FULL PAPER

BJC

British Journal of Cancer (2014) 111, 2242–2247 | doi: 10.1038/bjc.2014.530

Keywords: breast cancer; biomarkers; time dependency; relapse risk

Time dependence of biomarkers: non-proportional effects of immunohistochemical panels predicting relapse risk in early breast cancer

J Stephen*[,1], G Murray[1], D A Cameron[2], J Thomas[2], I H Kunkler[2], W Jack[2], G R Kerr[2], T Piper[2], C L Brookes[3], D W Rea[3], C J H van de Velde[4], A Hasenburg[5], C Markopoulos[6], L Dirix[7], C Seynaeve[8] and J M S Bartlett[2,9]

[1]*Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK;* [2]*Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK;* [3]*Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK;* [4]*Leiden University Medical Center, Leiden, The Netherlands;* [5]*Department of Gynecological Oncology, University Hospital, Freiburg, Germany;* [6]*Department of Surgery, Athens University Medical School, Athens, Greece;* [7]*Oncology Center, St Augustinus, Antwerp, Belgium;* [8]*Department of Medical Oncology, Erasmus MC-Daniel den Hoed Cancer Center, Rotterdam, The Netherlands and* [9]*Transformative Pathology, Ontario Institute for Cancer Research, Toronto, Canada*

Background: We investigated the impact of follow-up duration to determine whether two immunohistochemical prognostic panels, IHC4 and Mammostrat, provide information on the risk of early or late distant recurrence using the Edinburgh Breast Conservation Series and the Tamoxifen *vs* Exemestane Adjuvant Multinational (TEAM) trial.

Methods: The multivariable fractional polynomial time (MFPT) algorithm was used to determine which variables had possible non-proportional effects. The performance of the scores was assessed at various lengths of follow-up and Cox regression modelling was performed over the intervals of 0–5 years and >5 years.

Results: We observed a strong time dependence of both the IHC4 and Mammostrat scores, with their effects decreasing over time. In the first 5 years of follow-up only, the addition of both scores to clinical factors provided statistically significant information ($P<0.05$), with increases in R^2 between 5 and 6% and increases in D-statistic between 0.16 and 0.21.

Conclusions: Our analyses confirm that the IHC4 and Mammostrat scores are strong prognostic factors for time to distant recurrence but this is restricted to the first 5 years after diagnosis. This provides evidence for their combined use to predict early recurrence events in order to select those patients who may/will benefit from adjuvant chemotherapy.

Prediction of residual risk following adjuvant endocrine treatment in early breast cancer has become a critical component in the selection of treatment options. In addition, recent data on the impact of up to 10 years of adjuvant endocrine therapy in early breast cancer suggest that some women may benefit from extended treatment with tamoxifen(Goss *et al*, 2003; Davies *et al*, 2013; Gray *et al*, 2013). At present, patients and their clinicians may be presented with three time-related choices related to the treatment of luminal breast cancers: (1) whether it is beneficial to treat patients with adjuvant chemotherapy and therefore delay endocrine therapy; (2) whether to follow a switch strategy for endocrine therapy; and (3) whether to extend endocrine therapy after 5 years for a total of 10 years. Multiple diagnostic algorithms have been developed to provide information on residual risk for patients facing these choices;

*Correspondence: J Stephen; E-mail: j.stephen-2@sms.ed.ac.uk

Received 22 May 2014; revised 1 September 2014; accepted 10 September 2014; published online 14 October 2014

however, there is limited information on how time impacts the risk assessments provided by these diagnostic tools.

A recent study developed a residual risk model (IHC4; Cuzick *et al*, 2011) combining immunohistochemical (IHC) assessment of ER, PgR, Ki67 and HER2, which provided equivalent information on residual risk to the multiparameter, PCR-based, OncotypeDx test. The IHC4 + C (IHC markers plus clinicopathologic parameters) score also provides additional information on the residual risk of distant recurrence to ER-positive primary breast cancer patients receiving adjuvant endocrine therapy, supplementary to that provided by the Adjuvant online! and NPI intermediate-risk groups (Barton *et al*, 2012). Another IHC model (Mammostrat; Ring *et al*, 2006) has been developed combining five IHC biomarkers (p53, NDRG1,CEACAM5, SLC7A5 and HTF9C), which has been shown to significantly improve traditional prognostic factors in predicting outcome for ER-positive breast cancer patients (Ross *et al*, 2008; Bartlett *et al*, 2010, 2012).

Non-proportional effects of prognostic scores have been previously reported with better prediction for patients at high risk of early relapse rather than for those at risk of later recurrent disease (Buyse *et al*, 2006; Desmedt *et al*, 2007; Haibe-Kains *et al*, 2008). Haibe-Kains *et al* (2008) suggested that this could be due to three reasons: (i) different biological mechanisms for the occurrence of early and late relapses; (ii) the statistical methodology, with the scores being developed on cohorts based on median follow-up; (iii) the quality of survival data, where one could intuitively believe that the quality of survival data decreases with respect to the duration of follow-up as it is difficult to follow up patients during a long period resulting in a high level of censoring.

In this study we investigated the impact of follow-up duration on the IHC4 and Mammostrat scores to determine whether these two prognostic panels provide information on the risk of early or late recurrence.

MATERIALS AND METHODS

Materials. The Edinburgh Breast Conservation Series (BCS) represents a fully documented, consecutive cohort of 1812 patients treated by breast conservation surgery, axillary node sampling or clearance, and whole breast radiotherapy at the Edinburgh Cancer Centre between 1981 and 1998 (Thomas *et al*, 2009; Bartlett *et al*, 2010). Following ethical approval (Lothian Local Research Ethics 04) tissue blocks were retrieved from all cases and sufficient material was available from 1686 cases for assembly into tissue microarrays (TMAs) (Supplementary Figure 1A). For all cases with available tissue, tumours were regraded on whole sections by a single pathologist (Thomas *et al*, 2009).

The Tamoxifen Exemestane Adjuvant Multinational (TEAM) trial is a multinational randomised, open-label, phase III trial in postmenopausal women with hormone receptor-positive early breast cancer testing the efficacy of 5 years of exemestane (25 mg once per day) *vs* tamoxifen (20 mg once per day for 2.5–3 years) followed by exemestane for a total of 5 years (van de Velde *et al*, 2011). Five of nine participating countries provided paraffin-embedded tumour samples for pathology sub-studies (Bartlett *et al*, 2011a). Tissue blocks were received at a central laboratory and 4598 were found suitable for TMA construction (Supplementary Figure 1B).

Biomarker analysis. Immunohistochemical staining for a panel of biomarkers including ER, PgR, HER2, Ki67, HTF9C, CEACAM5, NDRG1, p53 and SLC7A5 and FISH (fluorescence *in situ* hybridisation) for HER2 was performed using either sextuplet (ER and PgR) or triplicate (all other markers) 0.6 mm^2 TMA cores. Results were derived from dual scoring by expert observers (as described by Kirkegaard *et al* (2006)) for the Edinburgh BCS cohort for all markers. For TEAM patients, ER, PgR and Ki67 scores were derived by quantitative image analysis using the Ariol system with algorithms validated against both whole sections and manual assessment (Faratian *et al*, 2009; Bartlett *et al*, 2011a). Data for ER were recorded as a histoscore (Kirkegaard *et al*, 2006) and for Ki67 and PgR as a percentage of positive cells (ATAC and Ki67 guidelines; Dowsett *et al*, 2011). Results for HER2 were scored according to the UK guidelines (Walker *et al*, 2008; Bartlett *et al*, 2011b), with cases regarded as HER2-amplified if any core showed amplification/overexpression. Positivity for p53, HTF9C (recently re-named TRIMT2A), CEACAM5, NDRG1 and SLC7A5 was recorded as previously described (Ring *et al*, 2006; Ross *et al*, 2008; Bartlett *et al*, 2010, 2012).

Generation of prognostic scores. The IHC4 model (Cuzick *et al*, 2011) utilised a linear combination of multiple markers: ER, PgR, HER2 and Ki67. Continuous marker scores were normalised prior to inclusion in the IHC4 model. ER histoscores were divided by 30, and PgR scores as a percentage of cells staining positive were divided by 10 to obtain continuous values between 0 and 10. Ki67 scores were represented as percentage positive cells and HER2 was treated as a dichotomous variable. The IHC4 risk score was generated according to the previously specified algorithm (Cuzick *et al*, 2011). The IHC4 score is analysed as a continuous risk score, except for Kaplan–Meier analyses, in which the IHC4 score is categorised into three groups using two cutoff points that correspond to a 10-year distant recurrence rate of 10% and 20% from the original study; however, these cutoffs have not been previously validated (Cuzick *et al*, 2011).

The Mammostrat model (Ring *et al*, 2006) used five IHC markers: SLC7A5, CEACAM5, NDRG1, HTF9C and p53. The Mammostrat risk score was generated by combining binary staining results for all markers as either positive or negative according to the previously specified algorithm (Ring *et al*, 2006; Ross *et al*, 2008; Bartlett *et al*, 2010, 2012). The Mammostrat score was categorised into low ($\leqslant 0$), medium (>0 and <0.7) and high ($\geqslant 0.7$) risk categories as previously specified (Ring *et al*, 2006; Ross *et al*, 2008; Bartlett *et al*, 2010, 2012).

Missing data. The known technical limitations of TMAs inevitably result in missing data (Voduc *et al*, 2008). A large amount of data were missing (32.1%) for the PgR variable measured as a percentage of positive cells in the Edinburgh BCS cohort. Therefore, multiple imputation was performed using all predictors plus the event indicator and the Nelson–Aalen estimator of the cumulative baseline hazard, as recommended by White and Royston (2009). We used the *mi impute chained* command in Stata to perform multiple imputation using chained equations to generate 42 imputed data sets, based on the rule of thumb suggested by White *et al* (2011). The results from analyses on each of the imputed data sets were combined using Rubin's rules to produce estimates and confidence intervals that incorporate the uncertainty of imputed values (Rubin, 1987).

Statistical analysis. The primary end point selected for this study was time to distant recurrence (TTDR) as this is the event that drives subsequent death from breast cancer. Time to distant recurrence was defined as the time to distant metastasis (van de Velde *et al*, 2011) or death with evidence of recurrent breast cancer, with patients censored at the time of last follow-up. Additional clinical variables used, when specified, were age (continuous), tumour size (continuous), number of positive nodes (continuous), histological grade (grade I–III), treatment (exemestane/tamoxifen in TEAM) and chemotherapy (yes/no). Investigation into the functional form of continuous covariates on the log-hazard identified tumour size and the number of positive nodes to be non-linear and were included as such in all analyses.

The primary analysis was based on ER-positive patients treated with endocrine therapy (without chemotherapy). Secondary analyses were performed on all ER-positive patients irrespective of treatment. Exploratory analyses were performed on node-negative and node-positive ER-positive patients treated with endocrine therapy (without chemotherapy) and on node-positive ER-positive patients irrespective of treatment.

The assumption of proportional hazards when performing Cox regression was tested by plotting scaled Schoenfeld residuals against time and testing for a zero slope. Cox modelling was performed including a covariate–log (time) interaction into the model. The multivariable fractional polynomial time (MFPT) algorithm proposed by Sauerbrei *et al* (2007) was used to determine which variables had possible non-proportional effects, as well as the best-fitting fractional polynomial to model these effects. We explored the effect of these derived risk scores before and after 5 years (as a decision point for continued endocrine therapy at this time) to determine whether either score provided information on the residual risk of early or late distant recurrence.

The performance of the IHC4 and Mammostrat risk scores was assessed along with conventional clinical risk factors using measures of calibration and discrimination, with follow-up censored at 5, 9 and 15 years for the Edinburgh BCS cohort and 3 and 5 years for the TEAM cohort. 'Model calibration' refers to how closely the estimates of survival from the model agree with the survival from the observed data (Altman *et al*, 2009; Moons *et al*, 2009). This was assessed for each decile of predicted risk, ensuring 10 equally sized groups, by producing a calibration plot (observed *vs* predicted probabilities of 5-year distant recurrence) and calculating the calibration slope. 'Discrimination' is the ability of a risk score to differentiate between patients who do and those who do not experience an event during the study period (McGeechan *et al*, 2008; Altman *et al*, 2009). Discrimination was evaluated using Royston and Sauerbrei's (2004) R^2 statistic based on their index of discrimination (D), with a difference in D of at least 0.1 indicating improved prognostic separation. All statistical analyses were carried out in Stata (version 12).

RESULTS

Data were available on 1449 (ER-negative and ER-positive) patients with 273 distant recurrences from the Edinburgh BCS cohort (median follow-up of 12.9 years) and on 3766 (ER-positive) patients with 548 distant recurrences from the TEAM cohort (median follow-up of 6.2 years). The distributions of the scores were different between the two cohorts (Supplementary Figure 2), with the median IHC4 scores being 44 and 27 in the Edinburgh BCS and TEAM cohorts, respectively. A larger proportion of patients (64%) were allocated to the low Mammostrat risk group in the Edinburgh BCS cohort compared with 43% in the TEAM cohort. This is as expected because of the TEAM cohort being a higher-risk population compared with the Edinburgh BCS cohort, with higher mean tumour size (23 mm *vs* 16 mm, respectively), a larger proportion of higher grade tumours (grade 3: 35% *vs* 19%, respectively) and higher mean number of positive nodes (1.9 *vs* 0.5, respectively).

Non-proportional effects in a multivariable model. In multivariable modelling, the MFPT algorithm determined IHC4 to have a significant time-by-covariate interaction for both patient subgroups in both cohorts with the best-fitting FP to be log of time. The parameter associated with this interaction was negative, suggesting the effect of a unit increase in IHC4 on TTDR decreased over time (Figure 1). The decrease over time was more prominent in the Edinburgh BCS cohort, with the adjusted HR crossing the value 1 (corresponding to a null effect) at approximately 6.5 years.

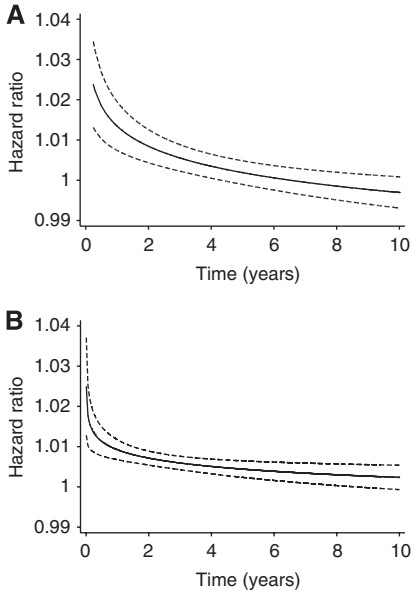

Figure 1. Time-dependent adjusted hazard ratio estimate (up to 10 years) with 95% CIs (dashed lines) for a unit increase in IHC4 score for all ER-positive patients in (A) the Edinburgh BCS cohort and (B) the TEAM cohort. Adjusted for age, grade, nodes positive, treatment and chemotherapy.

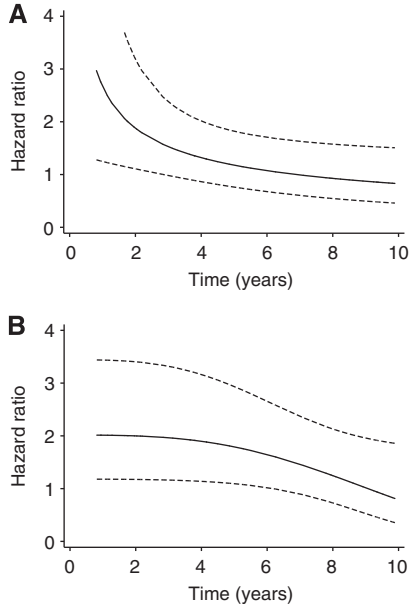

Figure 2. Time-dependent adjusted hazard ratio estimate with 95% CIs (dashed lines) for high-risk compared with low-risk Mammostrat score in the Edinburgh BCS cohort for (A) all ER-positive patients and (B) ER-positive patients treated with endocrine therapy only. Adjusted for age, grade, nodes positive, treatment and chemotherapy.

The Mammostrat score was determined to have a significant time-by-covariate interaction in the Edinburgh BCS cohort only for high risk *vs* low risk (Figure 2). There was uncertainty in the best-fitting FP to model the interaction, with log of time chosen for all ER-positive patients and time-cubed chosen for ER-positive patients treated with endocrine therapy only.

Impact of follow-up duration on model performance. We assessed the performance of the scores in addition to clinical factors at various lengths of follow-up. Measures of discrimination are given in Table 1 for full follow-up (using all available data

Table 1. Performance data on IHC4 and Mammostrat score in addition to clinical factors

	All ER-positive		ER-positive endocrine only	
Model	**% R^2 (95 % CI)**	**D-Statistic (95% CI)**	**% R^2 (95 % CI)**	**D-Statistic (95% CI)**
Edinburgh BCS				
Full follow-up				
Clinical	24.0 (16.2–31.7)	1.15 (0.91–1.39)	25.7 (16.7–34.6)	1.20 (0.92–1.48)
IHC4	24.4 (16.7–32.1)	1.16 (0.92–1.41)	26.3 (17.4–35.1)	1.22 (0.94–1.50)
Mammostrat	25.4 (17.6–33.2)	1.19 (0.95–1.44)	27.1 (18.1–36.1)	1.25 (0.97–1.53)
Combined	25.7 (18.0–33.5)	1.21 (0.96–1.45)	27.9 (19.0–36.8)	1.27 (0.99–1.56)
5 Years				
Clinical	31.1 (20.8–41.4)	1.51 (1.17–1.85)	35.3 (23.3–47.4)	1.51 (1.12–1.91)
IHC4	33.7 (23.6–43.8)	1.63 (1.28–1.98)	39.0 (27.2–50.7)	1.63 (1.23–2.04)
Mammostrat	33.6 (23.4–43.8)	1.58 (1.23–1.92)	38.5 (26.8–50.3)	1.62 (1.22–2.02)
Combined	36.5 (26.4–46.5)	1.69 (1.34–2.04)	41.3 (29.9–52.7)	1.72 (1.31–2.12)
TEAM				
Full follow-up				
Clinical	27.5 (23.00–32.0)	1.26 (1.12–1.40)	29.5 (23.6–35.3)	1.33 (1.14–1.51)
IHC4	31.0 (26.5–35.4)	1.37 (1.23–1.52)	32.8 (27.0–38.4)	1.43 (1.24–1.62)
Mammostrat	29.8 (25.3–34.2)	1.33 (1.19–1.48)	31.0 (25.1–36.7)	1.37 (1.18–1.56)
Combined	32.3 (27.8–36.7)	1.41 (1.27–1.56)	33.3 (27.5–38.9)	1.45 (1.26–1.63)
5 Years				
Clinical	29.0 (23.8–34.1)	1.31 (1.14–1.47)	30.5 (23.7–37.0)	1.36 (1.14–1.57)
IHC4	34.0 (28.9–38.9)	1.47 (1.31–1.63)	34.9 (28.3–41.2)	1.50 (0.29–1.71)
Mammostrat	31.5 (26.4–36.5)	1.39 (1.23–1.55)	32.2 (25.5–38.6)	1.41 (1.20–1.62)
Combined	35.4 (30.3–40.2)	1.51 (1.35–1.68)	35.5 (28.9–41.7)	1.52 (1.31–1.73)

Abbreviations: BCS = breast conservation series; CI = confidence interval; ER = estrogen receptor; IHC = immunohistochemical; IHC4 = IHC4 score. NOTE: Measures of discrimination for patients in the Edinburgh BCS and TEAM cohorts with full follow-up and follow-up censored at 5 years. A difference in D of at least 0.1 indicates improved prognostic separation. The models being assessed are: clinical factors, age, grade, tumour size, nodes positive, treatment and chemotherapy; IHC4, IHC4 in addition to clinical factors; Mammostrat, Mammostrat in addition to clinical factors; Combined, IHC4 and Mammostrat in addition to clinical factors.

Table 2. Period-specific multivariate Cox regression of IHC4 and Mammostrat score

	All ER-positive		ER-positive endocrine only	
	0–5 years	**5–10 years**	**0–5 Years**	**5–10 years**
Main effect[a]	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**
Edinburgh BCS				
	(97 events)	(61 events)	(70 events)	(45 events)
IHC4	2.09 (1.07–4.05)	1.02 (0.53–1.96)	1.79 (0.87–3.71)	1.20 (0.59–2.44)
MvL	1.84 (1.14–2.97)	1.50 (0.80–2.80)	2.00 (1.13–3.54)	1.50 (0.72–3.13)
HvL	1.65 (0.94–2.92)	1.36 (0.65–2.85)	1.87 (0.98–3.57)	1.68 (0.75–3.77)
TEAM				
	(416 events)	(132 events)	(242 events)	(74 events)
IHC4	1.69 (1.18–2.44)	1.21 (0.70–2.07)	1.34 (0.85–2.10)	0.89 (0.44–1.78)
MvL	1.44 (1.09–1.90)	1.38 (0.86–2.21)	1.41 (0.98–2.01)	0.96 (0.51–1.81)
HvL	2.01 (1.56–2.60)	1.57 (1.02–2.44)	1.72 (1.24–2.38)	1.18 (0.69–2.04)

Abbreviations: BCS = breast conservation series; CI = confidence interval; HvL = high v low risk Mammostrat score; IHC = immunohistochemical; IHC4 = IHC4 score; MvL = medium v low risk Mammostrat score. NOTE: Separate multivariate analysis of IHC4 and Mammostrat scores with conventional histopathologic variables for TTDR before 5 years and after 5 years for patients in different subgroups. Values represent estimated hazard ratios and 95% CIs calculated as between the IQR of the continuous IHC4 score and high and medium risk compared with low risk as categorised by the Mammostrat score for all patients with ER-positive breast cancer and ER-positive patients who received no chemotherapy in the Edinburgh BCS and TEAM cohorts.
[a]Adjusted for age, grade, tumour size, nodes positive, treatment and chemo.

rather than censoring at a specific time point) and follow-up censored at 5 years. In the Edinburgh BCS cohort the models performed statistically better with shorter follow-up compared with full follow-up with differences in D-statistic between 0.4 and 0.5 and R^2 between 7 and 13%. There was a small improvement in model performance with shorter follow-up in the TEAM cohort, with increases in R^2 between 1.5 and 3% and differences in D-statistic between 0.05 and 1. The calibration of the combined model was improved in the Edinburgh BCS cohort with follow-up censored at 5 years, with a calibration slope estimate of 1.0 (95% CI 0.8–1.1) for follow-up censored at 5 years *vs* 1.2 (0.8–1.5) for full follow-up for all ER-positive patients.

Prognostic value of scores within the first 5 years and beyond 5 years after diagnosis. To investigate whether either score provided prognostic information on TTDR beyond 5 years after diagnosis, Cox regression was performed with follow-up time divided into the intervals 0–5 years and 5–10 years. Period-specific Kaplan–Meier curves are displayed in Supplementary Figure 3. Both scores were significant independent predictors of outcome restricted to the first 5 years of follow-up, after which there was no evidence that the scores were associated with TTDR (Table 2). For example, the interquartile HR for IHC4 score was 2.1 (95% CI, 1.1–4.1) in the first 5 years after diagnosis compared with 1.0 (95% CI, 0.5–2.0) after 5 years for all ER-positive patients in the

Edinburgh BCS cohort. There was evidence of a prognostic effect after 5 years for Mammostrat high risk *vs* low risk for all ER-positive patients (Table 2) and ER-positive node-negative patients treated with endocrine therapy (Supplementary Table 1) in the TEAM cohort with HRs of 1.6 (95% CI, 1.0–2.4) and 3.3 (95% CI, 1.1–10.5), respectively. This effect of Mammostrat high *vs* low risk was also seen in the Edinburgh BCS cohort for ER-positive node-negative patients treated with endocrine therapy only (Supplementary Table 1) for the 5–10-year time period after diagnosis with an HR of 2.8 (95% CI, 1.0–7.8).

Comparison of IHC4 and Mammostrat. The IHC4 score provided additional prognostic information beyond that of clinical factors compared with the Mammostrat score for all ER-positive patients in both patient cohorts in the first 5 years of follow-up (Supplementary Table 2; increase in R^2: 6.6% *vs* 2.5% and D-statistic: 0.12 *vs* 0.07 in the Edinburgh BCS cohort and increase in R^2: 5.0% *vs* 2.5% and D-statistic: 0.16 *vs* 0.08 in the TEAM cohort). Similarly, for ER-positive patients treated with endocrine therapy in the TEAM cohort, the IHC4 score was the stronger predictor of outcome, whereas in the Edinburgh cohort the prognostic information provided by either score was similar (increase in R^2: 3.7% *vs* 3.2% and D-statistic: 0.12 *vs* 0.11).

The addition of both scores to clinical factors. The scores were entered simultaneously into a multivariate Cox regression model, and in the first 5 years of follow-up the addition of both scores to clinical factors provided statistically significant information ($P<0.05$) for both subsets of patients across both cohorts with increases in R^2 between 5 and 6% and increases in D-statistic between 0.16 and 0.21 (Supplementary Table 2). However, both scores only remained significant independent predictors of TTDR restricted to the first 5 years of follow-up when simultaneously entered into a multivariate Cox regression model for all ER-positive patients in the TEAM cohort with an interquartile HR for ICH4 score of 1.6 (95% CI, 1.1–2.4) and HRs for medium and high *vs* low Mammostrat score of 1.3 (95% CI, 1.0–1.8) and 1.7 (95% CI, 1.3–2.2), respectively (Supplementary Tables 3). Only the IHC4 score provided significant independent prognostic information on TTDR in the first 5 years of follow-up for all ER-positive patients treated with endocrine therapy in the TEAM cohort and for both patient subgroups in the Edinburgh BCS cohort. Although not statistically significant, the Mammostrat score provided some improvement in model discrimination over and above that provided by the IHC4 score and clinical factors with an increase in R^2 and D-statistic of 2.3% and 0.09, respectively, for ER-positive patients treated with endocrine therapy only in the Edinburgh BCS cohort. There was evidence of an effect of Mammostrat high risk *vs* low risk after 5 years of survival after adjustment for IHC4 and clinical factors in ER-positive, node-negative patients treated with endocrine therapy only in the TEAM and Edinburgh BCS cohorts with HRs of 3.2 (95% CI, 1.0–10.1) and 2.8 (95% CI, 1.0–7.7), respectively (Supplementary Table 4).

DISCUSSION

Our analyses confirm that the IHC4 and Mammostrat scores are strong prognostic predictors for TTDR, but this is restricted to the first 5 years after diagnosis. The prognostic effect of IHC4 and Mammostrat score on TTDR decreased with increasing follow-up time. A previous analysis by Sgroi *et al* (2013) also confirmed a significant prognostic ability for IHC4 for early distant recurrence only (0–5 years).

The performance of both scores was good, especially in the first 5 years of follow-up, with the combination of both scores significantly improving the ability to discriminate between events and non-events when compared with clinical factors only and good calibration between observed and predicted 5-year risk of TTDR. The IHC4 score provided more prognostic information on TTDR compared with the Mammostrat score in the first 5 years of follow-up except for all ER-positive patients in the larger TEAM cohort where the addition of both scores provided statistically significant information.

Despite the effects of the scores being strongest in the first 5 years after diagnosis, there was evidence of a prognostic effect of the Mammostrat score with respect to the risk of late recurrence (beyond 5 years after diagnosis). ER-positive, node-negative patients treated with endocrine therapy only who were classified as having a high-risk Mammostrat score had 2.8 (95% CI, 1.0–7.8) and 3.3 (95% CI, 1.1–10.5) times the risk for distant recurrence after 5 years compared with those classified as havng low risk in the Edinburgh BCS and TEAM cohorts, respectively. These results suggest the possible use of Mammostrat score to predict the risk of late recurrence, which will need to be investigated further on other patient cohorts with long-term follow-up.

Missing data were present, although the majority of patients had information on all risk factors. We used current recommended approaches with multiple imputation to overcome the biases that occur when performing complete-case analysis (Burton and Altman, 2004; Vergouwe *et al*, 2010).

The IHC4 is analysed as a continuous score, but for Kaplan–Meier analysis cutoff points are required. To avoid the biases that occur from choosing our own cutoff points, we used those from the original study (Cuzick *et al*, 2011). However, these did not validate well in our cohorts, allocating only a small number of patients to the low-risk group ($<10\%$).

In conclusion, the IHC4 and Mammostrat risk scores were significantly associated with risk for distant disease recurrence in the first 5 years after diagnosis and added prognostic information beyond that provided by clinical factors. Mammostrat may provide insights to patients and clinicians seeking to make informed decision about extended endocrine therapy after an initial 5 years of treatment and warrants further study.

CONFLICT OF INTEREST

JMSB has received travel support and speakers fees from GE Healthcare (who market Mammostrat). The remaining authors declare no conflict of interest.

REFERENCES

Altman DG, Vergouwe Y, Royston P, Moons KGM (2009) Prognosis and prognostic research: validating a prognostic model. *Br Med J* **338**: b605.

Bartlett JMS, Bloom KJ, Piper T, Lawton TJ, van de Velde CJH, Ross DT, Ring BZ, Seitz RS, Beck RA, Hasenburg A, Kieback D, Putter H, Markopoulos C, Dirix L, Seynaeve C, Rea D (2012) Mammostrat as an immunohistochemical multigene assay for prediction of early relapse risk in the tamoxifen versus exemestane adjuvant multicenter trial pathology study. *J Clin Oncol* **30**(36): 4477–4484.

Bartlett JMS, Brookes CL, Robson T, van de Velde CJH, Billingham LJ, Campbell FM, Grant M, Hasenburg A, Hille ETM, Kay C, Kieback DG, Putter H, Markopoulos C, Kranenbarg EM-K, Mallon EA, Dirix L, Seynaeve C, Rea D (2011a) Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: a prospectively powered pathology study in the tamoxifen and exemestane adjuvant multinational trial. *J Clin Oncol* **29**(12): 1531–1538.

Bartlett JMS, Starczynski J, Atkey N, Kay E, O'Grady A, Gandy M, Ibrahim M, Jasani B, Ellis IO, Pinder SE, Walker RA (2011b) HER2 testing in the UK: recommendations for breast and gastric *in-situ* hybridisation methods. *J Clin Pathol* **64**(8): 649–653.

Bartlett JMS, Thomas J, Ross DT, Seitz RS, Ring BZ, Beck RA, Pedersen HC, Munro A, Kunkler IH, Campbell FM, Jack W, Kerr GR, Johnstone L, Cameron DA, Chetty U (2010) Mammostrat (R) as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res* **12**(4): R47.

Barton S, Zabaglo L, A'Hern R, Turner N, Ferguson T, O'Neill S, Hills M, Smith I, Dowsett M (2012) Assessment of the contribution of the IHC4 + C score to decision making in clinical practice in early breast cancer. *Br J Cancer* **106**(11): 1760–1765.

Burton A, Altman DG (2004) Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer* **91**(1): 4–8.

Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris A, Bogaerts J, Therasse P, Floore A, Amakrane M, Piette F, Rutgers E, Sotiriou C, Cardoso F, Piccart MJ, Consortium T (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98**(17): 1183–1192.

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 (2011) 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* **29**(32): 4273–4278.

Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, Abraham M, Medeiros Alencar VH, Badran A, Bonfill X, Bradbury J, Clarke M, Collins R, Davis SR, Delmestri A, Forbes JF, Haddad P, Hou M-F, Inbar M, Khaled H, Kielanowska J, Kwan W-H, Mathew BS, Mittra I, Mueller B, Nicolucci A, Peralta O, Pernas F, Petruzelka L, Pienkowski T, Radhika R, Rajan B, Rubach MT, Tort S, Urrutia G, Valentini M, Wang Y, Peto R. Adjuvant Tamoxifen Longer A (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* **381**(9869): 805–816.

Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, D'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JGM, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clinical Cancer Research* **13**(11): 3207–3214.

Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, Ellis M, Henry NL, Hugh JC, Lively T, McShane L, Paik S, Penault-Llorca F, Prudkin L, Regan M, Salter J, Sotiriou C, Smith IE, Viale G, Zujewski JA, Hayes DF (2011) Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group. *J Natl Cancer Inst* **103**(22): 1656–1664.

Faratian D, Munro A, Twelves C, Bartlett JMS (2009) Membranous and cytoplasmic staining of Ki67 is associated with HER2 and ER status in invasive breast carcinoma. *Histopathology* **54**(2): 254–257.

Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB, Piccart MJ, Castiglione M, Tu D, Shepherd LE, Pritchard KI, Livingston RB, Davidson NE, Norton L, Perez EA, Abrams JS, Therasse P, Palmer MJ, Pater JL (2003) A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med* **349**(19): 1793–1802.

Gray RG, Rea D, Handley K, Bowden SJ, Perry P, Earl HM, Poole CJ, Bates T (2013) aTTom: Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer. In 2013 ASCO Annual Meeting.

Haibe-Kains B, Desmedt C, Piette F, Buyse M, Cardoso F, van't Veer L, Piccart M, Bontempi G, Sotiriou C (2008) Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* **9**: 394.

Kirkegaard T, Edwards J, Tovey S, McGlynn LM, Krishna SN, Mukherjee R, Tam L, Munro AF, Dunne B, Bartlett JMS (2006) Observer variation in immunohistochemical analysis of protein expression, time for a change? *Histopathology* **48**(7): 787–794.

McGeechan K, Macaskill P, Irwig L, Liew G, Wong TY (2008) Assessing new biomarkers and predictive models for use in clinical practice a clinician's guide. *Arch Intern Med* **168**(21): 2304–2310.

Moons KGM, Altman DG, Vergouwe Y, Royston P (2009) Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Br Med J* **338**: b606.

Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MCU, Yoder BJ, Budd GT, Nielsen TO, Hicks DG, Estopinal NC, Ross DT (2006) Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol* **24**(19): 3039–3047.

Ross DT, Kim C-y, Tang G, Bohn OL, Beck RA, Ring BZ, Seitz RS, Paik S, Costantino JP, Wolmark N (2008) Chemosensitivity and stratification by a five monoclonal antibody immunohistochemistry test in the NSABP B14 and B20 trials. *Clini Cancer Res* **14**(20): 6602–6609.

Royston P, Sauerbrei W (2004) A new measure of prognostic separation in survival data. *Stat Med* **23**(5): 723–748.

Rubin D (1987) *Multiple imputation for nonresponse in surveys*. Wiley.

Sauerbrei W, Royston P, Look M (2007) A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biom J* **49**(3): 453–473.

Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, Erlander MG, Dunbier A, Sidhu K, Lopez-Knowles E, Goss PE, Dowsett M (2013) 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* **14**(11): 1067–1076.

Thomas JSJ, Kerr GR, Jack WJL, Campbell F, McKay L, Pedersen H-C, Kunkler IH, Cameron DA, Chetty U, Bartlett JMS (2009) Histological grading of invasive breast carcinoma-a simplification of existing methods in a large conservation series with long-term follow-up. *Histopathology* **55**(6): 724–731.

van de Velde CJH, Rea D, Seynaeve C, Putter H, Hasenburg A, Vannetzel J-M, Paridaens R, Markopoulos C, Hozumi Y, Hille ETM, Kieback DG, Asmar L, Smeets J, Nortier JWR, Hadji P, Bartlett JMS, Jones SE (2011) Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): a randomised phase 3 trial. *Lancet* **377**(9762): 321–331.

Vergouwe Y, Royston P, Moons KGM, Altman DG (2010) Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* **63**(2): 205–214.

Voduc D, Kenney C, Nielsen TO (2008) Tissue Microarrays in clinical oncology. *Seminars in Radiation Oncology* **18**(2): 89–97.

Walker RA, Bartlett JMS, Dowsett M, Ellis IO, Hanby AM, Jasani B, Miller K, Pinder SE (2008) HER2 testing in the UK: further update to recommendations. *J Clin Pathol* **61**(7): 818–824.

White IR, Royston P (2009) Imputing missing covariate values for the Cox model. *Stat Med* **28**(15): 1982–1998.

White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* **30**(4): 377–399.

StataCorp (2011) *Stata Statistical Software: Release 12*. StataCorp LP: College Station, TX.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)