A Comparison of the Performance of EndoPredict Clinical and NHS PREDICT in 120 Patients Treated for ER-positive Breast Cancer

KINAN MOKBEL1,2, UMAR WAZIR1, HIBA EL HAGE CHEHADE1, AISLING MANSON1, CHRISTINA CHOY1, VICTORIA MOYE2 and KEFAH MOKBEL1

1The London Breast Institute, Princess Grace Hospital, London, U.K.; 2University of Exeter Medical School, Exeter, U.K.

Abstract. Background: Computational algorithms, such as NHS PREDICT, have been developed using cancer registry data to guide decisions regarding adjuvant chemotherapy. They are limited by biases of the underlying data. Recent breakthroughs in molecular biology have aided the development of genomic assays which provide superior clinical information. In this study, we compared the performance in risk stratification of EndoPredict Clinical (EPClin, a composite of clinical data and EndoPredict) and PREDICT in a cohort of patients with breast cancer considered potential candidates for chemotherapy by the clinicians. Materials and Methods: One hundred and twenty patients with biopsy-proven oestrogen receptor positive (ER+)/human epidermal growth factor receptor 2-negative (HER2−) breast cancer who underwent surgery were included. EPClin and PREDICT were determined for every tumour, and the results were compared. Results: Using EPClin scores performed on 120 tumours, the cohort was stratified into low-(n=60) and high-risk (n=60) groups leading to 50% reduction in total chemotherapy prescriptions. PREDICT differentiated the patients into low- (n=45), intermediate- (n=33), and high-risk groups (n=42). Discordance between scores was demonstrated for 50 (41.66%) tumours. Nine (20%) out of 45 patients with low PREDICT scores had high EPClin scores and would otherwise not have received chemotherapy if the NHS PREDICT tool had been used alone. Eight (19%) out of 42 patients at high risk by PREDICT were reclassified as being at low risk by EPClin and avoided adjuvant chemotherapy. The sensitivity, specificity, positive predictive value and negative predictive value for NHS PREDICT to predict the potential need for chemotherapy as determined by EPClin were 85%, 51%, 68% and 80%, respectively. Conclusion: To our knowledge, this is the first clinical study to compare EPClin and PREDICT. The data indicate that computational algorithms such as NHS PREDICT may not accurately predict the need for chemotherapy leading to overtreatment, undertreatment or uncertainty and anxiety in a significant proportion of patients. This underscores the importance of more personalized prognostic tools.

It is known that surgical resection for early breast cancer renders patients free of overt disease. Nonetheless, adjuvant therapy is also given to eradicate any residual microscopic disease, thus preventing recurrence (1). However, 65% of newly diagnosed breast cancer cases are lymph node (LN) negative and oestrogen receptor (ER) positive with a 5-year recurrence risk rate of less than 20% when treated with tamoxifen (2, 3). The addition of adjuvant chemotherapy would reduce the risk of recurrence by only an additional 2-10% (4). Therefore, it is crucial to continue to debate the current guidelines, which recommend chemotherapy for the majority of such cases, potentially overtreating the patients (5).

The addition of chemotherapy has traditionally been a decision based on clinical and pathological features (6). These factors have been incorporated into prognostic models, prognostic indices, predictive tools and risk scores (7-9). However, such decisions tend to be subjective and are seldom consistent (10-15). Therefore, computational and mathematical algorithms were subsequently developed using cancer registry data (16). NHS PREDICT is an example of these freely available, web-based prognostication tools. It provides 5- and 10-year survival estimates, as well as absolute treatment benefit (17, 18). Even though such tools have gained popularity in the UK and seem to be especially attractive for resource-constrained countries (19-21), its use may need to be tempered in view of the well-documented limitations of these algorithms, as well as the inherent biases of the underlying databases (1, 20).

Correspondence to: Kinan Mokbel, The London Breast Institute, Princess Grace Hospital, 45 Nottingham Place, London W1U 5NY, U.K. E-mail: kinamokbel@hotmail.com

Key Words: Genomic assays, prognosis, adjuvant therapy, breast cancer.
The recent breakthrough in molecular medicine which has characterized breast cancer as a heterogenous group of neoplastic processes has become the basis for novel prognostication molecular assays (22, 23). Popular examples of such multigene assays include MammaPrint (24), OncotypeDx (25), and EndoPredict (26).

EndoPredict is an 8-gene quantitative real-time polymerase chain reaction (q-RT PCR) assay performed on formalin-fixed breast tissue. There is level I evidence that suggests that EndoPredict is an independent prognostic parameter in ER-positive, human epidermal growth factor receptor 2 negative (HER2)-negative breast cancer (27). EndoPredict Clinical (EPClin) is a composite tool in which the results of EndoPredict are combined with clinical data to determine the risk of recurrence. Studies have shown that it provides superior clinical information (26, 27).

To our knowledge to date, the association between EPClin score and NSH PREDICT has not been tested in previous studies. Therefore, in our study, we analysed the correlation between EPClin and PREDICT scores in a cohort of 120 ER-positive HER2-negative breast cancer cases to compare the performance of both tools in risk stratification of these patients.

**Materials and Methods**

The study was conducted at the London Breast Institute, and patients were recruited between December 2015 and December 2016. The inclusion criteria were female patients newly diagnosed with ER-positive, HER2-negative, node-positive or -negative primary breast cancer who underwent breast-conserving surgery or mastectomy (with or without reconstruction) and sentinel lymph node biopsy or axillary lymph node clearance. Exclusion criteria were inoperable disease.

Pathological analysis was conducted at the Pathology Department of the London Breast Institute. EndoPredict was performed on paraffin-embedded tumor samples, and the EPClin score was calculated. The patients were stratified into low- (EPClin <3.3) and high-risk (EPClin ≥3.3) groups.

The patients recruited were also assessed using NHS PREDICT. As per its website (www.predict.nhs.uk), NHS PREDICT required the patients’ age, detection mode, tumour size, tumour grade, nodal involvement, receptor status (ER and HER2) and Ki67 status. We queried the absolute benefit of third-generation chemotherapy on survival for individual patients. Patients were thus classified into low- (score<3%), intermediate- (score 3-5%), and high-risk (score >5%) groups. Statistical analysis was performed using the MedCalc biomedical statistical software, version 16.4.3 (MedCalc software bvba, Ostend, Belgium).

**Results**

A total of 120 patients underwent mastectomy or breast-conserving surgery for ER-positive, HER2-negative breast cancer at The London Breast Institute The Princess Grace Hospital. The mean age was 55 years (median age=54 years, range=31- 77 years). The EndoPredict and the EPClin scores were determined in the 120 tumours from 120 patients. Based on the classification described by Filipit et al. (26), patients were divided into two risk groups: low risk (EPClin<3.3) and high risk (EPClin≥3.3). 60 (50%) out of the 120 patients were classified as being at high risk and 60 (50%) at low risk (range of scores=1.9-5.7). The latter group was advised to avoid chemotherapy. The median EPClin score for the low-risk group was 2.7 and for the high-risk group was 3.8. The online tool PREDICT was then used on the same patients to assess the cancer recurrence risk and the need for additional adjuvant chemotherapy. A total of 45, 33, and 42 patients fell into the low-, intermediate-, and high-risk categories with median scores of 1.9, 3.8 and 7.35, respectively, and an overall range of 1-19.1%. The EPClin analysis was performed in three patients with more than three lymph nodes outside the standard protocol. These three patients were very keen to avoid chemotherapy if possible despite significant lymph node involvement and initial advice to receive adjuvant chemotherapy on the basis of clinical assessment. The molecular score was calculated on the basis of three nodes being involved and was found to be elevated in all three cases. One patient (four affected nodes) decided to have adjuvant dose dense chemotherapy, one patient (6 affected nodes) opted for a shorter course of chemotherapy and the third patient (8 affected nodes) declined and opted to have dual endocrine therapy in the form of Zoladex and Letrozole. Two of these patients had infiltrating lobular carcinoma which is recognised to be less responsive to adjuvant chemotherapy. The results and clinicopathological characteristics are demonstrated in Tables I-VI.

**Table I. Patient clinicopathological features (n=120).**

<table>
<thead>
<tr>
<th>Clinicopathological feature</th>
<th>Number of patients</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>120</td>
<td>100%</td>
</tr>
<tr>
<td>Tumour size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1a (1-5 mm)</td>
<td>1</td>
<td>0.83%</td>
</tr>
<tr>
<td>T1b (5-10 mm)</td>
<td>15</td>
<td>12.50%</td>
</tr>
<tr>
<td>T1c (10-20 mm)</td>
<td>51</td>
<td>42.50%</td>
</tr>
<tr>
<td>T2 (20-50 mm)</td>
<td>45</td>
<td>37.50%</td>
</tr>
<tr>
<td>T3 (&gt;50 mm)</td>
<td>8</td>
<td>6.66%</td>
</tr>
<tr>
<td>Nodal status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>31</td>
<td>25.83%</td>
</tr>
<tr>
<td>Negative</td>
<td>89</td>
<td>74.16%</td>
</tr>
<tr>
<td>Tumour grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>6</td>
<td>5%</td>
</tr>
<tr>
<td>G2</td>
<td>88</td>
<td>73.33%</td>
</tr>
<tr>
<td>G3</td>
<td>26</td>
<td>21.66%</td>
</tr>
</tbody>
</table>

Comparison of the EPClin and NHS PREDICT score stratifications showed discrepant results. PREDICT scores
ranged between 1% and 19.1%, while EPClin scores ranged between 1.9 and 5.7. Concordance between the two scores was seen in 70 cases. Fifty cases (41.66%) demonstrated discordance where PREDICT scores were high in eight cases that had low EPClin scores. In this study, the worst PREDICT score was 11.1%, with a corresponding EPClin score of 3.26. On the other hand, nine cases with low PREDICT scores demonstrated high EPClin scores, with the lowest PREDICT score being 1.3% and a corresponding EPClin of 4. The 33 cases with intermediate PREDICT scores were classified as being at low risk in 16 and high risk in 17 cases by EPClin score.

Further correlation analysis demonstrated a positive association between EPClin and PREDICT scores (Figure 1). Although this association was statistically significant, the degree of correlation was only 60.25% (Pearson’s correlation coefficient), and this is consistent with our observed concordance rate of 58.3% and confirms that PREDICT cannot accurately predict the molecular profiling scores to be used as an alternative in clinical practice.

The sensitivity, specificity, positive predictive value and negative predictive value for NHS PREDICT to predict the need for chemotherapy as determined by EPClin were 85%, 51%, 68% and 80%, respectively. Figure 2 demonstrates the receiver operating characteristics (ROC) curve. The area under curve (AUC) was 0.799 with a standard error of 0.0403 and 95% confidence interval=0.716 to 0.867. The Z statistic was 7.413 (area=0.5; \( p<0.0001 \)). This indicates that NHS PREDICT can differentiate between high and low EPClin scores; however, its accuracy of predicting EPClin is only fair and is not good enough to replace the latter.

Nine out of 45 (20%) patients with low PREDICT scores were found to have high EPClin scores. These patients would have been deprived of chemotherapy if the NHS PREDICT score was utilised. Furthermore, eight out of 42 (19%) patients with high PREDICT scores were found to have low EPClin scores and these patients would have been overtreated with chemotherapy if the PREDICT score had been utilised. Therefore, NHS PREDICT would have resulted in nine patients being undertreated with chemotherapy, eight patients being overtreated with chemotherapy, and 33 patients being classified as intermediate leading to anxiety and uncertainty regarding the benefit of adjuvant chemotherapy.
Discussion

It is well established that breast cancer is the most common cancer in females. Nonetheless, the increase in incidence has been paralleled with improvement in survival rates over the past two decades (28). This evolution is attributed not solely to the advances in diagnostic measures, but also particularly to the revolution established by precision medicine (29). The traditional histopathological classification of breast cancer into invasive ductal carcinoma and other types (lobular, tubular, medullary, and metastatic carcinoma) (30) has been refined and in some instances replaced by molecular classification (31). Fortunately, breast cancer is one of the few tumour types in which molecular classification has been successfully used to implement targeted therapy, thus improving disease-free survival (32).

The heterogeneity in breast cancer makes it a challenging solid tumour to diagnose and treat (31). Therefore, a ‘one-size-fits all’ treatment is not a valid option anymore (30). After surgery for early breast cancer, the patient can be declared as being ‘clear’ of the disease (1). The physicians are then left to face the difficult decision as to whether any adjuvant therapy is needed to eradicate residual microscopic disease (1, 33). Owing to the fact that more than half of newly diagnosed cancer cases are ER-positive, lymph node-negative, and that adjuvant chemotherapy plays an additional protective role in a small proportion of these patients, the decision regarding chemotherapy should be carefully and individually established to spare the patients the unnecessary side-effects of this treatment (2-5).

Oncologists have depended on clinicopathological features to guide them in their clinical decisions. Features such as tumour size, lymph node status, and lymphovascular invasion are risk factors associated with prognosis. However, characteristics such as histological grade, ER, progesterone receptor (PR), and HER2 status, and proliferative index are associated with both prognosis and sensitivity to different treatment modalities (34). These factors were also incorporated into prognostic indices, such as the Nottingham Prognostic Index, and guidelines developed by the National Comprehensive Cancer Network and the Saint Gallen Consensus in order to guide oncologists in the management of patients with breast cancer (35-37). However, this approach was marred by inconsistency and subjectivity of the decisions (10-15). In particular, factors such as treatment toxicity, performance status, quality of life, patient’s perception of the treatment efficacy would play a pivotal role in clinical decision making, which may not be accounted for in the aforesaid guidelines (38-40).

This justified the efforts to build multivariate prognostic models such as the web-based tools Adjuvant Online! and NHS PREDICT (17, 18, 41) and multigene assays. Adjuvant Online! and PREDICT are widely used free online predictive models that aid with adjuvant chemotherapy decision-making based on data generated by large cancer registries (16, 42).

NHS PREDICT gained wide acceptance in the UK, and has been suggested to be a suitable option in resource-poor settings, where more expensive assays may not be available. However, concerns have been raised regarding the accuracy of this tool. For instance, Wong et al. found that NHS PREDICT overestimates survival in young patients (20). This may be due the limitations of the underlying cohort included in the registry, which would tend to be older than the general population. Similarly, it is unclear how reliable the results of NHS PREDICT would be in more ethnically
diverse areas within the UK, such as London or Birmingham. The tool has been updated (now called PREDICT Plus) (43), and over time further parameters have been added to the model (19, 43-45). Whether these changes will address the concerns regarding generalisation is as yet unclear.

The next step in cancer prognostication has been the development of genomic assays, such as Oncotype Dx and EndoPredict (46). There is significant evidence suggesting that these assays provide superior prognostication compared to other tools. However, in view of the capital required to deploy these tools, the question remains whether there is still a place for online, registry-based tools, in particular with regards to poorer countries. It was to explore this question that this study was performed.

The use of molecular profiling (EPClin) in this study stratified 50% of patients to chemotherapy and extended hormonal therapy, and 50% of patients to hormonal therapy alone. Whilst some discordance between tests could be expected, the degree of discordance observed in this study was significantly high at 41.7%. This is inconsistent with any expectation of equivalence between these tests. Furthermore, EPClin reclassified eight out of 42 (19%) patients, identified as being at high risk by PREDICT, as being at low risk, and these patients therefore avoided adjuvant chemotherapy with its associated toxicity. Contrasting this with the fact that nine out of 45 (20%) patients (low PREDICT and high EPClin scores) would not have received chemotherapy if tested with NHS PREDICT alone, it becomes difficult to accept online registry-based tools as an acceptable substitute for genomic testing, even in resource-poor settings. There is a growing body of evidence that small tumours often metastasise early and more efficiently than larger dense tumours driven by certain molecular and genetic alterations (47). Patients with such tumours are unlikely to be selected for adjuvant therapy by conventional assessment or by online tools, which makes genomic assay not simply the optimal assessment, but in fact a necessity in the current management of breast cancer.

The failure of NHS PREDICT to accurately predict the genomic profiling scores is due to the fact that the NHS PREDICT score is essentially based on the clinical and pathological parameters of the tumour whereas the EPClin score is primarily based on the genes which mediate responsiveness and resistance to endocrine therapy. Therefore, the latter provides a fingerprint of tumour biology in the context of response to endocrine therapy, whereas the former is based on conventional clinical and pathological parameters and on data derived from cancer registries which are recognised for inaccuracy. In addition, as alluded to earlier, tools such as NHS PREDICT may not be reliable in ethnically diverse areas within the UK, let alone elsewhere.

The only previous study which compared genomic profiling versus online tools is the MINDACT trial, which compared Mammaprint to Adjuvant Online! In the course of that study, it was noted that patients who were considered at high risk according to online tools but at low risk by Mammaprint and therefore did not receive chemotherapy had an excellent survival, demonstrating that the latter is more accurate in stratification of patients for chemotherapy or no chemotherapy (48). The trial demonstrated a concordance rate of 68%, which is higher than that observed (58.3%) in our study. This can be explained by the fact that the threshold for stratification as high-risk by Adjuvant Online! was set at a higher level than that for PREDICT and the stratification process was dichotomous without an intermediate group as was the case with PREDICT. The main limitation of our study relates to the assumption that molecular profiling with EPClin is prognostically superior and that we did not report clinical outcome data to assess which test is superior in stratifying patients for chemotherapy or no chemotherapy interventions. However, the assumption of prognostic superiority of EPClin in this patient population is reasonable and based on level I evidence (46).

It has to be emphasised that as a composite score, EPClin has some advantages over assay-only scores, such as Mammaprint and EndoPredict. Combining clinical information with a molecular score would hopefully provide a more robust prognostic tool (46).

Furthermore, an issue common to both genomic assays and online tools is the ambiguity regarding the intermediate-risk group. As would be expected, this is a cause for much patient anxiety. In our study, 33 (27.5%) tumours had an intermediate PREDICT score, a result which might have caused inconsistent clinical decision making between different physicians, as well as distress to already stressed patients. This is a concern which has been noted previously in the literature (49). By stratifying the patients into two risk groups, EPClin offers an elegant solution to this problem leading to an improvement in the emotional and functional wellbeing of patients being considered for adjuvant chemotherapy. The findings of our study are consistent with the conclusions of our recently published systematic literature review (50).

**Conclusion**

To our knowledge, the current study is the first to report on the correlation between EPClin and PREDICT in a cohort of referred patients with breast cancer. The data indicate that commonly used online prognostication tools, such as NHS PREDICT, may not accurately predict the survival rate, or indeed the need for adjuvant chemotherapy. The discrepancies observed between the two scores emphasize the importance of a more personalised prognostic tool, especially in this modern era of targeted therapy.
Acknowledgements

This review was funded by grants from the Breast Cancer Hope Foundation (London, UK).

Conflicts of Interest

The Author Kefah Mokbel provides medical advisory and consultancy to Myriad Genetics.

References