ARTICLE

Comparison of EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant Recurrence After Endocrine Therapy

Richard Buus, Ivana Sestak, Ralf Kronenwett, Carsten Denkert, Peter Dubsky, Kristin Krappmann, Marsel Scheer, Christoph Petry, Jack Cuzick, Mitch Dowsett

Affiliations of authors: The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research, London, UK (RB, MD); Academic Department of Biochemistry, Royal Marsden Hospital, London, UK (RB, MD); Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK (IS, JC); Sividon Diagnostics GmbH, Cologne, Germany (RK, KK, MS, CP); Institute of Pathology, Charité University Hospital and German Cancer Consortium (DKTK), Berlin, Germany (CD); Department of Surgery and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria (PD)

Correspondence to: Mitch Dowsett, The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research and Academic Department of Biochemistry, Royal Marsden Hospital, Fulham Road, London, SW3 6JJ, UK (e-mail: mitch.dowsett@icr.ac.uk).

Abstract

Background: Estimating distant recurrence (DR) risk among women with estrogen receptor–positive (ER+), human epidermal growth factor receptor 2 (HER2)–negative early breast cancer helps decisions on using adjuvant chemotherapy. The 21-gene Oncotype DX recurrence score (RS) is widely used for this. EndoPredict (EPclin) is an alternative test combining prognostic information from an eight-gene signature (EP score) with tumor size and nodal status. We compared the prognostic information provided by RS and EPclin for 10-year DR risk.

Methods: We used likelihood ratio χ^2 and Kaplan-Meier survival analyses to compare prognostic information provided by EP, EPclin, RS, and the clinical treatment score (CTS) of clinicopathologic parameters in 928 patients with ER+ disease treated with five years' anastrozole or tamoxifen. Comparisons were made for early (0-5 years) and late (5-10 years) DR according to nodal status. All statistical tests were two-sided.

Results: In the overall population, EP and EPclin provided substantially more prognostic information than RS (LRχ^2: EP = 49.3; LRχ^2: EPclin = 139.3; LRχ^2: RS = 29.1), with greater differences in late DR and in node-positive patients. EP and EPclin remained statistically significantly prognostic when adjusted for RS (ΔLRχ^2: EP+RS vs RS = 20.2; ΔLRχ^2: EPclin+RS vs RS = 113.8). Using predefined cut-offs, EPclin and RS identified 58.8% and 61.7% patients as low risk, with hazard ratios for non-low vs low risk of 5.99 (95% confidence interval [CI] = 3.94 to 9.11) and 2.73 (95% CI = 1.91 to 3.89), respectively.

Conclusions: EP and EPclin were highly prognostic for DR in endocrine-treated patients with ER+, HER2-negative disease. EPclin provided more prognostic information than RS. This was partly but not entirely because of EPclin integrating molecular data with nodal status and tumor size.

Breast cancer is the most common cancer in women. About 80% of primary breast cancers are estrogen receptor (ER)–positive disease. Patients with ER-positive disease receive adjuvant endocrine therapy after surgery that markedly improves their prognosis (1). A large proportion of patients receiving endocrine therapy have sufficiently low risk to safely avoid chemotherapy. Differentiating these patients from higher-risk patients who may benefit from adjuvant chemotherapy is a priority for breast cancer management (2).

Multigene expression prognostic assays may be used to estimate residual risk of recurrence following surgery and endocrine treatment to aid decisions on the appropriateness of chemotherapy treatment. The most widely used test is the Oncotype DX 21-gene recurrence score (RS) (3). Other prognostic scores to estimate residual risk in endocrine-treated patients include the PAM50 risk of recurrence (ROR) score (4), the Breast Cancer Index (BCI) (5), and the IHC4 test that is immunohistochemically based and is combined with the clinical treatment score (CTS) to integrate clinicopathological parameters (6). The amount of prognostic information provided for early (0-5 years) and late (beyond five years) recurrence varies across these tests (7).

The EndoPredict (EP) assay combines the expression of three proliferative and five ER-signaling/differentiation-associated genes and is normalized by three housekeeping genes (8). EP may be measured in formalin-fixed, paraffin-embedded tissue sections by quantitative real-time polymerase chain reaction (qRT-PCR) in decentralized laboratories (9) and provides a score that ranges between 0 and 15 after scaling. EPclin was derived from EP by incorporating nodal status and tumor size to create an integrated diagnostic algorithm for clinical decisions (8). Both EP and EPclin were trained on a cohort of 964 patients with ER-positive, human epidermal growth factor receptor 2 (HER2)–negative carcinomas treated with adjuvant endocrine therapy only (8). Thresholds for EP and EPclin to differentiate between patients at low or high risk corresponding to a 10% probability of distant recurrence (DR) at 10 years were set at 5 and 3.3, respectively. Both EP and EPclin were shown to be prognostic for early and late distant recurrence in the Austrian Breast and Colorectal Cancer Study Group (ABCSG)-6 and -8 trials (10).

TransATAC, the translational substudy of the Arimidex, Tamoxifen, Alone or in Combination trial (ATAC), served as a validation study for the Oncotype DX RS (11), PAM50 ROR (12), and BCI (13) scores and as a training set for a definition of PAM50 ROR cut-off values and for CTS and IHC4 scores (6).

Our aims were to assess the prognostic value of EP and EPclin for DR in postmenopausal women with hormone receptor–positive, HER2-negative primary breast cancer in TransATAC and to compare their prognostic ability with that of the Oncotype DX RS.

Methods

Patient Cohort, RNA Extraction

The ATAC trial evaluated efficacy and safety of anastrozole vs tamoxifen given for five years in postmenopausal women with localized primary breast cancer (14). TransATAC draws upon formalin-fixed, paraffin-embedded tumor samples from a subset of women randomized to the monotherapy arms. RNA was extracted by Genomic Health Inc. (GHI) (11), and residual RNA was available for 928 ER-positive, HER2-negative women. For this analysis, eligibility required hormone receptor–positive, HER2-negative, chemotherapy-naive disease where RS and at least 350 ng residual RNA were available. A pilot study was conducted that confirmed the suitability of TransATAC samples for EP assessment (described in the Supplementary Methods, available online). This study was approved by the South-East London Research Ethics Committee, and all patients included gave informed consent.

Procedures

EP genes' analysis by qRT-PCR was performed by Sividon, who were blinded to all clinical data. Fifty to 100 ng RNA was used to quantitate the eight cancer-related genes of interest (BIRC5, UBE2C, DHCR7, RBBP8, IL6ST, AZGP1, MGP, and STC2) and three reference genes (CALM2, OAZ1, and RPL37A). EP and EPclin scores were determined as previously described (8). The predefined cut-offs for diagnostic decisions corresponding to a 10% DR rate at 10 years were applied to stratify patients into low- or high-risk groups: EP low risk ($<$5), EP high risk ($\geq$5); EPclin low risk ($<$3.3), EPclin high risk ($\geq$3.3) (8). RS risk groups were determined as previously described, where cut-offs of 18 and 31 in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 trial cohort corresponded to approximately 11% and 20% of 10-year risk of DR (3). In addition to these predefined diagnostic cut-points, we compared DR between tertiles based on the genomic assays to allow a more detailed comparison. CTS was derived as reported previously (6) and calculated with the prespecified algorithm: CTS = 100x{0.417N$_{1-3}$ + 1.566N$_{4+}$ + 0.930(0.497T$_{1-2}$ + 0.882T$_{2-3}$ + 1.838T$_{>3}$ + 0.559Gr$_2$ + 0.970Gr$_3$ + 0.130Age$_{\geq 65}$ − 0.149Ana)}.

Study Endpoints

The primary endpoint was distant relapse–free survival (DRFS), which was the time from diagnosis until DR. DR was defined as metastasis from the primary tumor at distant organs, excluding contralateral disease and locoregional and ipsilateral recurrences. Death before DR was treated as a censoring event.

Statistical Analysis

Our stepwise primary objectives were to assess whether EPclin had statistically significant prognostic information for 10-year DR in postmenopausal women with breast cancer given either Tamoxifen or Anastrozole monotherapy. If so, we would test if EPclin or EP added statistically significant prognostic information to RS and whether EP/EPclin provided statistically significant additional information to CTS. Secondary analyses included determining the prognostic ability of EP and EPclin in early (0-5 years) and late (>5 years) settings, in patients divided into subgroups by nodal status, and the additional prognostic information provided by tests in multivariable comparisons.

The statistical analysis plan was approved by the Long-term Anastrozole vs Tamoxifen Treatment Effects (LATTE) committee and Sividon before data analysis took place and is described in the Supplementary Methods (available online). All statistical tests were two-sided, and a P value of less than .05 was regarded as statistically significant. All statistical analyses were performed with STATA version 13.1 (College Station, TX).

Results

Sample availability is shown in Figure 1. Values for RS, EP, and EPclin scores were calculated for 928 patients. Demographics of the population are shown in Supplementary Table 1 (available online). A total of 128 DRs was recorded within the 10-year median follow-up period. In node-negative women (n = 680), there were 59 DRs; in node-positive women (n = 248), 69 DRs were recorded.

Univariate Analyses

Results for EP, EPclin, RS, and CTS are presented in Table 1. Both EP and EPclin were highly prognostic across 10 years (LRχ^2:

EP = 49.3; LRχ^2: EPclin = 139.3), with EPclin being statistically significantly more prognostic than EP in all time windows and subgroups, except for node-negative patients in years 0 to 5. Both EP and EPclin provided substantially more information than RS in years 0 to 10 (LRχ^2: RS = 29.1). EP had similar prognostic power to RS in years 0 to 5 in all subgroups. In node-negative patients, both EP and EPclin performance were very similar to that of RS (LRχ^2: EP = 15.5; LRχ^2: EPclin = 17.0; LRχ^2: RS = 18.7). In contrast, in node-positive patients, EPclin outperformed RS. EP and EPclin were superior to RS in years 5 to 10, where RS was particularly weak regardless of nodal status (LRχ^2: EP = 23.6; LRχ^2: EPclin = 59.3; LRχ^2: RS = 5.6).

Figure 2 shows the DR rate over 10 years for each of EP, EPclin, and RS for the overall population when divided into tertiles of their respective scores. The hazard ratio (HR) for the comparison between the lowest and highest tertiles of each score was 4.72 (95% confidence interval [CI] = 2.78 to 8.02), 18.01 (95% CI = 7.87 to 41.19), and 2.41 (95% CI = 1.59 to 3.64),
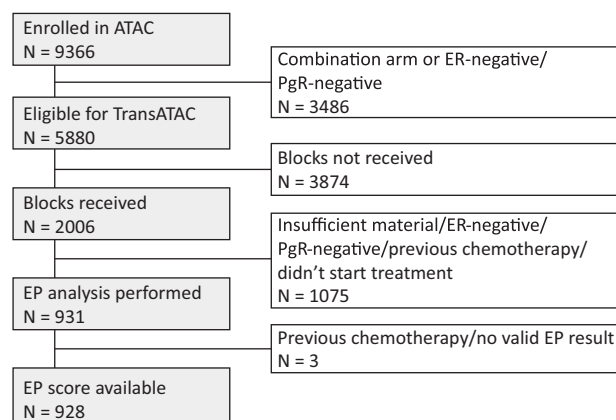
respectively. For EPclin, the lowest tertile had a DR rate of only 2.1% (95% CI = 1.0 to 4.7) at 10 years while the highest tertile had a DR rate of 31.5% (95% CI = 26.4 to 37.4). Similar plots of EP, EPclin, and RS score tertiles for the separate node-negative and node-positive populations are shown in Supplementary Figures 2 and 3 (available online). EPclin identified a third of patients in the node-negative population, in which only one of 227 patients had a DR over 10 years, corresponding to a 10-year relapse rate of 0.5% (95% CI = 0.1 to 3.4). For EP and RS, the equivalent tertiles had DRs of 1.5% (95% CI = 0.5 to 4.7) and 7% (95% CI = 4.2 to 11.5), respectively, over the same time period.

Supplementary Figure 4 (available online) shows the continuous relationship between EPclin score and the estimated 10-year DR rate in TransATAC according to the proportion of the scores contributed by each nodal group. In this cohort, about half of patients (52.6%) with EPclin scores of 3.3 or higher (high-risk) were node-positive; only 8.6% of patients with scores of less than 3.3 were node-positive.

Figure 1. CONSORT diagram of the availability of samples for analysis from the Arimidex, Tamoxifen, Alone or in Combination trial. ATAC = Arimidex, Tamoxifen, Alone or in Combination; ER = estrogen receptor; PgR = progesterone receptor.

Multivariable Analyses

Multivariable comparisons are shown in Table 1. Both EP and EPclin provided statistically significant prognostic value when added to the RS across 10 years (LRχ^2: RS = 29.1; ΔLRχ^2: EP+RS vs RS = 20.2; ΔLRχ^2: EPclin+RS vs RS = 113.8) (Table 1). For EP, this was because of its additional information beyond RS in five to 10 years only. EPclin added statistically significant prognostic information to RS both before and beyond five years, except in the node-negative subgroup of patients in years 0 to 5.

For the overall population, statistically significant prognostic information beyond that of the CTS was provided in years 0 to 10 by EP, EPclin, and RS; however, it was greater for EP and EPclin than for RS. Similar results were observed within node-negative and -positive subgroups (Table 1). The better performance of EP and EPclin in years 0 through 10 was because of its greater prognostic value in years 5 to 10, where RS added no statistically significant prognostic information to CTS (LRχ^2: CTS = 64.7; ΔLRχ^2: EP+CTS vs CTS = 9.8; ΔLRχ^2: EPclin+CTS vs CTS = 9.9; ΔLRχ^2: RS+CTS vs CTS = 2.3).

Table 1. Likelihood (χ^2) for distant recurrence for all prognostic scores in all patients and subgroups*

Patient group	No. of patients	No. of DRs	EPclin		EP		RS		EPclin + RS vs RS†		EP + RS vs RS†		CTS		EPclin + CTS vs CTS†		EP + CTS vs CTS†		RS + CTS vs CTS†	
			LRχ^2	P	LRχ^2	P	LRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P	LRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P
All patients																				
0–10 y	928	128	139.3	<.001	49.3	<.001	29.1	<.001	113.8	<.001	20.2	<.001	149.8	<.001	20.3	<.001	16.4	<.001	12.8	<.001
0–5 y	928	61	80.0	<.001	25.7	<.001	26.1	<.001	54.0	<.001	3.1	.08	85.0	<.001	10.5	.001	6.9	.009	11.8	<.001
5–10 y	820	67	59.3	<.001	23.6	<.001	5.6	.02	59.6	<.001	21.6	<.001	64.7	<.001	9.9	.002	9.8	.002	2.3	.13
Node-negative patients																				
0–10 y	680	59	39.7	<.001	30.8	<.001	21.3	<.001	18.3	<.001	9.7	.002	35.6	<.001	12.5	<.001	11.9	<.001	8.4	.004
0–5 y	680	24	17.0	<.001	15.5	<.001	18.7	<.001	1.6	.2	0.7	.4	19.0	<.001	3.6	.06	5.2	.02	8.1	.004
5–10 y	623	35	22.7	<.001	15.5	<.001	4.8	.03	20.9	<.001	12.4	<.001	16.9	<.001	9.0	.003	6.6	.01	1.4	.24
Node-positive patients																				
0–10 y	248	69	48.3	<.001	14.5	<.001	8.0	.005	44.8	<.001	6.5	.01	61.6	<.001	8.3	.004	5.4	.02	4.1	.04
0–5 y	248	37	32.2	<.001	7.9	.005	8.0	.005	25.9	<.001	0.9	.33	35.2	<.001	6.4	.01	2.3	.13	3.7	.05
5–10 y	197	32	16.1	<.001	6.6	.01	1.0	.32	18.3	<.001	7.1	.008	26.4	<.001	2.3	.13	3.4	.06	0.7	.39

*Both univariate and multivariable analyses are presented for years 0 to 10, years 0 to 5, and years 5 to 10 separately. Likelihood ratio test based on Cox proportional hazard models for univariate and multivariable analyses. Differences in likelihood ratio values (ΔLRχ^2) were used. CTS = clinical treatment score; DR = distant relapse; EP = EndoPredict; LR = likelihood ratio; RS = recurrence score.

†Denotes multivariable comparisons; eg, the EPclin + RS vs RS comparison assesses the extra prognostic information that EPclin contributes when combined with the RS. All statistical tests were two-sided. All scores are continuous variables.
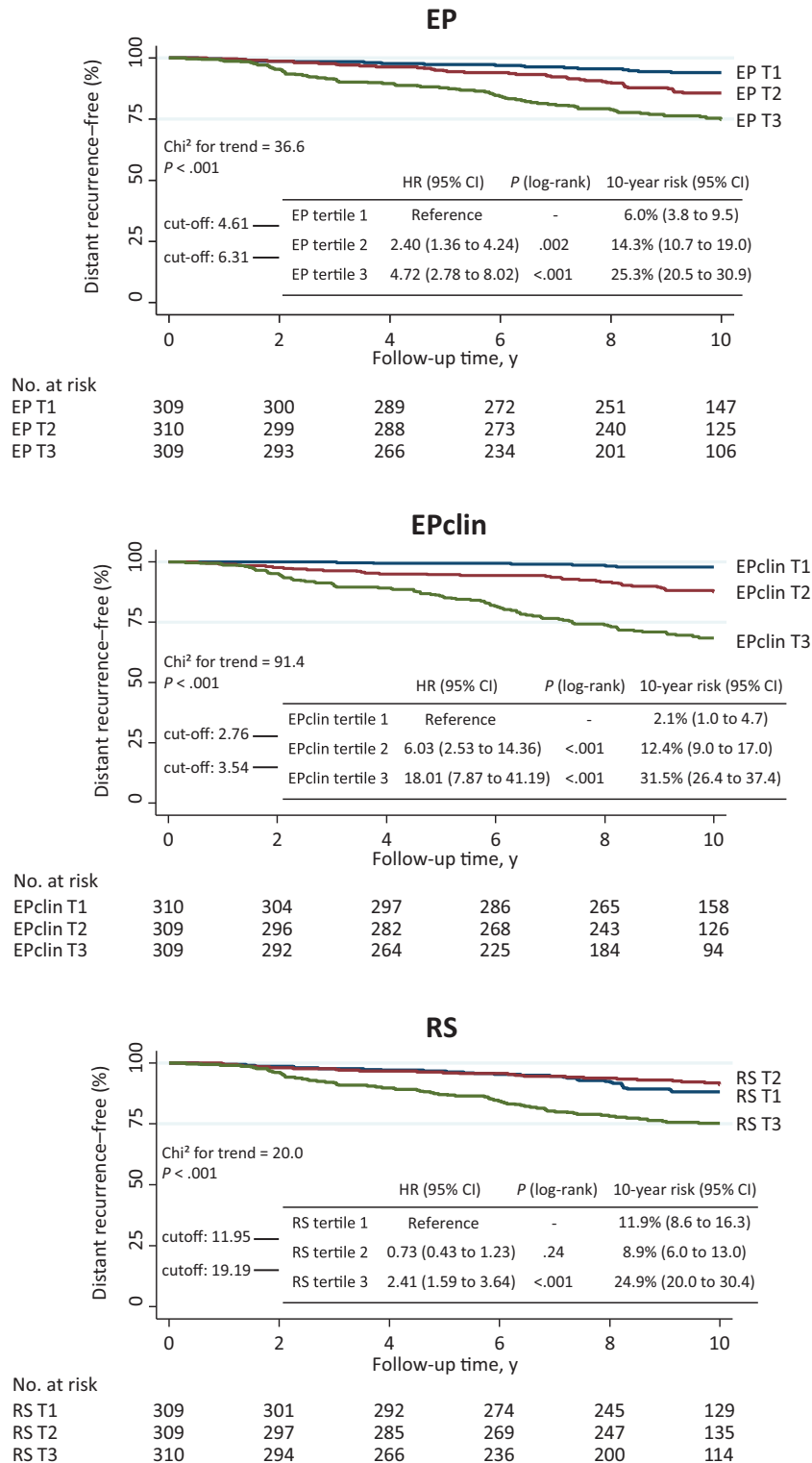
Figure 2. Kaplan-Meier estimates for 10-year distant recurrence according to EP, EPclin, and recurrence score, split into tertiles in all patients. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.

Risk Stratification

For RS, the percentage of patients recurring over 10 years was 5.3% (95% CI = 3.5 to 8.2), 14.3% (95% CI = 9.8 to 20.6), and 25.1%

(95% CI = 15.8 to 38.3) for the low-, intermediate-, and high-risk groups in node-negative patients and 25.1% (95% CI = 18.2 to 33.9), 34.8% (95% CI = 24.9 to 47.2), and 48.6% (95% CI = 31.4 to 69.2) for the node-positive group (Supplementary Figure 5,

available online). These are similar to rates observed over years 0 through 9 in 1178 TransATAC patients in our earlier report of RS' performance (11). To compare directly the recurrence rates in these categories with the low-/high-risk categories of EP and EPclin, we pooled the RS intermediate- and high-risk groups to create an RS non-low-risk group. More patients were stratified to the low-risk group by RS and EPclin than by EP (573 vs 546 vs 386 corresponding to 61.7%, 58.8%, and 41.6% of the cohort). The hazard ratio between the high-/non-low- vs low-risk groups was marginally greater for EP (HR = 2.98, 95% CI = 1.94 to 4.58, $P < .001$) than for RS (HR = 2.73, 95% CI = 1.91 to 3.89, $P < .001$) and substantially greater for EPclin (HR = 5.99, 95% CI = 3.94 to 9.11, $P < .001$) (Figure 3).

EPclin's superior ability to classify patients as low risk was further demonstrated by the similar number of patients classified as low risk by RS coupled with a substantially lower 10-year recurrence rate (EPclin: 5.8%, 95% CI = 4.0 to 8.3; RS: 10.1%, 95% CI = 7.7 to 13.1) (Figure 3). A greater absolute separation of the

DR rate was found between the risk groups for EPclin (23.0%) than for RS (13.4%). EPclin performed particularly well at stratifying node-positive patients where absolute separation at 10 years for DR rate was 31.9% compared with the 14.1% in node-negative patients (Supplementary Figures 6 and 7, available online).

For most cases, EPclin and RS categorization of risk agreed; however, 117 (12.6%) cases were EPclin low/RS non-low and 144 (15.5%) were EPclin high/RS low (kappa = 0.41, $P < .001$). Classification by EPclin aligned more closely with the observed risks: Pairwise comparison of EPclin high/RS low vs EPclin low/RS non-low (HR = 2.75, 95% CI = 1.39 to 5.44, $P = .002$) (Figure 4). The Net Reclassification Index (NRI) for EPclin vs RS was 17.5% ($P < .001$). In recurrent cases, the EPclin upgraded three times more cases into high-risk groups than the RS (McNemar's odds ratio = 3.00, 95% CI = 1.16 to 7.89, $P = .01$) whereas for noncases upgrading/downgrading was similar for these two scores.
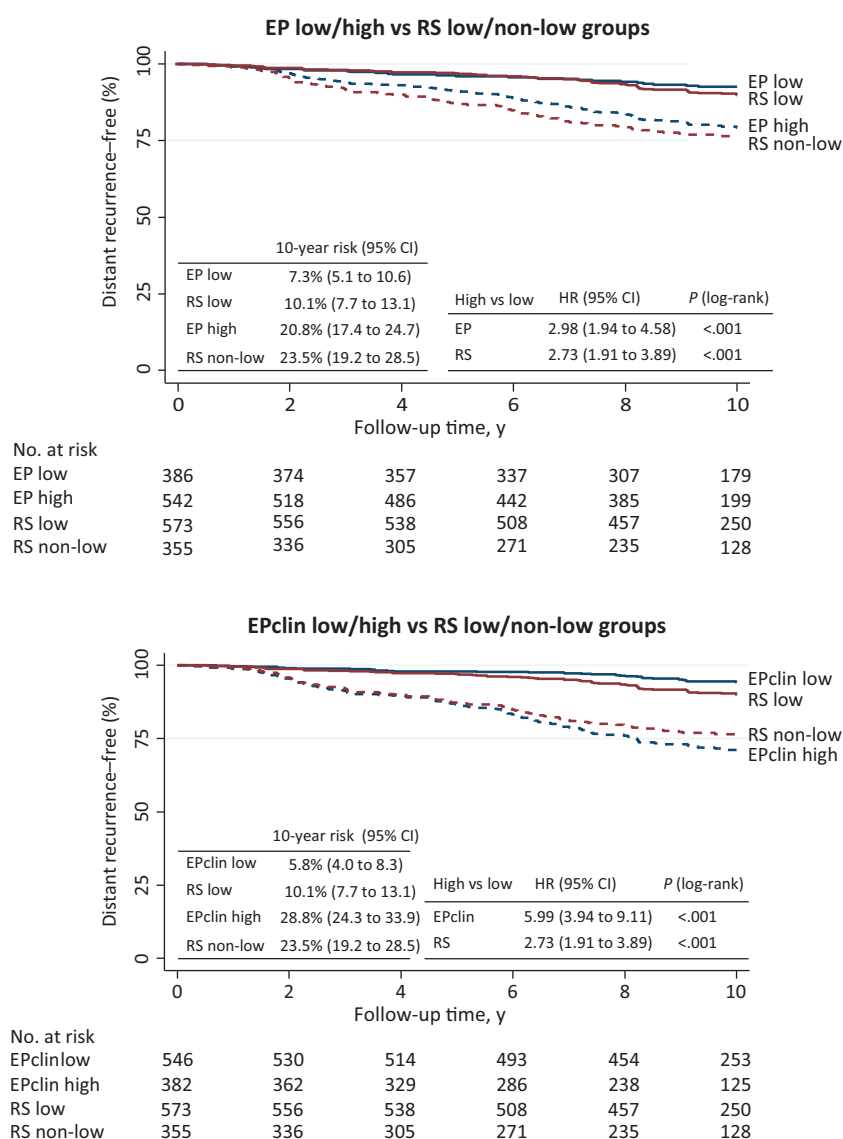

EP low/high vs RS low/non-low groups

	10-year risk (95% CI)	High vs low	HR (95% CI)	P (log-rank)
EP low	7.3% (5.1 to 10.6)			
RS low	10.1% (7.7 to 13.1)	EP	2.98 (1.94 to 4.58)	<.001
EP high	20.8% (17.4 to 24.7)	RS	2.73 (1.91 to 3.89)	<.001
RS non-low	23.5% (19.2 to 28.5)			

No. at risk

	0	2	4	6	8	10
EP low	386	374	357	337	307	179
EP high	542	518	486	442	385	199
RS low	573	556	538	508	457	250
RS non-low	355	336	305	271	235	128

EPclin low/high vs RS low/non-low groups

	10-year risk (95% CI)	High vs low	HR (95% CI)	P (log-rank)
EPclin low	5.8% (4.0 to 8.3)			
RS low	10.1% (7.7 to 13.1)	EPclin	5.99 (3.94 to 9.11)	<.001
EPclin high	28.8% (24.3 to 33.9)	RS	2.73 (1.91 to 3.89)	<.001
RS non-low	23.5% (19.2 to 28.5)			

No. at risk

	0	2	4	6	8	10
EPclin low	546	530	514	493	454	253
EPclin high	382	362	329	286	238	125
RS low	573	556	538	508	457	250
RS non-low	355	336	305	271	235	128

Figure 3. Kaplan-Meier plots for 10-year distant recurrence according to EP, EPclin, and recurrence score in all patients, stratified by cut-offs used for clinical decision-making. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.
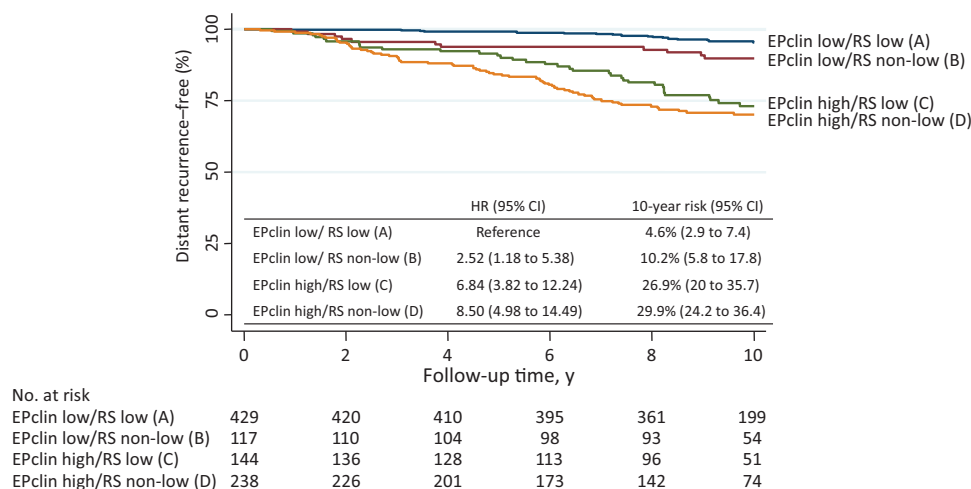
Figure 4. Kaplan-Meier plot of risk groups classified by EPclin and recurrence score for 10-year distant recurrence in all patients. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.

Discussion

In this TransATAC population, we found that both EP and EPclin were highly prognostic across the 10 years of follow-up and both scores also identified early and late relapse events. This is in agreement with previous reports in ER-positive, HER2-negative patient cohorts from the ABCSG-6 and -8 trials (8,10). Moreover, EP and EPclin were prognostic in all assessed subgroups.

We also compared the prognostic information provided by EP and EPclin with that of the widely used Oncotype DX RS. This study is the first direct comparison of the clinical performance of EP/EPclin with RS. EPclin, as opposed to RS, includes information from clinical factors, making it more clinically useful but also making fair comparisons with RS complicated. Therefore, as well as direct comparisons, we conducted analyses to determine how much information was added by the respective scores to CTS.

We found that EP was similar to RS in years 0 to 5 but was superior in years 5 to 10. EPclin markedly outperformed RS across the 10-year follow-up period and also in all additional univariate analyses, except in node-negative patients in the early time window. These findings suggest that: 1) In years 0 to 5, EPclin predicts recurrence better in the overall population than RS because of the clinical components included in EPclin; and 2) in years 5 to 10, the superior performance of EPclin compared with RS is partly because of the inclusion of clinical variables in EPclin but also because of a molecular component that predicted late recurrences better. The latter is also reinforced by the very similar prognostic value of EP in the early and late follow-up periods, in marked contrast with RS, where performance diminished beyond five years.

EP's overall better performance over RS might be attributed to the differences in the training populations. The EP algorithm was trained on a HER2-negative, mixed node-negative and -positive population, unlike RS, which was optimized on a mixed HER2-negative and -positive, node-negative population. These differences may explain the better prognostic ability of EP, in particular in node-positive patients and in patients at risk of a late relapse.

A previous analysis of EP components in ABCSG-6 and -8 trial samples showed that proliferative genes contributed to early prediction and ER-signaling genes provided prognosis beyond five years (10). Recently, we reported our analysis of RS components in ER+/HER2- TransATAC patients that pointed to the loss of prediction by the ER module as the main reason for its weak performance after five years while the proliferation RS module was prognostic throughout the 10 years (15). The different behavior of the proliferation and ER-associated genes in the two scores may be because of the different identity of genes used and their weighting in the respective algorithms. Further analysis is necessary to understand fully the differing behavior of these prognostic scores.

The integration of nodal status into the EPclin score allows the algorithm to be used in both node-negative and node-positive patients, supported by the observed DR rates in the populations identified as low risk in the respective nodal groups. It was notable, however, that when the algorithm was applied as a continuous variable in the node-negative population it identified one-third of the node-negative population with an extremely low DR of just 0.5% at 10 years. Categorization of a patient in such a low-risk group could be highly reassuring. Our earlier publication showed the differing relationship of RS with risk of DR according to nodal status (11). The current data emphasize that RS should not be used in node-positive patients to estimate recurrence risk without appropriate calibration of the relationship of RS with DR for such patients.

Of note, a recent report from the TAILORx trial described the very low risk of DR rate in patients with RS of 10 or lower (16); this was, however, only over the first five years of follow-up. Generally, patients in a low-risk group would not be recommended to receive chemotherapy treatment because of their perceived low recurrence risk. Previously, ABCSG-6 and -8 observed 10-year DR rates by EPclin classification that were 4% in the low groups for both studies, 28% and 22% in the high-risk groups for the two trials, respectively (8). Our analysis showed similar 10-year recurrence rates at 5.8% and 28.8% in the low and high EPclin groups of TransATAC, respectively, in contrast with 10.1% and 23.5% observed for the RS low and non-low groups. An NRI favorable to EPclin indicated that EPclin classification aligned better with observed risk than RS and therefore provided superior risk stratification when compared with RS. If results are available from both assays yet disagree with one

another, more weight should be assigned to the EPclin risk estimate.

Previously, the importance of integrating clinical and molecular variables to create a more accurate prognostic index for RS (17) and for IHC4 (6) was reported. The superiority of EPclin over EP that resulted from such integration is probably best demonstrated by the DR rates in the highest and lowest tertiles of the respective scores.

Recently, GHI began providing an online Recurrence Score Pathology-Clinical (RSPC) calculator for use in node-negative patients that combines RS with clinicopathological variables including age, tumor size, grade, and planned adjuvant hormonal therapy. Tang et al. reported a greater separation of low- and high-risk patients and reduced number of patients in the intermediate-risk group when classified by RSPC (17). Nevertheless, GHI recommends that RSPC should only be used as an "educational tool" together with RS result to enhance the understanding of the score in the assessment of DR risk (18). It should be noted that while integration of clinicopathological factors with molecular features greatly enhances the prognostic power of risk assessments, this has not been shown to increase predictive information regarding chemotherapy benefit (17).

Our study has strengths and limitations. Strengths include the large patient cohort with long-term follow-up from a well-documented clinical trial and well-characterized set of samples. For this comparison, the same batch of RNA was used, reducing intrasample variation. EP data was obtained by personnel blinded to the clinical data and the results of previous assays in TransATAC. Nonetheless, limitations included the low event rate in ER-positive, endocrine-treated patients and CTS trained on the TransATAC cohort, slightly overestimating its performance compared with what we would expect in independent validation cohorts. Generalizability of the results may be limited by the analysis of patients from the United Kingdom only. Additionally, an unintended sample selection bias might have occurred as the assessment of samples for EP could only be performed where larger amounts of residual RNA were available. Although this might have been expected to relate to fewer smaller tumors in this study than in our earlier report on RS (11), the proportion of tumors 2 cm or smaller was identical at 67% in both. By necessarily restricting the performance of the scores to patients not receiving chemotherapy, this cohort is likely to be biased toward lower risk in the spectrum of ER-positive patients. Lastly, in a number of cases, multiple comparisons were made and caution is needed in interpreting those results. However, for our primary sequential objectives, all tests and comparisons were highly statistically significant at the 1% level, even after correction for multiple comparisons (nominal P value < .001). For subgroup analyses, heterogeneity tests are more important (19) and no heterogeneity was observed between subgroups.

In summary, this study has confirmed the independent prognostic ability of EP and EPclin in postmenopausal women with ER+/HER2- primary disease. EPclin provided more prognostic information than RS partly because of its integration with node and tumor size information but also because of a superior molecular component able to predict late events better than RS. Our data highlights the importance of the inclusion of clinicopathological factors in overall estimate of risk assessments.

Funding

Notes

References

1. Dowsett M, Forbes JF, Bradley R, et al. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet.* 2015;386(10001):1341–1352.
2. Dowsett M, Goldhirsch A, Hayes DF, et al. International Web-based consultation on priorities for translational breast cancer research. *Breast Cancer Res.* 2007;9(6):R81.
3. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004;351(27):2817–2826.
4. 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.* 2010;16(21):5222–5232.
5. Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer.* 2011;104(11):1762–1769.
6. Cuzick J, Dowsett M, Pineda S, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol.* 2011;29(32):4273–4278.
7. Sestak I, Dowsett M, Zabaglo L, et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst.* 2013;105(19):1504–1511.
8. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res.* 2011;17(18):6012–6020.
9. Denkert C, Kronenwett R, Schlake W, et al. Decentral gene expression analysis for ER+/Her2- breast cancer: results of a proficiency testing program for the EndoPredict assay. *Virchows Arch.* 2012;460(3):251–259.
10. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer.* 2013;109(12):2959–2964.
11. Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol.* 2010;28(11):1829–1834.
12. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol.* 2013;31(22):2783–2790.
13. Sgroi DC, Sestak I, Cuzick J, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol.* 2013;14(11):1067–1076.
14. 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.* 2010;11(12):1135–1141.
15. Dowsett M, Sestak I, Buus R, et al. Estrogen Receptor Expression in 21-Gene Recurrence Score Predicts Increased Late Recurrence for Estrogen-Positive/HER2-Negative Breast Cancer. *Clin Cancer Res.* 2015;21(12):2763–2770.
16. Sparano JA, Gray RJ, Makower DF, et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med.* 2015;373(21):2005–2014.
17. Tang G, Cuzick J, Costantino JP, et al. Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol.* 2011;29(33):4365–4372.
18. Genomic Health Inc. Oncotype DX® tools for Genomic Education. https://breast-cancer.oncotypedx.com/en-US/Professional-Invasive/Resources/ODX-Tools-RSPC.aspx. Accessed December 23, 2015.
19. Cuzick J. Forest plots and the interpretation of subgroups. *Lancet.* 2005;365(9467):1308.