the optimal AET regimen for postmenopausal patients consists of either 5 years of treatment with an aromatase inhibitor (AI) or 2-3 years of tamoxifen followed by 2-3 years of an AI.[2] Tamoxifen monotherapy for 5 years has been proven an inferior treatment.[3]

The majority of distant recurrences (DRs) occur after the first 5 years from diagnosis (ie, after AET has been stopped).[4] Increasingly, attention is being drawn to the challenge of preventing late relapse in ER-positive breast cancer. Extending AET to a total of 10 years has proven beneficial for a minority of patients, if the initial treatment consisted of tamoxifen monotherapy,

both for extended treatment with tamoxifen and for switching to an AI.[5,6] However, studies so far have not shown a clinically relevant benefit of extended therapy if the initial treatment included an AI.[7-10] AET is accompanied by significant toxicity, and extended therapy should only be prescribed after carefully weighing harms and benefits.[7,11,12]

Over the past decades, the field of estimating disease recurrences has mostly shifted toward gene expression profiles. For example, gene expression profiling can be used to identify luminal A and luminal B tumors, 2 subtypes of ER-positive breast cancer that reflect a different tumor biology and disease prognosis. Neither subtype, however, is predictive for a better response to endocrine therapy.[13-18] Moreover, gene expression profiles are expensive and not universally

CONTEXT

Key Objective

Most breast cancer recurrences occur after 5 years from diagnosis, when adjuvant therapy has been discontinued. Predicting which patients are most likely to experience late recurrences is crucial for optimizing treatment strategies. The Clinical Treatment Score post-5 years (CTS5) aims to predict this risk. This study externally validated the CTS5 using independent trial data and examined whether the CTS5 can predict benefit of extended adjuvant therapy.

Knowledge Generated

The CTS5 provides accurate risk estimation for patients with low and intermediate risks but overestimates the risk of recurrence in high-risk patients in the TEAM and IDEAL cohorts. The CTS5 could not predict benefit of 5 years over 2.5 years of extended adjuvant therapy in the IDEAL cohort.

Relevance

External validation of the CTS5 in the TEAM and IDEAL trials indicated overestimation of recurrence risk in patients deemed high risk. In this population, using the CTS5 would create a potential risk of overtreatment.

available. The value of gene expression profiling in routine clinical practice and, more specifically, the identification of patient subgroups where they should be used, remains challenging.[19] With this rationale, the Clinical Treatment Score post-5 years (CTS5) has been described using data from the ATAC and BIG-1-98 trials.[20-22] The CTS5 is a prognostic tool that aims to estimate the risk for late DRs and categorizes postmenopausal patients into low-, intermediate-, and high-risk groups on the basis of commonly reported clinicopathologic parameters and is therefore cheap and easy to use.

Prognostic models, such as the CTS5, are important because they provide additional information about the disease course and can help to guide the optimal treatment strategy. A crucial aspect of implementing prognostic models into daily clinical practice is the external validation of these models because using unvalidated models could lead to inappropriate modifications of treatment regimens. The CTS5 was created with patient data from 2 trial cohorts and was validated using the combined patient cohorts of these 2 trials[20-22] and, therefore, requires further validation. This study aims to externally validate the CTS5 as a prognostic tool for late DRs and to assess it as a predictive tool for choosing the optimal duration of extended AET using patient data from 2 prospective randomized clinical trials: the Tamoxifen and Exemestane Adjuvant Multicenter (TEAM) trial[2] and the Investigation on the Duration of Extended Adjuvant Letrozole (IDEAL) trial.[7]

METHODS

Algorithm

The CTS5 categorizes postmenopausal patients with ER-positive breast cancer who are disease free after 5 years of standard AET into low- (< 5%), intermediate- (5%-10%), and high- (> 10%) risk groups for developing a DR between 5 and 10 years from diagnosis.[21] This categorization

is based on several clinical parameters described in the Appendix (online only). The online CTS5 calculator provides both the risk category and an absolute risk of developing a late DR for individual patients.[23]

TEAM Study Cohort

TEAM is a phase III trial that randomly assigned postmenopausal patients with ER-positive breast cancer to either 5 years of exemestane or 2-3 years of tamoxifen followed by 2-3 years of exemestane. All patients were randomly assigned within 1 month after diagnosis. Details of the trial have been reported previously.[2,24] For the analyses of this study, all patients were included who were disease free at 5 years after random assignment and for whom all clinicopathologic data were available.

IDEAL Study Cohort

IDEAL is a phase III randomized trial that investigated the optimal duration of letrozole after standard AET in postmenopausal patients with ER-positive early breast cancer. Patients needed to be disease free after 5 years of standard AET before random assignment to either 2.5 or 5 years of extended treatment with letrozole. More than 90% of patients were randomly assigned within 6 months after stopping standard AET. No significant difference in disease-free survival was found between the treatment arms after a median follow-up of 6.6 years after random assignment. Details of the trial have been reported previously.[7] For this study, all patients were included for whom all clinicopathologic data were available.

Validation as Prognostic Tool

The CTS5 is validated as a prognostic tool if it is able to significantly differentiate patients into low-, intermediate-, and high-risk groups for developing DRs and if the predicted absolute risks for developing DRs correspond with the observed DR rates (ie, is properly calibrated).

Validation as Predictive Tool

The CTS5 is validated as a predictive tool for choosing between 2.5 years and 5 years extended AET if there is statistically significant interaction between the risk categories and the treatment arms of the IDEAL trial.

Statistical Analyses

The primary end point for the CTS5 is a late DR, defined as the occurrence of a DR between 5 and 10 years after diagnosis. For this study, this is translated to the occurrence of a DR at 10 years after random assignment in the TEAM cohort and at 5 years after random assignment in the IDEAL cohort.

Kaplan-Meier survival estimates and cumulative incidence survival estimates were used to determine the discriminative prognostic performance of CTS5. Hazard ratios (HRs) and corresponding 95% CIs and P values for Kaplan-Meier survival estimates were estimated from univariable Cox regression models. HRs and 95% CIs for cumulative incidence survival estimates were derived from Fine and Gray analyses.[25,26]

To examine the calibration, the observed DR rates were compared with the predicted number of DRs (calculations described in the Appendix). To test the predictive value of the CTS5 for extended AET, Cox regression was used to test for interaction between the risk categories and treatment arms of the IDEAL cohort. Treatment arm and risk category were used as coefficients in the Cox regression interaction model. $P < .05$ was considered statistically significant. Analyses were performed using SPSS version 23.0 (IBM Corporation, Chicago, IL) and R 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria) statistical software.

Ethical Standards

The study was conducted in compliance with the guidelines of the Declaration of Helsinki, International Conference on Harmonization and Good Clinical Practice. The TEAM trial was approved by the medical ethics committee of the Erasmus Medical Center Rotterdam (198.231/2001/7). The IDEAL trial was approved by the medical ethics committee of the Leiden University Medical Center (P06.217).

RESULTS

Patients in the Study Cohorts

The TEAM trial consists of 9,779 patients. Of those patients, 2,366 were excluded because they withdrew informed consent, experienced a relapse, died, or were lost to follow-up before 5 years from random assignment. Another 391 patients were excluded because they continued treatment in the IDEAL trial. For 1,127 of the remaining 7,022 patients, a CTS5 and accompanying risk prediction could not be calculated because of missing clinicopathologic parameters. Thus, the cohort used in this study consisted of 5,895 patients (hereafter called the TEAM cohort; Fig 1A). For the TEAM cohort, the observed late DR rate at 10 years after random assignment was 8.7%.

The IDEAL trial consists of 1,824 patients. For 233 patients, a CTS5 and risk prediction could not be calculated, which left 1,591 patients for this analysis (hereafter called the IDEAL cohort; Fig 1B). For the IDEAL cohort, the observed late DR rate at 5 years after random assignment was 6.8%.

Comparison of Cohorts

The observed late DR rates were 7.0% for the ATAC cohort and 5.5% for the BIG-1-98 cohort used to train and test the CTS5.[21] In the TEAM cohort, 2,113 patients (35.8%) were categorized into the low-risk group, 2,159 (36.6%) into the intermediate-risk group, and 1,623 (27.5%) into the high-risk group. Patients from the TEAM cohort were comparable to patients in the ATAC and BIG-1-98 cohorts with regard to baseline characteristics, treatment strategy (use of chemotherapy and type of AET), and overall late DR rates[20-22] (Table 1). Patients in the TEAM cohort had slightly larger tumors (43% of patients had T2+ tumors v 32% and 35% in the ATAC and BIG-1-98 cohorts, respectively; $P < .001$), and more patients had N+ disease (40% v 32% and 40%, respectively; $P < .001$).

In the IDEAL cohort, 343 patients (21.6%) were categorized into the low-risk group, 649 (40.8%) into the intermediate-risk group, and 599 (37.6%) into the high-risk group. Patients in the IDEAL cohort had some baseline characteristics that would classify them at higher risk than the patients in the ATAC and BIG-1-98 cohorts, such as age at diagnosis and number of positive lymph nodes (Table 1). This resulted in a different distribution of patients over the risk categories and is reflected in a significantly higher proportion of patients who received chemotherapy. Nevertheless, within the 3 CTS5 risk categories, there were no significant differences among the cohorts in baseline characteristics and risk factors apart from age. In addition, the overall late DR rates at 10 years were comparable between the IDEAL cohort and the ATAC and BIG-1-98 cohorts.

Prognostic Validation

Discrimination. In the TEAM cohort, the DR-free interval (DRFi) was 97% (95% CI, 96% to 98%) in the low-risk group, 93% (95% CI 91% to 94%) in the intermediate-risk group, and 86% (95% CI, 84% to 88%) in the high-risk group. The DRFi was significantly lower in the high-risk group (HR, 5.7; 95% CI, 3.6 to 8.8; log-rank $P < .001$) and the intermediate-risk group (HR, 2.8; 95% CI, 1.7 to 4.4; log-rank $P < .001$) compared with the low-risk group (Fig 2A).

In the IDEAL cohort, the DRFi was 98% (95% CI, 96% to 99%) in the low-risk group, 95% (95% CI, 93% to 97%) in the intermediate-risk group, and 89% (95% CI, 87% to 92%) in the high-risk group. The DRFi was significantly lower in the high-risk group (HR, 4.8; 95% CI, 2.3 to 10.2; log-rank $P < .001$) and intermediate-risk group (HR, 2.2; 95% CI, 1.0 to 4.8; log-rank $P = .037$) compared with the low-risk group as well (Fig 2B).
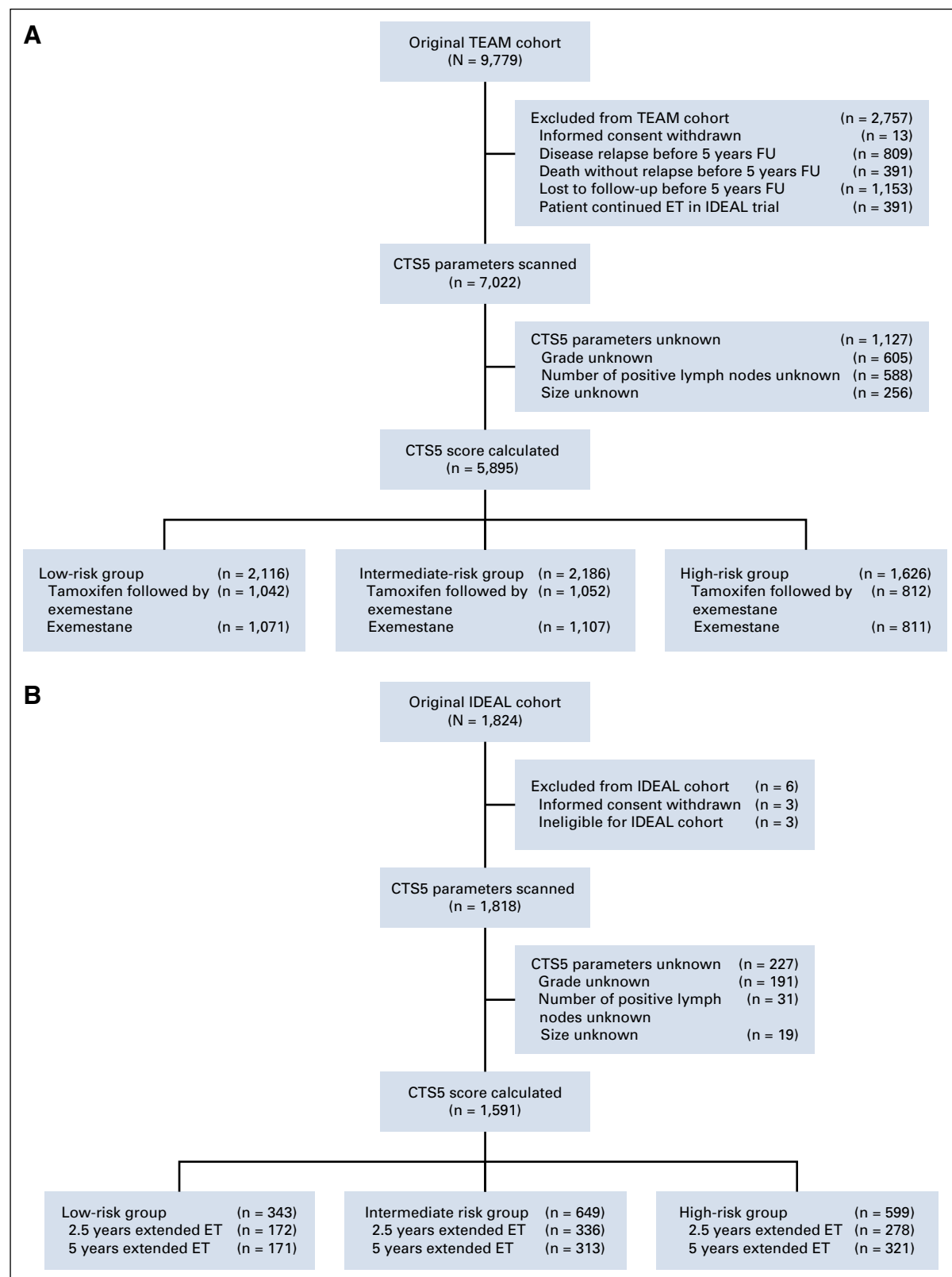
FIG 1. (A) CONSORT diagram to account for missing patients in the TEAM cohort.[2] (B) CONSORT diagram to account for missing patients in the IDEAL cohort.[7] CTS5, Clinical Treatment Score post-5 years[21]; DR, distant recurrence; ET, endocrine treatment.

TABLE 1. Overview of Clinicopathologic and Demographic Characteristics of the Patients in the TEAM and IDEAL Cohorts

Cohort	Total	Risk Group, No. (%)		
		Low	Intermediate	High
TEAM				
No. of patients	5,895	2,113 (35.8)	2,159 (36.6)	1,623 (27.5)
Median age, years (IQR)	63 (57-70)	61 (56-66)	63 (58-70)	66 (59-74)
Tumor size, mm				
< 10	376 (6.4)	343 (16.2)	30 (1.4)	3 (0.2)
10-20	2,968 (50.3)	1,688 (79.9)	1,037 (48.0)	243 (15.0)
> 20	2,551 (43.3)	82 (3.9)	1,092 (50.6)	1,377 (84.8)
Histologic grade				
1	1,163 (19.7)	757 (35.8)	304 (14.1)	102 (6.3)
2	3,176 (53.9)	1,195 (56.6)	1,221 (56.6)	760 (46.8)
3	1,556 (26.4)	161 (7.6)	634 (29.4)	761 (46.9)
Positive lymph nodes				
0	3,501 (59.4)	1,855 (87.8)	1,359 (62.9)	287 (17.7)
1	1,182 (20.1)	239 (11.3)	572 (26.5)	371 (22.9)
2-3	722 (12.2)	19 (0.9)	208 (9.6)	495 (30.5)
4-9	377 (6.4)	—	20 (0.9)	357 (22.0)
> 9	113 (1.9)	—	—	113 (7.0)
Prior chemotherapy	1,929 (32.7)	398 (18.8)	789 (36.5)	742 (45.7)
Allocated ET 5 years EXE	2,989 (50.7)	1,071 (50.7)	1,107 (51.3)	811 (50.0)
2-3 years TAM, 2-3 years EXE	2,906 (49.3)	1,042 (49.3)	1,052 (48.7)	812 (50.0)
IDEAL				
No. of patients	1,591	343 (21.6)	649 (40.8)	599 (37.6)
Median age, years (IQR)	55 (49-61)	52 (47-58)	55 (48-62)	56 (50-62)
Tumor size, mm				
< 10	56 (3.5)	47 (13.7)	8 (1.2)	1 (0.2)
10-20	607 (38.2)	251 (73.2)	267 (41.1)	89 (14.9)
> 20	928 (58.3)	45 (13.1)	374 (57.6)	509 (85.0)
Histologic grade				
1	277 (17.4)	128 (37.3)	105 (16.2)	44 (7.3)
2	760 (47.8)	161 (46.9)	310 (47.8)	289 (48.2)
3	554 (34.8)	54 (15.7)	234 (36.1)	266 (44.4)
Positive lymph nodes				
0	614 (38.6)	236 (68.8)	344 (53.0)	34 (5.7)
1	388 (24.4)	94 (27.4)	192 (29.6)	102 (17.0)
2-3	301 (18.9)	13 (3.8)	98 (15.1)	190 (31.7)
4-9	220 (13.8)	—	14 (2.2)	206 (34.4)
> 9	68 (4.3)	—	1 (0.2)	67 (11.2)
Prior chemotherapy	1,085 (68.2)	230 (67.1)	403 (62.1)	452 (75.5)
Initial ET 5 years TAM	180 (11.3)	44 (12.8)	82 (12.6)	54 (9.0)
5 years AI	465 (29.2)	85 (24.8)	189 (29.1)	191 (31.9)
2-3 years TAM, 2-3 years AI	946 (59.5)	214 (62.4)	378 (58.2)	354 (59.1)
Allocated extended ET 2.5 years	786 (49.4)	172 (50.1)	336 (51.8)	278 (46.4)
5 years	805 (50.6)	171 (49.9)	313 (48.2)	321 (53.6)

Abbreviations: AI, aromatase inhibitor; ET, endocrine therapy; EXE exemestane; IQR, interquartile range; TAM, tamoxifen.
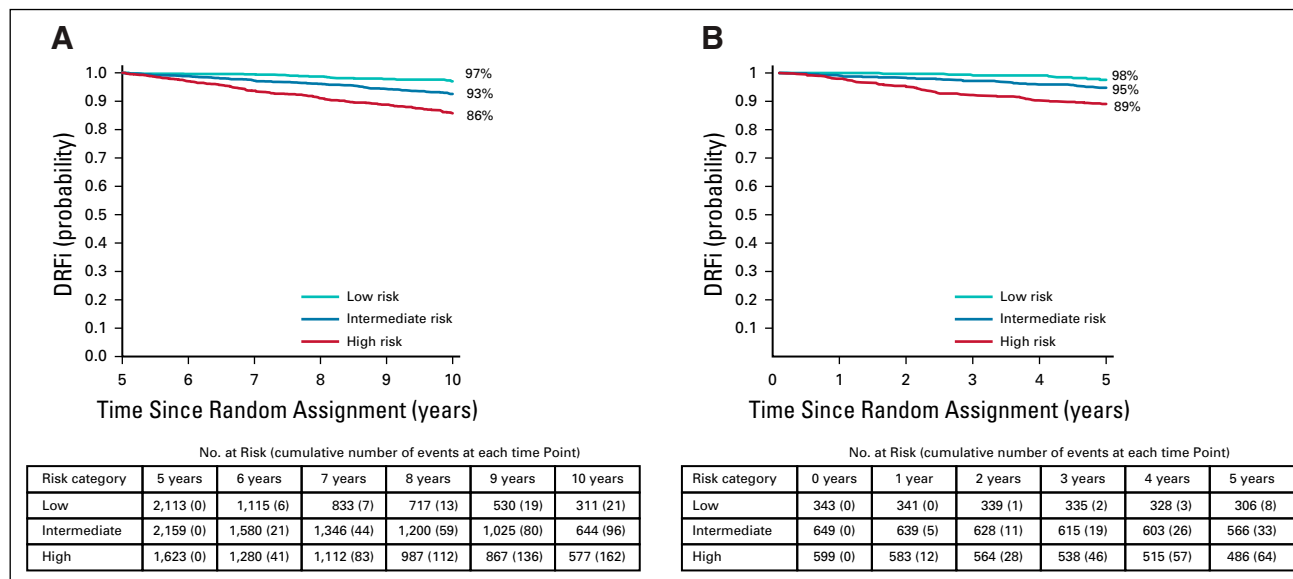
FIG 2. (A) Kaplan-Meier survival estimates of observed late distant recurrence (DR) rates with accompanying risk table, indicating discriminatory prognostic value of the Clinical Treatment Score post-5 years (CTS5) for DRs in patients from the TEAM cohort. (B) Kaplan-Meier survival estimates of observed late DR rates with accompanying risk table, indicating discriminatory prognostic value of the CTS5 for DRs in patients from the IDEAL cohort. DRFi, distant recurrence–free interval.

Calibration. In patients from the TEAM cohort with a predicted late DR risk up to 7%, the observed late DR rates corresponded with the predicted rates. However, in patients with higher predicted risks, the observed late DR rates were significantly lower than the predicted rates (Table 2; Fig 3A). In the highest decile, the absolute difference between the predicted DR rate and the observed rate was 10% (29.0% v 19.3% [95% CI, 15.3% to 23.0%], respectively; Appendix Fig A1A, online only).

In patients from the IDEAL cohort with predicted risk up to 8%, the predicted rate was higher than the observed rates, although not statistically significant. However, in patients with a higher predicted risk, the predicted rate was significantly higher than the observed rates (Table 2; Fig 3B). In the patients with the highest predicted risk, the predicted DR rate was 31.6%, while the observed rate was only 15.5% (95% CI, 9.6% to 21.0%; Appendix Fig A1B). The observed late DR rates using the cumulative incidence

method did not differ significantly from the rates obtained with the Kaplan-Meier method (data not shown).

Predictive Validation

In the IDEAL cohort, there was no statistically significant interaction between the risk categories and the treatment allocation ($P = .5$). No difference in observed DR rates was seen between 2.5 and 5 years of extended AET in the low-risk (HR, 0.6; 95% CI, 0.1 to 2.5), intermediate-risk (HR, 1.7; 95% CI, 0.8 to 3.3), or high-risk (HR, 1.2; 95% CI, 0.7 to 1.9) groups (Appendix Table A1, online only). Thus, there was no risk category in which the treatment allocation had a significant effect on the occurrence of late DR.

DISCUSSION

Considering the substantial risk of late DRs for patients with ER-positive breast cancer, even after 5 years of AET, and the potential benefit of extended AET, there is a demand for prognostic and predictive models.[4,27] The CTS5 was developed for this purpose. It aims to estimate the DR rate between 5 and 10 years from diagnosis for postmenopausal patients with ER-positive breast cancer who remain disease free after 5 years of standard AET.[21]

As applied to the TEAM and IDEAL cohorts, the CTS5 is able to discriminate postmenopausal patients with ER-positive breast cancer into 3 risk categories with respect to late DR. In the 2 large cohorts of trial patients used in these analyses, the risk of late DR as predicted by the CTS5 corresponded to the observed DR rates in low-risk patients, but the CTS5 overestimated the observed risk of DR in patients with a higher predicted risk. Furthermore, the

TABLE 2. Observed and Predicted Probability of Late Distant Recurrences by Risk Category in the TEAM and IDEAL Cohorts

Cohort and Risk Category	Predicted, %	Observed, % (95% CI)
TEAM		
Low	3.4	3.0 (1.6 to 4.3)
Intermediate	7.2	7.4 (5.8 to 8.6)
High	19.0	14.2 (12.1 to 16.3)
IDEAL		
Low	3.6	2.4 (0.8 to 4.1)
Intermediate	7.2	5.2 (3.5 to 6.9)
High	19.5	10.9 (8.4 to 13.4)
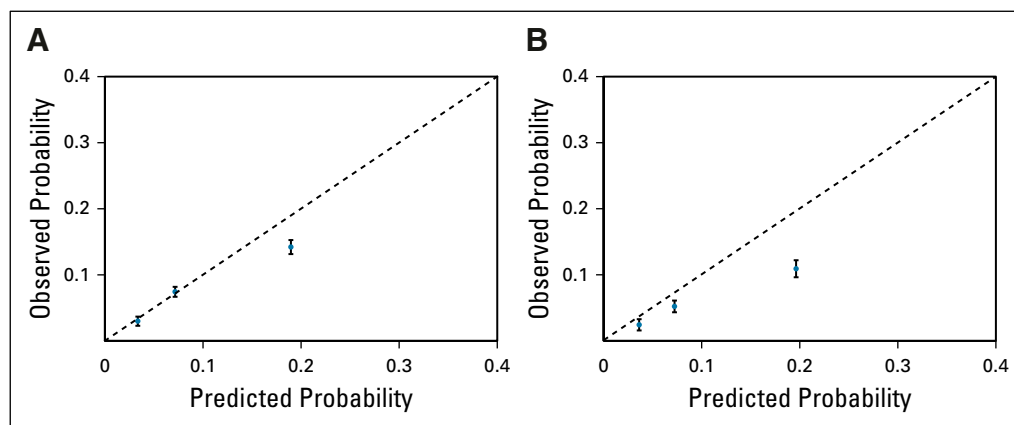
FIG 3. (A) Observed and predicted probability of distant recurrences (DRs) with accompanying 95% CIs in the 3 risk categories in the TEAM cohort. (B) Observed and predicted probability of DRs with accompanying 95% CIs in the 3 risk categories in the IDEAL cohort.

CTS5 could not predict benefit of 5 years over 2.5 years of extended AET using data from the IDEAL cohort.

The discrimination of the CTS5 has also been tested and validated in 1 retrospective cohort of nontrial patients with ER-positive breast cancer; however, the calibration was not tested in that study.[28] Although our study is one of the first that aimed to externally validate it, the CTS5 is already being used in clinical practice through the online calculator.[23]

Patients in the TEAM cohort, who had ER-positive breast cancer and were disease free for 5 years on either exemestane monotherapy or tamoxifen followed by exemestane, were comparable to patients in both cohorts used to create the CTS5 with regard to most risk factors, treatment strategy, and overall DR rates. The proportional distribution of patients into the 3 risk categories was also similar.[20-22] Because there was no difference in the definition of DR, it was expected that the predicted late DR risk would not differ from the observed DR rates. This was confirmed in low- and intermediate-risk patients. However, in the high-risk patients, the predicted risk was significantly higher than the observed late DR rate (19% and 14%, respectively). Although 14% would still be categorized as high risk, a more valid numerical estimation can lead to more accurately tailored treatment advice.

Patients in the IDEAL cohort who had ER-positive breast cancer and were disease free for 5 years on any AET had comparable baseline characteristics within the risk categories to both CTS5 cohorts and had comparable overall DR rates. However, IDEAL patients differed from the CTS5 cohorts with regard to treatment strategy; patients in the IDEAL cohort were treated with extended AET for either 2.5 or 5 years. Preliminary results from the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) meta-analysis on extended AET showed that extending AET beyond the first 5 years may yield an absolute risk reduction of late DR of 1%-5% at 10 years after diagnosis, with the effect size depending on the type of AET given during the first 5 years.[29]

When applying the CTS5 to the IDEAL cohort, the predicted DR rates were similar to the observed DR rates in low-risk patients, but in the highest risk decile, the CTS5 predicted a DR rate of 31.6%, while we observed a DR rate of 15.5% (95% CI, 9.6% to 21.0%). It is possible that the difference in treatment strategy led to a slight reduction in late DR rates (up to 5% according to the preliminary data of the EBCTCG), but extended treatment is unlikely to account for the 16% risk difference that was observed in this analysis.

A potential explanation for the difference in observed and predicted DR rates is the discrepancy in the years of diagnosis between the cohorts. While patients in the ATAC and BIG-1-98 cohorts were diagnosed between 1996 and 2003, patients in the TEAM and IDEAL cohorts were diagnosed between 2001 and 2006. Over this period, significant improvements in diagnostic accuracy (reliability of hormone receptor and human epidermal growth factor receptor 2 [HER2] status determination[30]) and systemic therapy (chemotherapy regimens and anti-HER2 medication) were made. Moreover, in the ATAC and BIG-1-98 cohorts, 20% and 24% of patients were treated with chemotherapy, while this was 33% and 68% in the TEAM and IDEAL cohorts, respectively. This could explain the observation that the largest difference between observed and predicted late DR rates was seen in high-risk patients.

The most common reason for excluding patients from the TEAM and IDEAL trial cohorts to create the cohorts used in this study was unknown grade (Fig 1). This was mostly due to the difficulty in scoring the histologic grade of lobular carcinomas.[31] Because there were no differences in late DR rates between the included patients and patients who were excluded on the basis of unknown grade (data not shown), this is not likely to bias the results of our analyses.

Furthermore, an aspect to keep in mind is the difference between prognostic and predictive tools.[32] Prognostic tools aim to distinguish patients with an inherently worse prognosis from those with a better prognosis, while predictive tools aim to distinguish those patients who will respond well to treatment from those who will not. Often, prognostic tools are not predictive and should not be used as such because patients with a worse prognosis are not necessarily the

same patients who benefit from more extensive therapy.[33] In patients treated in the IDEAL trial, the CTS5 was not able to select patients who benefit from 5 years of extended AET as opposed to 2.5 years. Because of the design of the IDEAL trial, it was not possible to investigate whether the CTS5 is predictive of the type of extended AET (ie, tamoxifen *v* AI *v* no extended AET).

Future prognostic and predictive models will most likely focus on genetic and/or functional profiles, and although these have shown promising results so far, accessibility and availability in daily clinical practice need to be prioritized as well.[19] This is one major advantage of the CTS5 because it is based on clinico-pathologic parameters that are available for all patients and is cheap and easy to implement across all health care settings.

In conclusion, the CTS5 as applied to patients treated in the TEAM and IDEAL cohorts categorizes patients into low-, intermediate-, and high-risk groups. In low-risk patients, the predicted late DR risks correspond with the observed DR rates, but the CTS5 overestimates the risk of late DRs in high-risk patients from the TEAM and IDEAL cohorts. Using patient data from the IDEAL cohort, the CTS5 cannot be validated as a predictive tool for extended AET either. Especially in high-risk patients, an unrealistic assessment of the risk to develop a late DR could potentially lead to overtreatment. Therefore, the numerical risk assessment from the CTS5 calculator in its current form should be interpreted with caution when used in daily clinical practice, particularly when used in high-risk patients.

AFFILIATIONS

[1]Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands
[2]Department of Medical Oncology, Leiden University Medical Center, Leiden, the Netherlands
[3]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
[4]Department of Medical Statistics, Leiden University Medical Center, Leiden, the Netherlands
[5]Department of Surgery, Netherlands Cancer Institute, Amsterdam, the Netherlands
[6]Department of Medical Oncology, Erasmus Medical Center, Rotterdam, the Netherlands
[7]Diagnostic Development Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada
[8]Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK
[9]Department of Radiation Oncology, Clinique Hartmann, Neuilly-sur-Seine, France
[10]Department of Medical Oncology, University of Birmingham, Birmingham, UK
[11]Department of Gynecology and Obstetrics, University Hospital Mainz, Mainz, Germany
[12]Department of Medical Oncology, University Hospitals Leuven, Leuven, Belgium
[13]Medical School, National and Kapodistrian University of Athens, Athens, Greece
[14]Department of Breast and Endocrine Surgery, University of Tsukuba Hospital, Ibaraki, Japan

CORRESPONDING AUTHOR

Gerrit-Jan Liefers, MD, PhD, Department of Surgery, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands; e-mail: g.j.liefers@lumc.nl.

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 365:1687-1717, 2005

2. Derks MGM, Blok EJ, Seynaeve C, et al: Adjuvant tamoxifen and exemestane in women with postmenopausal early breast cancer (TEAM): 10-Year follow-up of a multicentre, open-label, randomised, phase 3 trial. Lancet Oncol 18:1211-1220, 2017

3. Early Breast Cancer Trialists' Collaborative Group: Aromatase inhibitors versus tamoxifen in early breast cancer: Patient-level meta-analysis of the randomised trials. Lancet 386:1341-1352, 2015

4. Pan H, Gray R, Braybrooke J, et al: 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 377:1836-1846, 2017

5. Davies C, Pan H, Godwin J, et al: 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, 2013

6. Goss PE, Ingle JN, Martino S, et al: A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. N Engl J Med 349:1793-1802, 2003

7. Blok EJ, Kroep JR, Meershoek-Klein Kranenbarg E, et al: Optimal duration of extended adjuvant endocrine therapy for early breast cancer; results of the IDEAL trial (BOOG 2006-05). J Natl Cancer Inst 110:djx134, 2018

8. Gnant M, Steger G, Greil R, et al: A prospective randomized multi-center phase-III trial of additional 2 versus additional 5 years of anastrozole after initial 5 years of adjuvant endocrine therapy–results from 3,484 postmenopausal women in the ABCSG-16 trial. San Antonio Breast Cancer Symposium, San Antonio, TX, December 5-9, 2017 (abstr GS3-01)

9. Mamounas EP, Bandos H, Lembersky BC, et al: Use of letrozole after aromatase inhibitor-based therapy in postmenopausal breast cancer (NRG Oncology/NSABP B-42): A randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 20:88-99, 2019

10. Tjan-Heijnen VCG, van Hellemond IEG, Peer PGM, et al: Extended adjuvant aromatase inhibition after sequential endocrine therapy (DATA): A randomised, phase 3 trial. Lancet Oncol 18:1502-1511, 2017

11. Blok EJ, Kroep JR, Meershoek-Klein Kranenbarg E, et al: Treatment decisions and the impact of adverse events before and during extended endocrine therapy in postmenopausal early breast cancer. Eur J Cancer 95:59-67, 2018

12. Niravath P, Rimawi MF, Osborne CK: Aromatase inhibitor adverse effects: Are we sweeping them under the rug? J Clin Oncol 32:3779, 2014

13. Ellis MJ, Suman VJ, Hoog J, et al: Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: Clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. J Clin Oncol 29:2342-2349, 2011

14. Nielsen TO, Parker JS, Leung S, et al: A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res 16:5222-5232, 2010

15. Parker JS, Mullins M, Cheang MC, et al: Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27:1160-1167, 2009

16. Prat A, Cheang MC, Martín M, et al: Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. J Clin Oncol 31:203-209, 2013

17. Prat A, Ellis MJ, Perou CM: Practical implications of gene-expression-based assays for breast oncologists. Nat Rev Clin Oncol 9:48-57, 2011

18. Viale G, de Snoo FA, Slaets L, et al: Immunohistochemical versus molecular (BluePrint and MammaPrint) subtyping of breast carcinoma. Outcome results from the EORTC 10041/BIG 3-04 MINDACT trial. Breast Cancer Res Treat 167:123-131, 2018

19. Blok EJ, Bastiaannet E, van den Hout WB, et al: Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. Cancer Treat Rev 62:74-90, 2018

20. Cuzick J, Sestak I, Baum M, et al: Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-Year analysis of the ATAC trial. Lancet Oncol 11:1135-1141, 2010

21. Dowsett M, Sestak I, Regan MM, et al: 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 36:1941-1948, 2018

22. Regan MM, Neven P, Giobbie-Hurder A, et al: 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, 2011

23. Queen Mary University of London: CTS5 calculator. https://www.cts5-calculator.com

24. van de Velde CJH, Rea D, Seynaeve C, et al: Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): A randomised phase 3 trial. Lancet 377:321-331, 2011

25. Andersen PK, Geskus RB, de Witte T, et al: Competing risks in epidemiology: Possibilities and pitfalls. Int J Epidemiol 41:861-870, 2012

26. Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94:496-509, 1999

27. Davies C, Godwin J, Gray R, et al: Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: Patient-level meta-analysis of randomised trials. Lancet 378:771-784, 2011

28. Richman J, Ring AE, Dowsett M, et al: Clinical validity of CTS5 for estimating risk of late recurrence in unselected, non-trial patients with early ER+ breast cancer. J Clin Oncol 37, 2019 (suppl; abstr 514)

29. Gray R: Effects of prolonging adjuvant aromatase inhibitor therapy beyond five years on recurrence and cause-specific mortality: An EBCTCG meta-analysis of individual patient data from 12 randomised trials including 24,912 women. San Antonio Breast Cancer Symposium, San Antonio, TX, December 4-7, 2018 (abstr GS3-03)

30. Wolff AC, Hammond MEH, Allison KH, et al: Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. J Clin Oncol 36:2105-2122, 2018

31. Barroso-Sousa R, Metzger-Filho O: Differences between invasive lobular and invasive ductal carcinoma of the breast: Results and therapeutic implications. Ther Adv Med Oncol 8:261-266, 2016

32. Ballman KV: Biomarker: Predictive or prognostic? J Clin Oncol 33:3968-3971, 2015

33. Sechidis K, Papangelou K, Metcalfe PD, et al: Distinguishing prognostic and predictive biomarkers: An information theoretic approach. Bioinformatics 34:3365-3376, 2018

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Overestimation of Late Distant Recurrences in High-Risk Patients With ER-Positive Breast Cancer: Validity and Accuracy of the CTS5 Risk Score in the TEAM and IDEAL Trials

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

John M. S. Bartlett
Honoraria: Oncology Education, NanoString Technologies, Biotheranostics
Consulting or Advisory Role: Insight Genetics, BioNTech AG, Biotheranostics, RNA Diagnostics, Pfizer, OncoXchange
Research Funding: NanoString Technologies, Stratifyer, Genoptix, Thermo Fisher Scientific, Agendia, Biotheranostics
Patents, Royalties, Other Intellectual Property: Five pending patents: January 2017: Methods and Devices for Predicting Anthracycline Treatment Efficacy, US utility–15/325,472, EPO–15822898.1, Canada–not yet assigned (Inst); January 2017: Systems, Devices and Methods for Constructing and Using a Biomarker, US utility–15/328,108, EPO–15824751.0, Canada–not yet assigned (Inst); October 2016: Histone Gene Module Predicts Anthracycline Benefit, PCT/CA2016/0002474 (Inst); December 2016: 95-Gene Signature of Residual Risk Following Endocrine Treatment, PCT/CA2016/0003045 (Inst); December 2016: Immune Gene Signature Predicts Anthracycline Benefit, PCT/CA2016/000305 (Inst); Invention: A Molecular Classifier for Personalized Risk Stratification for Patients With Prostate Cancer, August 21, 2019
Travel, Accommodations, Expenses: Biotheranostics, NanoString Technologies

Daniel W. Rea
Honoraria: Roche, Novartis, Pfizer, Eli Lilly, Daiichi Sankyo
Consulting or Advisory Role: Genomic Health, MSD Oncology
Research Funding: Celgene (Inst), Roche (Inst), Biotheranostics (Inst)
Travel, Accommodations, Expenses: Pfizer, Daiichi Sankyo, Eisai, Novartis

Annette Hasenburg
Honoraria: AstraZeneca, MedUpdate, Pfizer, Roche, Streamedup, Tesaro, MedConcept
Consulting or Advisory Role: PharmaMar, Tesaro, Roche
Travel, Accommodations, Expenses: AstraZeneca, MedConcept, Roche, Streamedup, Tesaro, MedUpdate, Pfizer

Christos J. Markopoulos
Honoraria: Genomic Health
Travel, Accommodations, Expenses: Genomic Health

Yasuo Hozumi
Honoraria: AstraZeneca, Chugai Pharma, Pfizer, Eli Lilly

Judith R. Kroep
Consulting or Advisory Role: AstraZeneca (Inst), Novartis (Inst), Pfizer (Inst), Tesaro (Inst), Eli Lilly (Inst), Vitroscan (Inst)
Research Funding: AstraZeneca (Inst), Novartis (Inst), Philips Research (Inst)

Gerrit-Jan Liefers
Research Funding: Biotheranostics (Inst)

No other potential conflicts of interest were reported.

APPENDIX

Methods

Clinical treatment score post-5 years algorithm. The Clinical Treatment Score post-5 years (CTS5) categorizes postmenopausal patients with estrogen receptor (ER)–positive breast cancer who are disease free after 5 years of standard adjuvant endocrine therapy (AET), into low- ($<$ 5%), intermediate- (5%-10%), and high- ($>$ 10%) risk groups for developing a distant recurrence (DR) between 5 and 10 years from diagnosis.[21] This categorization is based on the number of positive lymph nodes, grade, size and quadratic size of the tumor, and age at diagnosis. Consistent with the method used in the original CTS5 publication,[21] tumor size was capped at 30 mm. Nodal status was divided into 5 groups: 0, node negative; 1, 1 positive lymph node; 2, 2-3 positive nodes; 3, 4-9 positive nodes; and 4, $>$ 9 positive nodes. The CTS5 score is calculated using the following equation:

$$CTS5 = 0.438 \times nodes + 0.988 \times (0.093 \times size - 0.001 \times size^2 + 0.375 \times grade + 0.017 \times age)$$

The cutoff values were established at $<$ 3.13 for the low-risk group, 3.13-3.86 for the intermediate-risk group, and $>$ 3.86 for the high-risk group.[21]

Statistical analyses. To calculate the predicted late DR rates, 6 random patients from the TEAM cohort and 6 random patients from the IDEAL cohort were entered in the online CTS5 calculator.[23] The risk scores that were provided by the calculator for these patients were used to calculate the cumulative baseline hazard at 5 years used in the algorithm behind the online calculator, using the following equation:

$$"Baseline hazard" = -\ln(1 - "predicted risk")/e^{("CTS5 score")}$$

These 12 baseline hazards were then averaged. This average baseline hazard (0.00223) was used to calculate the predicted risks for the other patients, using the following equation:

$$"Predicted risk" = 1 - (e^{\wedge}(-0.00223 * (e^{(CTS5 score)})))$$

These calculated risk scores were then cross-checked with the risk scores provided by the online calculator for another 12 random patients, 6 from the TEAM cohort and 6 from the IDEAL cohort. These were identical.

The observed DR rates were determined using Kaplan-Meier survival estimates and cumulative incidence survival estimates[25] to account for death as a competing risk. To examine the calibration, the observed DR rates were compared with the predicted number of DRs for 10 equal deciles.
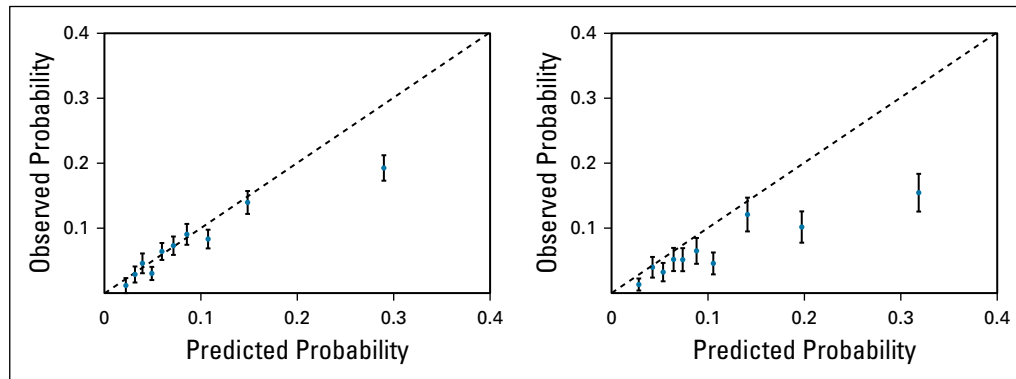
FIG A1. (A) Observed and predicted probability of distant recurrences (DRs) with accompanying 95% CIs in 10 equal deciles in the TEAM cohort. (B) Observed and predicted probability of DRs with accompanying 95% CIs in 10 equal deciles in the IDEAL cohort.

TABLE A1. Absolute Late Distant Recurrence Rates in the IDEAL Cohort Stratified by Treatment Arm and Risk Group

	Treatment Arm, % (95% CI)		
Risk Group	**2.5 Years**	**5 Years**	**P***
Low	3.1 (0.4 to 5.7)	1.8 (0.0 to 3.8)	.49
Intermediate	4.0 (1.9 to 6.1)	6.5 (3.7 to 9.2)	.15
High	10.0 (6.3 to 13.5)	11.7 (8.1 to 15.2)	.50

*Derived from univariable Cox regression models.