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

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 $\chi^2$ 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 ($LR^2$: EP = 49.3; EPclin = 139.3; $LR^2$: 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 ($LR^2$: EP + RS vs RS = 20.2; $LR^2$: 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.
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 

\[ \text{CTS} = 100 \times 0.417N_{1-3} + 1.566N_{4} + 0.930(0.497T_{1-2} + 0.882T_{2-3} + 1.838T_{3} + 0.559G_{2} + 0.970G_{3} + 0.130A_{E} + 0.457A_{S} - 0.149A_{N}) \]

**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 (LH, p<
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 (LR2: EP = 29.1; ALR2: EP + RS vs RS = 20.2; ALR2: 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 (LR2: CTS = 64.7; ALR2: EP + CTS vs CTS = 9.8; ALR2: EPclin + CTS vs CTS = 9.9; ALR2: RS + CTS vs CTS = 2.3).

Table 1. Likelihood (v2) 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† |
|---------------|----------------|-----------|--------|----|----|-------------------|----------------|-----|---------------------|-----------------|------------------|
| All patients  |                |           |        |    |    |                    |                 |     |                     |                 |                  |
| 0–10 y        | 928            | 128       | 139.3  | .001| 49.3| .001               | 29.1            | .001| 113.8               | .001            | 20.2             |
| 0–5 y         | 928            | 61        | 80.0   | .001| 25.7| .001               | 26.1            | .001| 54.0                | .001            | 3.1              |
| 5–10 y        | 623            | 35        | 59.3   | .001| 23.6| .001               | 56.6            | .002| 9.8                 | .002            | 2.3              |
| Node-negative patients |           |             |        |    |    |                    |                 |     |                     |                 |                  |
| 0–10 y        | 680            | 59        | 39.7   | .001| 30.8| .001               | 21.3            | .001| 18.3                | .001            | 9.7              |
| 0–5 y         | 680            | 24        | 17.0   | .001| 15.5| .001               | 18.7            | .001| 1.6                 | .001            | 0.7              |
| 5–10 y        | 523            | 35        | 22.7   | .001| 15.5| .001               | 4.8             | .03 | 20.9                | .001            | 12.4             |
| Node-positive patients |           |             |        |    |    |                    |                 |     |                     |                 |                  |
| 0–10 y        | 248            | 69        | 48.3   | .001| 14.5| .001               | 8.0             | .005| 44.8                | .001            | 6.5              |
| 0–5 y         | 248            | 37        | 32.2   | .001| 7.9 | .005               | 8.0             | .005| 25.9                | .001            | 0.9              |
| 5–10 y        | 197            | 32        | 16.1   | .001| 6.6 | .001               | 1.0             | .32 | 18.3                | .001            | 7.1              |

*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 (ALR2) 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.
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, Table 1).
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.

| 10-year risk (95% CI) | EP low/high vs RS low/non-low groups |
|-----------------------|----------------------------------|
| 7.3% (5.1 to 10.0)    | EP low                            |
| 10.1% (7.7 to 13.1)   | RS low                            |
| 20.8% (17.4 to 24.7)  | EP high                           |
| 23.5% (19.2 to 28.5)  | RS non-low                        |

| Follow-up time, y |
|-------------------|
| 0                 |
| 2                 |
| 4                 |
| 6                 |
| 8                 |
| 10                |

| No. at risk |
|-------------|
| 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         |

| 10-year risk (95% CI) | EPclin low/high vs RS low/non-low groups |
|-----------------------|-----------------------------------------|
| 5.8% (4.0 to 8.3)     | EPclin low                              |
| 10.1% (7.7 to 13.1)   | RS low                                  |
| 28.8% (24.3 to 33.9)  | EPclin high                             |
| 23.5% (19.2 to 28.5)  | RS non-low                              |

| Follow-up time, y |
|-------------------|
| 0                 |
| 2                 |
| 4                 |
| 6                 |
| 8                 |
| 10                |

| No. at risk |
|-------------|
| 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.
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 relationships 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
Breast Cancer: Now granted award to MD (CTR-Q4-Y1) and the Cancer Research UK grant awarded to JC (CS69/A16891).

Notes
The study sponsor had no role in design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

This project was developed during discussions at the Biedenkopf meeting October 31, 2013, through November 2, 2013, supported by the BANSS foundation, Biedenkopf/Lahn, Germany.

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(10002):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(2):5222-5232.
5. Jerevea P, 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(12):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. 2012;30(12):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, Wales C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive post-menopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. J Clin Oncol. 2010;28(11):1849-1854.
12. Dowsett M, Sestak I, Lopez Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotyp 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(13):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, Leung S, 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(9477):1308.

Funding
This work was supported by the Royal Marsden National Institutes of Health Biomedical Research Centre and the Institute of Cancer Research, London. This work was supported by the Royal Marsden National Institutes of Health Biomedical Research Centre and the Institute of Cancer Research, London.