Predicting Oncotype DX Recurrence Scores using locally available immunohistochemical markers:

Experience in a District General Hospital

Running Title: Ki67 & PR can be used to predict Oncotype scores

List of authors:

Katherine Humphris (Department of Pathology, Harrogate District Hospital, Lancaster Park Road, Harrogate, HG2 7SX)

John Stephenson (School of Human and Health Sciences, University of Huddersfield, Huddersfield HD1 3DH)

Vidya Kumaraswamy (Department of Pathology, Calderdale and Huddersfield NHS Trust Salterhebble Halifax HX3 0PW)

The authors confirm that there is no conflict of interest.

Word count 2028 excluding ancillary data

Abstract:

Abstract:

Aims: Oncotype DX testing is a reliable widely used gene assay to determine whether chemotherapy is of additional value in ER positive HER2 negative node negative breast cancer, but the high cost of the test can be a barrier for optimal therapy guidance for a substantial proportion of eligible patients around the world. We aimed to determine whether the commonly available immunohistochemical markers Ki67 and PR can predict Oncotype DX RS scores in a district general hospital setting.

Methods: The Oncotype DX RS scores from 58 tumours were regressed against corrected Ki67 values in a simple regression model, and against ER- and PR-derived indices and corrected Ki67 values in a multiple model. Model portability was assessed using leave-one-out cross-validation (LOOCV).

Results: All terms in both regression models were significantly associated with RS scores at the 5% significance level (p<0.001 for all parameters). The multiple model was a better fit to the data (adjusted-R^2=0.784), and performed better under LOOCV (root mean square error = 7.26), suggesting good predictive capability and model portability.

Conclusion: Locally available, cheaper alternatives to multigene assays to determine therapy in ER positive HER2 negative patients is of benefit both from patient management and financial perspectives. A model has been derived with high capability to predict RS scores accurately from linear combinations of predictive biomarkers in a DGH setting which should show good properties when applied to other samples.

Keywords: Pathology (Surgical), Chemotherapy/Cancer/Regional Perfusion, Statistics

Introduction:

Breast cancer is the most common type of cancer in women in the UK, accounting for approximately 15% of all newly diagnosed cancers. Breast cancer therapy is increasingly more tailored to tumour characteristics, with multigene assays being used to determine whether chemotherapy is of additional value in ER positive HER2 negative node negative breast cancer.

The Oncotype DX test uses reverse transcription polymerase chain reaction (RT-PCR) to quantify the expression of 21 genes (16 cancer-related genes and 5 reference genes) from ribonucleic acid (RNA) extracted from breast tissue samples. Oncotype DX has been approved by within the UK by the National Institute for Health and Care Excellence (NICE), and has also been incorporated into the American Joint Committee on Cancer (AJCC), 8th Edition, as a factor to assign Pathological Prognostic Stage Group. The result of the Oncotype DX assay is expressed as a Recurrence Score (RS). The RS provides both predictive and prognostic information, with scores greater than 25 indicating a marked benefit from adding chemotherapy for treatment.

Oncotype DX is an expensive test performed in a central laboratory in the USA. In comparison, routine immunohistochemical (IHC) markers for breast cancer (ER, PR and HER2) are available locally in most pathology laboratories at a fraction of the price. The cost of each Oncotype DX test, whilst substantially less than the cost of chemotherapy for an individual patient, is still more than a hundredfold the cost of an individual immunohistochemistry test, and this can be a barrier for optimal therapy guidance for a substantial proportion of eligible patients around the world. Additional issues concerning the Oncotype DX are related to difficulties of replicating the assay[1]. Some studies have also questioned the validity of Oncotype DX testing in rarer forms of cancer [2].

Several studies [3-8] have been performed to assess whether the RS scores and/or the associated prognostic information derived from the Oncotype DX assay can be predicted using locally available clinicopathological data and immunohistochemical tests. Probably the best determinants of

predicting Oncotype scores have been found to be the progesterone receptor (PR) expression levels and the Ki67 index [3,4]. ER and PR are routinely performed biomarkers on breast cancers, and the methodology of staining and assessing positivity is well established. The Ki67 immunohistochemical marker is used in some laboratories to check the proliferation rate of tumours, with levels of expression of the marker showing a prognostic and predictive use in breast cancer [9]. Ki67 assessment has been known to show inter-laboratory and inter-observer variability, due to various pre-analytical and analytical conditions, and Ki67 immunohistochemistry is known to have issues with reproducibility [10-12]. Among other factors, the variation is related to the technical procedure of the staining and the intrinsic heterogeneous nature of tumours, with areas showing higher levels of positive nuclei compared to others within the same tumour. The International Ki67 Working Group has been working over recent years to increase concordance of Ki67 scoring, and various training guides are available on the website[13]. Figure 1 below illustrates the heterogeneity of Ki67 expression: what is visible in the photographs may be less than 0.1% of all the tumour visible on the slide.

*****INSERT FIGURE 1 ABOUT HERE*****

The aim of the current study was to evaluate the use of the ER, PR and Ki67 biomarkers to predict RS scores in a group of symptomatic patients in a district general hospital (DGH) setting. Objectives set to meet these aims were: to determine how well Ki67 (alongside parameters derived from the ER and PR biomarkers) predicts RS in this setting; whether a predictive model derived in a specific sample of patients can be generalised to other samples, and whether these routine markers can identify sub-groups of patients with ER positive HER2 negative node negative breast cancer who will not gain additional information from Oncotype DX analysis.

Methods

Laboratory methods:

Since the approval of Oncotype DX as a diagnostic test in routine practice within England, eligible patients' cancers have been tested through the CHFT Histopathology Department. The patients for this study were identified through the Genomic Health database, and the population consisted of symptomatic patients rather than a select population of patients enrolled in clinical trials. We used core biopsies for the study, in preference to the excision specimens, for three reasons: fixation is better on core biopsy with the resultant immunohistochemistry of better quality; the smaller amount of tissue on the core biopsy slide made Ki67 counting easier; and finally, ER and PR were already routinely done and available on core biopsies. The Oncotype DX testing was done on the best tumour block from the excision specimen.

Patient demographics and tumour characteristics, including ER and PR expression and HER2 status was collected. ER10 and PgR10 indices for use in the models were directly derived from the ER and PR biomarkers. The ER10 score is calculated as the H score divided by 10; the PgR10 score is calculated as the percentage of PR positive cells divided by 10. IHC for the Ki67 antigen was performed using the Roche Ventana rabbit monoclonal antibody CONFIRM anti-Ki67 (30-9). The concentration used was approximately 2 μg/mL of primary antibody diluted in a buffer containing carrier protein and preservative. Slides were baked for 30 minutes in a 60°C oven before being loaded on to the Ventana BenchMark Ultra staining platform. Antigen retrieval used heat and CC1 (Tris buffer). Antibody incubation time was 24 minutes with antibody diluent but no casein.

The Ki67-stained slides were scanned, and representative areas were photographed. The representative areas including as much variation as possible, i.e. they were global rather than hot spots. 1000 tumour cell nuclei were counted to derive the Ki67 labelling index. The recommendations from the International Ki67 in Breast Cancer Working Group were used[9]. The slides were then visually inspected to check whether the Ki67 index was representative of the entire biopsy. Where discrepant, repeat scoring was carried out using microscopy.

Statistical methods:

The sample was summarised descriptively. Two candidate intrinsically linear regression models of RS were derived and applied to the entire cohort. Model 1 included adjusted values of the Ki67 biomarker as the sole predictor. Model 2 additionally included the ER10 and PgR10 derived quantities as predictors. Parameter estimates, associated confidence intervals and *p*-values were derived for all parameters in both models. Model fit was assessed using adjusted-R^2 statistics. Regression assumptions were checked by visual inspection of plots of standardised residuals against predicted values, and by normality of standardised residuals.

Following derivation of models on the complete cohort, to ensure that any selected model could be applied to other samples, cross-validation was conducted on both models using the leave-one-out cross-validation (LOOCV) procedure. Cross-validation was assessed using the root mean square error (RMSE) statistic, with lower values of the RMSE indicating a model with potentially good portability. RMSE values may also be interpreted in absolute terms by comparison with the variance of the data.

A scatter plot of RS scores against fitted values was constructed for the model identified to show optimum performance (based on considerations of model fit and cross-validation properties). Any influential data points were identified by inspection of standardised residuals and Cook's distances.

As a subsidiary analysis to investigate the goodness-of-fit of models on specific cases within the ER positive / HER2 negative group, all models were fitted to subgroups of patients defined by PR score; assessing cases with PR=8 and cases with PR<8. Adjusted-R^2 statistics from the resulting models were derived for both subgroups. Cross-validation procedures were not conducted on these subgroups due to low sample sizes.

Results

58 breast cancers from 55 patients were analysed in this study representing all cases seen over the analysis period. All cases had ER, PR and HER2 staining on the core biopsy. Histological parameters were accessed from the excision specimens.

The sample is summarised descriptively in Table 1.

Table 1: descriptive summary of data

Variable	**Mean (SD; range)**
Patient age (years)	56.1 (11.8; 30-77)
Size of tumour (mm)	29.6 (16.6; 7-100)
Whole tumour size (mm)	35.4 (21.0; 14-100)
PR Allred/Quick score	5.31 (3.16; 0-8)
ER Allred/Quick score	7.86 (0.540; 5-8)
ER10 value	9.29 (1.99; 0.6-10.0)
PgR10 value	4.60 (4.01; 0-10)
Ki67 labelling index (%)	27.2 (18.5; 3-85)
RS scores	23.0 (14.8; 5-78)
Variable	**Frequency (valid %)**
Laterality	
Left	29 (50.0%)
Right	29 (50.0%)
Surgery	
Central excision	1 (1.7%)
Mastectomy	21 (36.2%)
Wide local excision	36 (62.1%)
Type	

Invasive ductal carcinoma	45 (77.6%)
Invasive lobular carcinoma	12 (20.7%)
Mixed ductal/lobular carcinoma	1 (1.7%)
Grade	
Grade 2	45 (77.6%)
Grade 3	13 (22.4%)
Lymphovascular invasion	
No invasion	33 (56.9%)
Suspected invasion	2 (3.4%)
Invasion	23 (39.7%)
Metastasis (*n*=24)	
Less than 2 mm	13 (54.2%)
Over 2 mm	11 (45.8%)
Margins	
Less than 1 mm	17 (29.3%)
Over 1 mm	41 (70.7%)
HER2 category	
0	11 (19.0%)
1+	29 (50.0%)
2+	18 (31.0%)

Simple linear regression analyses conducted to assess the corrected Ki67 biomarker (Model 1) revealed it to be significantly associated with RS scores at the 5% significance level (p<0.001). The model was a good fit to the data (adjusted-R^2=0.626).

A multiple linear regression including the corrected Ki67, ER10 and PgR10 biomarkers (Model 2) revealed all terms to be significantly associated with RS scores (controlling for the concurrent effects

of other factors) at the 5% significance level (p<0.001 for each parameter). This model was a very good fit to the data (adjusted-R^2=0.784).

No evidence was revealed for violation of regression assumptions in either model, with all residuals found to be approximately normally distributed and no obvious patterns observable in plots of standardised residuals against standardised predicted values. Both models were considered suitable for further assessment using cross-validation procedures. RMSE values derived in the LOOCV procedure were low (under 4% of model variance in both cases) indicating good portability for both models, with a slightly lower RMSE value derived from Model 2.

Regression and cross-validation parameters from the models are summarised in Table 2.

Table 2: regression and cross-validation parameters (both models)

Model	Biomarkers/index	*P*-value	Parameter estimate (B)	95% CI for B	Adjusted-R^2	RMSE
1	Ki67[1]	<0.001	0.945	(0.752, 1.14)	0.626	9.39
2	Ki67[1] ER10 PgR10	<0.001 <0.001 <0.001	0.672 -2.44 -0.906	(0.501, 0.843) (-3.52, 1.36) (-1.38, -0.428)	0.784	7.26

Hence Model 2 was considered to be preferable to Model 1 in terms of both goodness-of-fit and transferability properties. A scatter plot of RS scores obtained against predicted values for Model 2 (Figure 2) illustrates high goodness-of-fit, with the majority of points clustered around a region of RS scores between about 10 and about 40. Examination of standardised residuals and Cook's distances revealed no evidence for influential data points.

INSERT FIGURE 2 ABOUT HERE

Hence results obtained from this model appear to be valid at least in patients with an RS score in the range 0-30, and possibly beyond. Uncertainty in the estimates of patients with extremely low, or,

particularly, extremely high RS scores will be greater due to the relatively low numbers of patients in the data set used in the prediction of these values.

Partitioning the data set by PR scores for the subsidiary analysis yielded models which fitted the data fairly well for cases with PR<8 but inadequately for PR=8 (Table 3).

Table 3: goodness-of fit statistics (all models): partitioned samples

Model	**Adjusted-R^2 statistics: PR score**	
	PR<8	PR=8
1	0.663	0.421
2	0.803	0.429

Discussion

The key finding of this research is that biomarkers may be used to predict RS scores with high levels of predictability. With an abundance of conflicting studies available, we thought it important to determine the use of Ki67 in our DGH setting. In our setting we have used Ki67 routinely mainly in a neoadjuvant setting with Grade 2 ER positive HER2 negative cancers to help decide whether neoadjuvant chemotherapy (NACT) might be of benefit. We compared our Ki67 scores (with staining and counting done locally) with staining and counting done at the Royal Marsden Hospital. Although our counts were higher by about 30%, they correlated with the Royal Marsden scores, indicting verification of our results after comparison to a national standard.

Our study revealed Model 2 to be the better performing model on the sample data under both goodness-of-fit and cross-validation criteria; although both models were revealed to fit well to the data, and cross-validation procedures indicated good transferability of both models to other samples with minimal expected loss of predictive capability. Alongside the Ki67 biomarker, the selected Model 2 additionally included only quantities derived from biomarkers that are routinely performed on

breast cancers, and there are no other biomarkers that are routinely performed on breast cancers other than those included in this model. Furthermore, the high correlation for this model suggests that in any case it would be unlikely that further biomarkers substantively improving model fit would be identified.

Preliminary evidence from the analysis of data split by PR score suggests that biomarkers may be effective predictors of RS scores in cases with PR<8, but much less so in cases in which PR=8. There were insufficient numbers of either case to pursue this line of enquiry or investigate cross-validation properties in the current investigation; however, findings suggest that further refinements of the model may be necessary if it is required to target more specifically defined groups of patients for predictive modelling.

Conclusions

Locally available, cheaper alternatives to multigene assays to determine therapy in ER positive HER2 negative patients is of benefit both from patient management and financial perspectives. A model has been derived with high capability to predict RS scores accurately from linear combinations of predictive biomarkers in a DGH setting. The model shows promising cross-validation properties, suggesting that outcomes from the Oncotype DX test are predictable from biomarkers in other samples. However, this may need to be tested further with larger samples. Further work is needed to assess the reliability of measures derived from the interpretation of Ki67 staining.

Funding: None declared.

Acknowledgements: Professor Johnathan Joffe for suggesting the initial concept for the study; Professor Mitch Dowsett and Dr Lila Zabaglo at Royal Marsden Hospital for assessing our Ki67 results;

Julie Kilner, Claire Peacock and Katy Bruffell for immunohistochemical staining of Ki67 slides; Robert Steed for scanning slides with Ki67 staining.

Author contributions: VK and JS agreed the methodology, drafted the manuscript, conducted the analysis, interpreted the results, fed back comments and read and approved the final manuscript. KH conducted the analysis.

Competing Interests

The authors declare no competing interests.

Ethical approval

The study did not involve human participants. Ethical approval was not required.

Author contributions: VK and JS agreed the methodology, drafted the manuscript, conducted the analysis, interpreted the results, fed back comments and read and approved the final manuscript. KH conducted the laboratory analysis.

Key messages

We aimed to determine whether the commonly available immunohistochemical markers Ki67 and PR can predict Oncotype DX RS scores in a district general hospital setting. We derived a model with to predict RS scores accurately from linear combinations of predictive biomarkers in a DGH setting. Our model revealed good predictive properties, with evidence for transferability to other samples, suggesting that locally available, cheaper alternatives to multigene assays to determine therapy in ER positive HER2 negative patients would be of benefit.

References

1. Schildgen V, Warm M, Brockmann M, Schildgen O. Oncotype DX Breast Cancer recurrence score resists inter-assay reproducibility with RT^2-Profiler Multiplex RT-PCR. Sci Rep. 2019 Dec 30;9(1):20266. doi: 10.1038/s41598-019-56910-0. PMID: 31889145; PMCID: PMC6937305.
2. Tiberi D, Masucci L, Shedid D, et al. Limitations of Personalized Medicine and Gene Assays for Breast Cancer. Cureus. 2017;9(3):e1100. Published 2017 Mardoi:10.7759/cureus.1100
3. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, Zabaglo L, Mallon E, Green AR, Ellis IO, Howell A, Buzdar AU, Forbes JF. 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 Journal of Clinical Oncology 2011 29:32, 4273-4278
4. Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis Molly E Klein, David J Dabbs, Yongli Shuai, Adam M Brufsky, Rachel Jankowitz, Shannon L Puhalla and Rohit Bhargava Modern Pathology (2013) 26, 658–664; doi:10.1038/modpathol.2013.36; published online 15 March 2013
5. Predicting Oncotype D X scores using clinicopathologic features: A report from the National Cancer Database. DOI: 10.1200/JCO.2018.36.15_suppl.551 Journal of Clinical Oncology 36, no. 15_suppl (May 20, 2018) 551-551.
6. Prediction of Oncotype DX recurrence score using deep multi-layer perceptrons in estrogen receptor-positive, HER2-negative breast cancer Baltres, A., Al Masry, Z., Zemouri, R. et al. Breast Cancer 27, 1007–1016 (2020).
7. Prediction of Oncotype Dx Recurrence Score Using Clinical Parameters: A Comparison of Available Tools and a Simple Predictor Based on Grade and Progesterone Receptor. Hematology/Oncology and Stem Cell Therapy. Thibodeau, Stephane & Voutsadakis, Ioannis. (2019).

8. Prediction of Oncotype Dx Recurrence Score Using Clinical Parameters: A Comparison of Available Tools and a Simple Predictor Based on Grade and Progesterone Receptor. Hematology/Oncology and Stem Cell Therapy. Thibodeau, Stephane & Voutsadakis, Ioannis. (2019).
9. Assessment of Ki-67 in breast cancer: Recommendations from the International Ki-67 in breast cancer working group. JNCI 2011; 103: 1-9
10. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer Diagnostics guidance Published: 19 December 2018 www.nice.org.uk/guidance/dg34
11. Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, Obermann E, Ohlschlegel C, Padberg B, Rakozy C, Sancho Oliver S, Schobinger-Clement S, Schreiber-Facklam H, Singer G, Tapia C, Wagner U, Mastropasqua MG, Viale G, Lehr HA. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. PLoS One. 2012;7(5):e37379. doi: 10.1371/journal.pone.0037379. Epub 2012 May 25. PMID: 22662150; PMCID: PMC3360682.
12. Namuche F, Ruiz R, Flores C, Gomez HL, Aguilar A. Ki67 cut offs and oncotype DX recurrence score in early breast cancer [abstract]. In: Proceedings of the 2017 San Antonio Breast Cancer Symposium; 2017 Dec 5-9; San Antonio, TX. Philadelphia (PA): AACR; Cancer Res 2018;78(4Suppl):Abstract nr P6-09-11.
13. International Ki67 in Breast Cancer Working Group. Available from URL https://www.ki67inbreastcancerwg.org/