

**Keywords:** breast cancer; prognostic model; young onset; HER2

# An evaluation of the prognostic model PREDICT using the POSH cohort of women aged $\leq 40$ years at breast cancer diagnosis

T Maishman<sup>\*1</sup>, E Copson<sup>1</sup>, L Stanton<sup>1</sup>, S Gerty<sup>1</sup>, E Dicks<sup>2</sup>, L Durcan<sup>1</sup>, G C Wishart<sup>3</sup>, P Pharoah<sup>2</sup>, POSH Steering Group<sup>4</sup> and D Eccles<sup>1</sup>

<sup>1</sup>Cancer Sciences Academic Unit and University of Southampton Clinical Trials Unit, Faculty of Medicine, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK; <sup>2</sup>Department of Oncology, Strangeways Research Laboratory, University of Cambridge, Worts Causeway, Cambridge CB1 8RN, UK and <sup>3</sup>Faculty of Medical Science, Anglia Ruskin University, Cambridge CB1 1PT, UK

**Background:** Breast cancer is the most common cancer in younger women (aged  $\leq 40$  years) in the United Kingdom. PREDICT (<http://www.predict.nhs.uk>) is an online prognostic tool developed to help determine the best available treatment and outcome for early breast cancer. This study was conducted to establish how well PREDICT performs in estimating survival in a large cohort of younger women recruited to the UK POSH study.

**Methods:** The POSH cohort includes data from 3000 women aged  $\leq 40$  years at breast cancer diagnosis. Study end points were overall and breast cancer-specific survival at 5, 8, and 10 years. Evaluation of PREDICT included model discrimination and comparison of the number of predicted versus observed events.

**Results:** PREDICT provided accurate long-term (8- and 10-year) survival estimates for younger women. Five-year estimates were less accurate, with the tool overestimating survival by 25% overall, and by 56% for patients with oestrogen receptor (ER)-positive tumours. PREDICT underestimated survival at 5 years among patients with ER-negative tumours.

**Conclusions:** PREDICT is a useful tool for providing reliable long-term (10-year) survival estimates for younger patients. However, for more accurate short-term estimates, the model requires further calibration using more data from young onset cases. Short-term prediction may be most relevant for the increasing number of women considering risk-reducing bilateral mastectomy.

Breast cancer is the most common cancer in women in the United Kingdom, with around 50 000 newly diagnosed cases each year (Cancer Research UK, 2014). Approximately 4% of cases are in younger women (aged  $\leq 40$  years at diagnosis) yet it remains the most frequent malignancy in women of this age group (Cancer Research UK, 2014).

Determining the long-term outcome and potential benefits from systemic adjuvant treatments for early-stage breast cancer has been improved through the use of a number of currently available

predictive tools, including the Nottingham Prognostic Index (NPI), Adjuvant! and PREDICT. These tools have become increasingly sophisticated; incorporating a growing number of prognostic factors, enabling a move from categorising patients into broad prognostic groups to providing survival estimates at a patient level. However, many existing prognostic tools have only been validated in a small number of younger women, in which an overestimation of up to 30% in overall survival (OS) is reported in younger women (Engelhardt *et al*, 2014).

\*Correspondence: T Maishman; E-mail: [t.maishman@soton.ac.uk](mailto:t.maishman@soton.ac.uk)

<sup>4</sup>See Appendix

revised 10 December 2014; accepted 18 January 2015; published online 12 February 2015

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

## MATERIALS AND METHODS

**Existing tools.** The original NPI was based on 387 patients treated in a single institution and used tumour size, grade, and lymph node status to provide a prognostic index (Haybittle *et al*, 1982). The tool was later validated (Todd *et al*, 1987; D'Eredita *et al*, 2001) and updated to provide survival estimates by NPI group (Blamey *et al*, 2007a), and at a patient level (Blamey *et al*, 2007b).

Adjuvant! is an online decision aid tool, which was developed in 2001, based on 34 252 women aged 36–69 years from multiple institutions in the United States, and incorporated the patient's age, tumour size and grade, oestrogen receptor (ER) status, and number of positive nodes, to provide individual 10-year outcomes (Ravdin *et al*, 2001). The tool was later validated in patients from Canada (Olivotto *et al*, 2005) and the United Kingdom (Campbell *et al*, 2009). Adjuvant! makes an adjustment for women under 35 years, based on data from 415 patients; however, this applies only to patients with ER-positive tumours (Aebi *et al*, 2000).

The online breast cancer prognostic and treatment benefit tool, PREDICT, is used to estimate survival for individual patients based on known pathological prognostic factors, including age at diagnosis, mode of detection, tumour grade, tumour size, ER and nodal status, and type of therapy. The tool was developed in the United Kingdom, based on 5694 women aged 23–95 years (401 aged  $\leq 40$  years), from multiple institutions in East Anglia, and validated in two further cohorts: 5468 women aged 22–93 years from the West Midlands Cancer Intelligence Unit (Wishart *et al*, 2010); and the same cohort used for the validation of Adjuvant! (Wishart *et al*, 2011).

An updated version of PREDICT (<http://www.predict.nhs.uk>) now incorporates human epidermal growth factor receptor 2 (HER2) status (Wishart *et al*, 2012). The updated tool was validated using 1653 patients with information on HER2 from the same cohort of 3140 women from Canada used in the previous validations of PREDICT and Adjuvant!, and was shown to provide accurate OS and breast cancer-specific survival (BCSS) estimates at 10 years. The results demonstrated that the incorporation of HER2 status improved the tool, with predicted overall and breast cancer-specific deaths within 8.4% and 2.5% of those observed ( $P=0.05$  and  $P=0.60$ , respectively). However, the results indicated that PREDICT underestimated the number of deaths in women aged 20–35 years by 32% for both OS and BCSS (Wishart *et al*, 2012).

Although PREDICT was successfully validated in a large cohort, only 159 (9.6%) of these women from Canada were aged  $\leq 40$  years at diagnosis. However, young age at diagnosis is associated with lower survival rates compared with older patients (El Saghir *et al*, 2006; Adami *et al*, 1986). The reasons for the differing effects remain unclear; studies have shown that lower survival rates coupled with higher relapse rates in younger patients are independent of other known prognostic factors (de la Rochefordiere *et al*, 1993; Nixon *et al*, 1994). Younger patients may receive less significant survival benefits from hormonal therapy after adjuvant chemotherapy than older patients (Ahn *et al*, 2007). Furthermore, studies of premenopausal and perimenopausal women with breast cancer have found that chemotherapy alone was insufficient for younger patients, and perhaps tailored treatments should be investigated (Aebi *et al*, 2000; Colleoni *et al*, 2006).

The aim of this study was to investigate the performance of the updated PREDICT tool in a large cohort of women aged  $\leq 40$  years at breast cancer diagnosis in terms of model calibration and discrimination for both OS and BCSS.

**Study population.** POSH is a multicentre prospective observational cohort study of 3000 young women diagnosed with breast cancer in the United Kingdom between 2000 and 2008 (<http://www.southampton.ac.uk/medicine/research/posh.page>). The detailed study protocol was published in 2007 (Eccles *et al*,

2007), and the cohort has previously been described (Copson *et al*, 2013a).

Information obtained in the POSH cohort included age at diagnosis, ethnicity (Caucasian/white, black, Asian, other, missing/unknown), menopausal status (pre-, peri-, and post-menopausal, or missing/unknown), family history of breast cancer, ER, progesterone receptor, and HER2 status, histology, histological grade, tumour size, number of positive lymph nodes, lymphovascular invasion status, focality (localised, multifocal, missing/unknown), presentation, gene status (BRCA1, BRCA2, other, not tested/missing/unknown), and type of adjuvant therapy. This article presents analyses conducted on follow-up data from the POSH cohort received until 22 October 2013.

Study end points were OS and BCSS at 5, 8, and 10 years, where OS is defined as time from breast cancer diagnosis to death from any cause, and BCSS as time to death from breast cancer (deaths from other causes were censored at the time of last follow-up). The cohort includes data from 2827, 1843, and 597 patients with key prognostic information needed for the analyses for OS at 5, 8, and 10 years, respectively. Cause of death was missing for 5 patients leaving 2822, 1841, and 595 patients for analyses for BCSS, respectively.

Predicted OS and BCSS were calculated for each patient using PREDICT by investigators blinded to actual patient outcomes. Age at diagnosis, tumour size, and number of positive lymph nodes were entered as continuous variables. Categorical variables were used for presentation (screen-detected, symptomatic, or unknown), histological grade (1, 2, 3, or unknown), ER status (negative (ER $-$ ) or positive (ER $+$ )), HER2 (negative (HER2 $-$ ), positive (HER2 $+$ ), or unknown), and chemotherapy regimen (second generation or third generation).

**Model calibration and discrimination.** Using a similar approach to the methods applied in previous validations of PREDICT (Wishart *et al*, 2011, 2012), predicted OS and BCSS were compared with the corresponding observed OS and BCSS at 5, 8, and 10 years. Model calibration, a comparison between the predicted and observed mortality, was evaluated for the complete data set, by quartiles of the predicted mortality and also within strata (Wishart *et al*, 2012). As evidence suggests that existing prognostic tools do not perform well in Asian patients (Engelhardt *et al*, 2014), and black ethnicity was found to be an independent risk factor for reduced survival in younger women (Copson *et al*, 2013b), model calibration was also evaluated across ethnic groups. Goodness-of-fit tests were performed using  $\chi^2$  tests based on the number of predicted and observed events (4 d.f.). Model discrimination was assessed by calculating the area under the receiver-operator characteristic curve (AUC) and corresponding 95% confidence intervals for 5-, 8-, and 10-year predicted all-cause mortality and breast cancer-specific mortality. The AUC is the probability that the predicted mortality of a randomly selected patient who died is higher than that of a randomly selected patient who survived; the higher the AUC, the better the model is at identifying patients with a worse survival (Wishart *et al*, 2012).

All analyses were performed using STATA v12.1 (StataCorp LP, College Station, TX, USA).

## RESULTS

**Model calibration.** The demographic, tumour, and treatment information at baseline together with the predicted and observed all-cause OS at 5, 8, and 10 years are shown in Table 1. Overall, PREDICT did not perform well at 5 years, with an underestimation of the total number of deaths by 25% (455 vs 607,  $P<0.001$ ), and within most subgroups; most notably in patients with grade 2 tumours (58%; 67 vs 161,  $P<0.001$ ), 0–10 mm tumours (52%; 20 vs 42,  $P=0.001$ ), ER $+$  tumours (56%; 158 vs 362,  $P<0.001$ ), and

**Table 1. Observed and predicted 5-, 8-, and 10-year all-cause mortality by demographical, tumour, and treatment characteristics**

Characteristic	5 Years					8 Years					10 Years				
	N	O	P	D	%	N	O	P	D	%	N	O	P	D	%
Total	2827	607	455	-152	-25	1843	454	430	-24	-5.3	597	152	164	12	7.9
<b>Age at diagnosis (years)</b>															
18-25	40	7	5	-2	-28.6	21	4	4	0	0	8	1	1	0	0
26-30	258	62	46	-16	-25.8	167	48	43	-5	-10.4	55	15	16	1	6.7
31-35	864	210	152	-58	-27.6	580	168	146	-22	-13.1	203	60	57	-3	-5
36-40	1665	328	252	-76	-23.2	1075	234	237	3	1.3	331	76	90	14	18.4
<b>Menopausal status</b>															
Premenopause	2771	591	444	-147	-24.9	1801	440	420	-20	-4.5	581	148	162	14	9.5
Post-menopause	7	2	1	-1	-50	4	2	1	-1	-50	1	1	0	-1	-100
Perimenopause	5	2	1	-1	-50	5	2	2	0	0	3	1	1	0	0
Unknown	44	12	9	-3	-25	33	10	8	-2	-20	12	2	2	0	0
<b>Morphology</b>															
Ductal	2447	531	410	-121	-22.8	1606	396	388	-8	-2	520	132	149	17	12.9
Lobular	129	27	10	-17	-63	79	21	12	-9	-42.9	30	8	4	-4	-50
Other	216	42	31	-11	-26.2	133	32	27	-5	-15.6	40	12	10	-2	-16.7
Unknown	35	7	4	-3	-42.9	25	5	4	-1	-20	7	0	1	1	.
<b>Grade</b>															
1	156	5	4	-1	-20	98	5	4	-1	-20	31	3	2	-1	-33
2	929	161	67	-94	-58.4	614	121	75	-46	-38	200	44	31	-13	-30
3	1676	429	378	-51	-11.9	1097	322	346	24	7.5	351	103	129	26	25.2
Not graded/missing	66	12	7	-5	-41.7	34	6	5	-1	-16.7	15	2	2	0	0
<b>LV invasion</b>															
Negative	1374	182	156	-26	-14.3	873	142	144	2	1.4	290	47	59	12	25.5
Positive	1253	385	272	-113	-29.4	845	287	264	-23	-8	270	96	99	3	3.1
Unknown	200	40	27	-13	-32.5	125	25	22	-3	-12	37	9	7	-2	-22
<b>Node status</b>															
Negative	1370	161	128	-33	-20.5	869	111	115	4	3.6	266	39	42	3	7.7
Positive	1431	439	324	-115	-26.2	959	339	312	-27	-8	327	112	121	9	8
Unknown	26	7	4	-3	-42.9	15	4	3	-1	-25	4	1	1	0	0
<b>Tumour size</b>															
0-10	265	42	20	-22	-52.4	161	29	17	-12	-41.4	48	14	7	-7	-50
11-20	930	125	100	-25	-20	629	99	99	0	0	221	41	45	4	9.8
21-50	1229	302	233	-69	-22.8	798	228	219	-9	-3.9	244	78	79	1	1.3
50+	244	99	85	-14	-14.1	171	75	83	8	10.7	54	13	28	15	115.4
Unknown	159	39	16	-23	-59	84	23	12	-11	-47.8	30	6	5	-1	-16.7
<b>ER status</b>															
Negative	965	245	297	52	21.2	642	183	242	59	32.2	231	64	94	30	46.9
Positive	1862	362	158	-204	-56.4	1201	271	188	-83	-30.6	366	88	70	-18	-20.5
<b>Local Rx</b>															
BCS+RT	1310	188	168	-20	-10.6	872	146	159	13	8.9	275	59	59	0	0
Mast+RT	1001	313	228	-85	-27.2	648	231	213	-18	-7.8	208	70	85	15	21.4
Mast alone	445	87	48	-39	-44.8	277	65	47	-18	-27.7	101	22	18	-4	-18.2
Other	71	19	12	-7	-36.8	46	12	10	-2	-16.7	13	1	3	2	200
<b>Systemic Rx</b>															
None	46	4	5	1	25	28	3	5	2	66.7	6	1	1	0	0
Hormone	227	16	6	-10	-62.5	157	15	8	-7	-46.7	50	6	3	-3	-50
Chemo	986	261	297	36	13.8	659	198	248	50	25.3	237	71	94	23	32.4
Both	1568	326	147	-179	-54.9	999	238	170	-68	-28.6	304	74	66	-8	-10.8
<b>HER2 status</b>															
Negative	1773	383	255	-128	-33.4	1059	273	234	-39	-14.3	327	82	83	1	1.2
Positive	679	183	159	-24	-13.1	434	141	141	0	0	140	50	53	3	6
Borderline	40	10	6	-4	-40	33	10	7	-3	-30	14	5	4	-1	-20
Unknown	335	31	35	4	12.9	317	30	49	19	63.3	116	15	24	9	60
<b>Ethnicity</b>															
Caucasian/White	2582	547	413	-134	-24.5	1692	411	392	-19	-4.6	557	144	156	12	8.3
Black	110	36	21	-15	-41.7	74	27	20	-7	-25.9	17	5	4	-1	-20
Asian	84	15	14	-1	-6.7	55	11	13	2	18.2	16	3	3	0	0
Other	19	3	3	0	0	12	3	3	0	0	3	0	0	0	.
Unknown	32	6	5	-1	-16.7	10	2	2	0	0	4	0	1	1	.

Abbreviations: BCS = breast-conserving surgery; D = difference; ER = oestrogen receptor status; HER2 = human epidermal growth factor receptor 2; LV = lymphovascular invasion; N = number of patients; O = number of observed events; P = number of predicted events; RT = radiotherapy; Rx = Treatment.

patients receiving both hormone and chemotherapy (55%; 147 vs 326,  $P < 0.001$ ). Conversely, PREDICT overestimated the number of deaths by 21% (297 vs 245,  $P = 0.0009$ ) in patients with ER – tumours and patients receiving adjuvant trastuzumab (data not shown) by 25% (86 vs 69,  $P = 0.041$ ). Across HER2 and ethnicity subgroups, PREDICT was found to underestimate all-cause mortality in HER2 – tumours, Caucasian/white and Black ethnicity (33%; 255 vs 383,  $P < 0.001$ , 25%; 413 vs 547,  $P < 0.001$ , and 42%; 21 vs 36,  $P = 0.012$  respectively), and slightly underestimated for HER2 + and borderline tumours. In addition, when looking at chemotherapy regimen (data not shown) the tool was found to underestimate the number of deaths by 28% (297 vs 410,  $P < 0.001$ ) and by 18% (146 vs 177,  $P = 0.020$ ) for patients receiving second- and third-generation chemotherapy, respectively. Despite a poor performance across most subgroups at 5 years, the tool did perform well in certain subgroups; notably in patients with tumours  $> 50$  mm (14% underestimation; 85 vs 99,  $P = 0.159$ ), patients receiving both radiotherapy and breast conserving surgery (11% underestimation; 168 vs 188,  $P = 0.145$ ), and in Asian patients (7% underestimation; 14 vs 15,  $P = 0.796$ ).

At 8 years, the performance of PREDICT considerably improved, with the difference between the predicted deaths within 6% of those observed across the entire cohort (430 vs 454,  $P = 0.260$ ), and the tool performing well across most subgroups. Notable improvements to the predicted all-cause mortality from 5 to 8 years were found in patients aged 36–40 years (23%,  $P < 0.001$ , to 1%,  $P = 0.845$ ), patients with a ductal morphology (23%,  $P < 0.001$ , to 2%,  $P = 0.688$ ), negative nodal status (21%,  $P = 0.009$ ,

to 4%,  $P = 0.704$ ), 21–50 mm tumours (23%,  $P < 0.001$ , to 4%,  $P = 0.551$ ), patients who had radiotherapy and a mastectomy (27%,  $P < 0.001$ , to 8%,  $P = 0.236$ ), and patients receiving second- or third-generation chemotherapy (28%,  $P < 0.001$ , to 7%,  $P = 0.201$  and 18%,  $P = 0.020$ , to 4%,  $P = 0.705$ , respectively). The performance of the tool had also greatly improved across all ethnicity subgroups and in patients with HER2 + and borderline tumours. The number of predicted deaths remained much lower than those observed for patients with grade 2 tumours, 0–10 mm tumours, patients with a lobular morphology, and patients receiving both hormone and chemotherapy. There was also greater disparity in the predicted vs observed number of deaths at 8 years compared with 5 years when looking at patients receiving adjuvant trastuzumab and at ER subgroups, with all-cause mortality overestimated for patients receiving adjuvant trastuzumab by 58% (52 vs 33,  $P = 0.001$ ), and overestimated for ER – tumours by 32% (242 vs 183,  $P < 0.001$ ) and underestimated for ER + tumours by 31% (188 vs 271,  $P < 0.001$ ).

In contrast to the underestimation of the total number of deaths at 5 and 8 years, PREDICT was found to overestimate the number of deaths at 10 years by 8% (164 vs 152,  $P = 0.330$ ). The number of deaths remained overestimated for patients with ER – tumours (47%; 94 vs 64,  $P < 0.001$ ), grade 3 tumours (25%; 129 vs 103,  $P = 0.010$ ), and tumours larger than 50 mm (115%; 28 vs 13,  $P < 0.001$ ). In addition, the tool was found to overestimate all-cause mortality in patients aged 36–40 years (18%; 90 vs 76,  $P = 0.108$ ), patients with grade 3 tumours (25%; 129 vs 103,  $P = 0.010$ ) and tumours without lymphovascular invasion (26%; 59 vs 47,

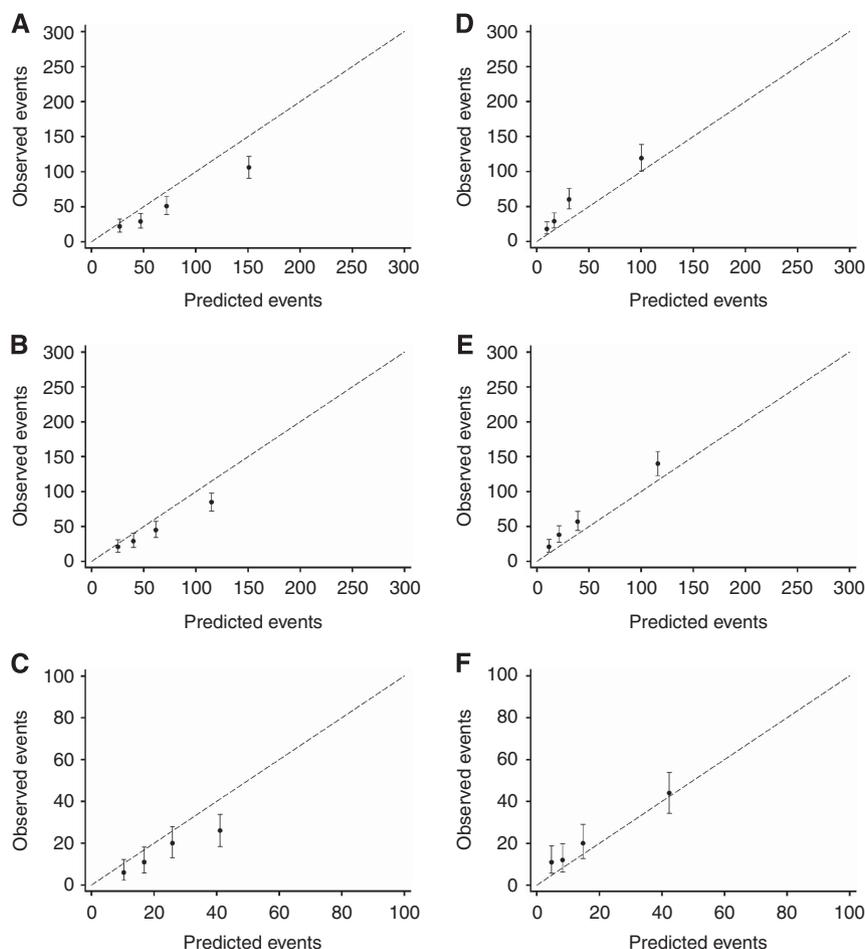


Figure 1. Calibrations plots of observed OS outcomes with 95% confidence intervals against predicted outcomes by quartiles of the predicted risk, by ER status. OS outcomes for patients with ER – tumours at (A) 5 years, (B) 8 years, and (C) 10 years, and OS outcomes for patients with ER + tumours at (D) 5 years, (E) 8 years, and (F) 10 years.

$P=0.080$ ), patients receiving chemotherapy alone (32%; 94 vs 71,  $P=0.006$ ), and patients receiving third-generation chemotherapy (31%; 42 vs 32,  $P=0.077$ ). Nevertheless, the 10-year estimates predicted by the tool had much improved across most subgroups, including ethnicity subgroups and HER2 subgroups.

The number of predicted vs observed breast cancer-specific deaths was very similar to the results for OS, with no notable differences between the OS and BCSS results when comparing these both by year (5, 8, and 10 years) and across subgroups (Supplementary Table 1).

The comparison between the predicted and observed all-cause mortality by quartiles of the predicted risk both across years and split by ER status is presented in Figure 1. PREDICT was found to overestimate the total number of deaths for patients with ER- tumours but underestimate for the ER+ subgroup. Goodness-of-fit test  $P$ -values identified that PREDICT performed better at 10 years compared with 5 and 8 years ( $P<0.001$ ,  $P=0.00284$ , and  $P=0.0295$  for ER- tumours at 5, 8, and 10 years, respectively, and  $P<0.001$ ,  $P<0.001$  and  $P=0.0183$  for ER+ tumours).

The findings were almost identical when comparing the predicted and observed breast cancer-specific mortality by year and ER status (Supplementary Figure 1). Moreover, when comparing the predicted and observed all-cause mortality by ER status for patients with HER2- tumours only (Figure 2) and HER2+ tumours only (Figure 3), the pattern of overestimating mortality in patients with ER- tumours and underestimating

mortality in ER+ tumours remained evident across years. The overestimation of the tool was also apparent when comparing the number of predicted vs observed all-cause deaths in patients with both ER- and HER2+ tumours (data not shown), with an overestimation at 5, 8, and 10 years of 38% (101 vs 73,  $P=0.001$ ), 39% (78 vs 56,  $P=0.003$ ), and 68% (32 vs 19,  $P=0.003$ ), respectively. For patients with both ER- and HER2- tumours, the overestimation was not as large (8% ( $P=0.301$ ), 18% ( $P=0.060$ ), and 22% ( $P=0.188$ ), respectively). For patients with both ER+ and HER2- tumours, large differences between the predicted and observed deaths could be seen at 5 and 8 years but not at 10 years. Underestimation was prominent in patients with both ER+ and HER2+ tumours, with an underestimation at 5, 8, and 10 years of 46% (59 vs 101,  $P<0.001$ ), 26% (63 vs 85,  $P=0.017$ ), and 32% (21 vs 31,  $P=0.072$ ), respectively.

**Model discrimination.** PREDICT provided a reasonably high degree of discrimination for OS across years (data not shown) and when splitting by ER and HER2 status (Figure 4). Model discrimination was slightly better for patients with ER+ tumours across all years (AUC: 0.718 vs 0.730, 0.709 vs 0.748, 0.694 vs 0.724 for ER- vs ER+ tumours at 5, 8, and 10 years, respectively). Conversely, model discrimination was better for patients with HER2- tumours, particularly at 10 years (AUC: 0.724 vs 0.592 for HER2- vs HER2+ tumours). Similar findings were apparent for BCSS (Supplementary Figure 2).

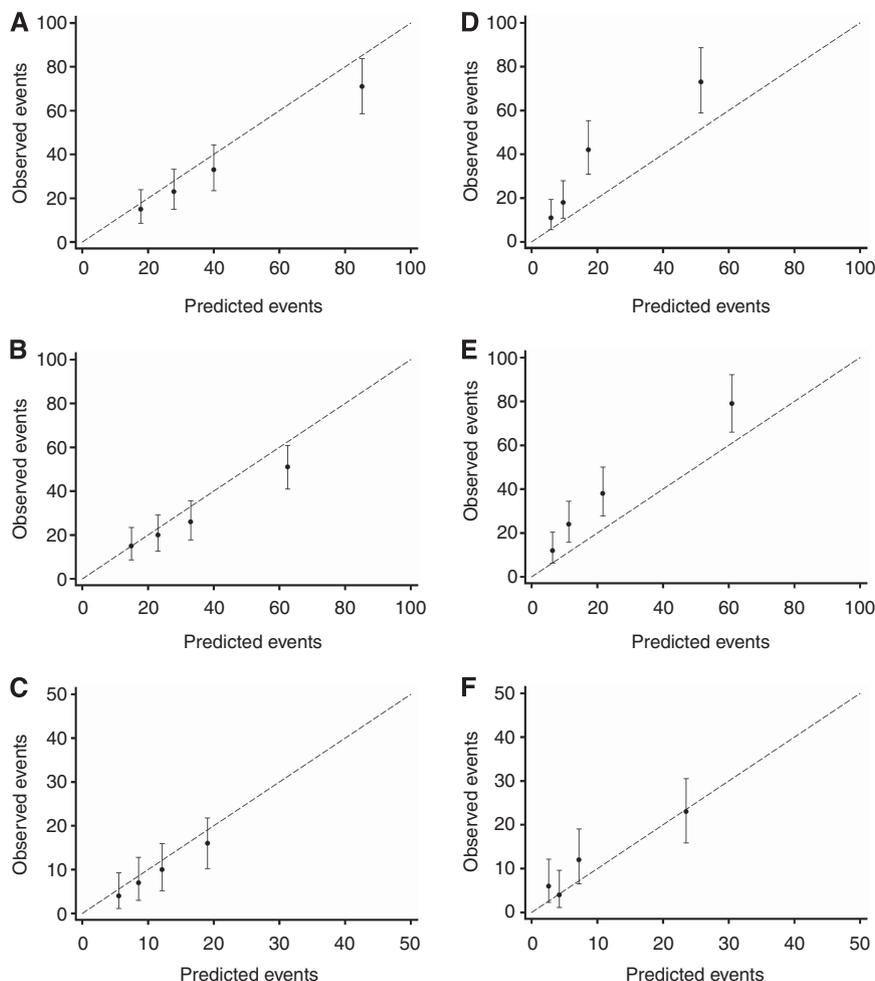


Figure 2. Calibrations plots of observed OS outcomes with 95% confidence intervals against predicted outcomes by quartiles of the predicted risk for HER2- patients only, by ER status. OS outcomes for patients with ER- tumours at (A) 5 years, (B) 8 years, and (C) 10 years, and OS outcomes for patients with ER+ tumours at (D) 5 years, (E) 8 years, and (F) 10 years.

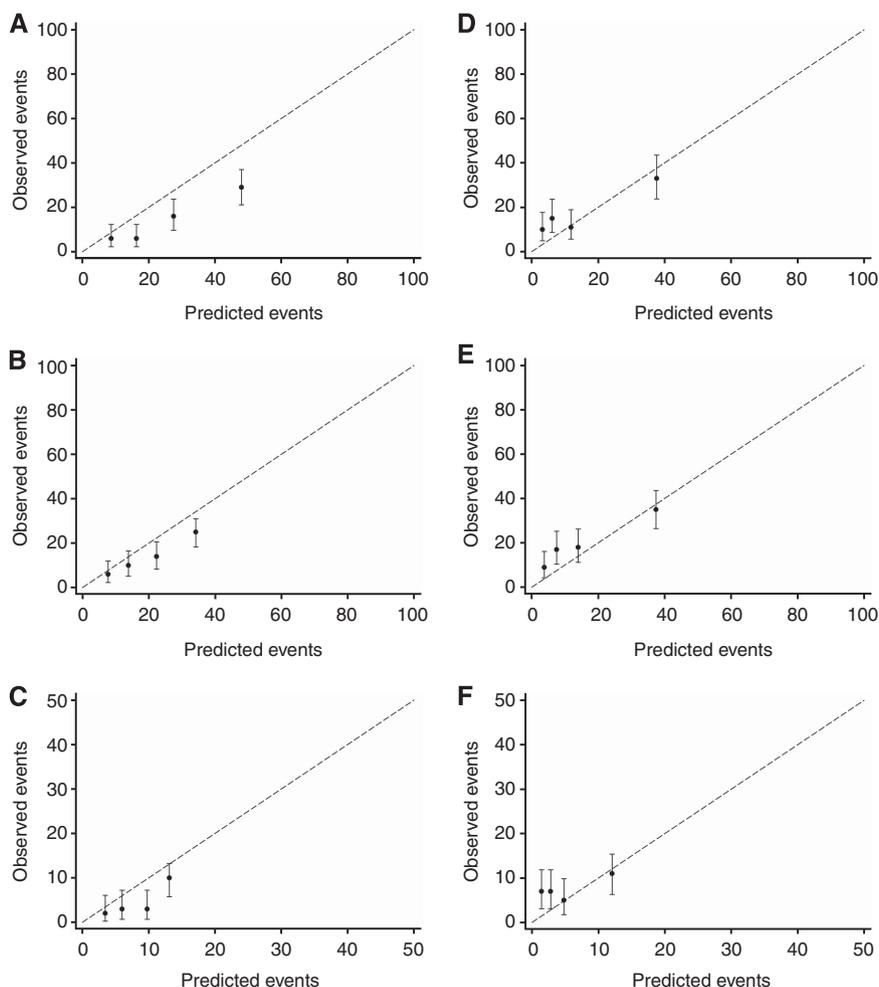


Figure 3. Calibrations plots of observed OS outcomes with 95% confidence intervals against predicted outcomes by quartiles of the predicted risk for HER2+ patients only, by ER status. OS outcomes for patients with ER- tumours at (A) 5 years, (B) 8 years, and (C) 10 years, and OS outcomes for patients with ER+ tumours at (D) 5 years, (E) 8 years, and (F) 10 years.

## DISCUSSION

This study has demonstrated that the prognostic tool PREDICT is able to provide accurate long-term (8- and 10-year) outcomes for younger women with breast cancer but provides only limited accuracy regarding short-term (5-year) survival, for both OS and BCSS.

High performance in terms of accurate predictions in the number of deaths was prominent in a number of subgroups across both years and survival type. These subgroups included patients receiving both radiotherapy and breast conserving surgery, patients with grade 1 tumours, HER2+ tumours, and patients with both ER- and HER2- tumours. Although the number of Asian patients included in the analyses at 5, 8, and 10 years was small (84, 55, and 16, respectively), PREDICT performed well in this subgroup of the POSH cohort, which was contrary to previous findings on a number of prognostic tools (Engelhardt *et al*, 2014). Our findings also demonstrated that in Caucasian/white and Black ethnicity subgroups, PREDICT was able to provide accurate long-term, but not short-term, estimates.

A key area in which PREDICT could improve its prognostic ability is in patients with ER- tumours. The number of predicted deaths across both years and survival type was overly pessimistic in this subgroup as demonstrated by the calibration plots (Figure 1 and Supplementary Figure 1). These results were contrary to the previously published validation of PREDICT for patients aged

20–85 years, in which the tool was found to accurately predict 10-year OS and BCSS in the ER- subgroup (Wishart *et al*, 2012).

Splitting the ER- subgroup by HER2 status, it was evident that there is some disparity between the PREDICT estimates for HER2- and HER2+ patients in this subgroup, with the tool providing accurate predictions for patients with both ER- and HER2- tumours across years and survival type, while consistently overestimating the number of deaths for patients with both ER- and HER2+ tumours. In contrast, the tool underestimated the number of deaths in patients with ER+ tumours, when looking not only at the subgroup as a whole but also when further splitting by HER2 status.

In relation to HER2 subgroups overall, the tool was able to predict the number of deaths for both OS and BCSS to within 1% of those observed for the HER2- subgroup, with corresponding 10-year AUCs of 0.724 and 0.718, respectively. PREDICT was not quite as reliable in terms of identifying patients with a worse survival in the HER2+ subgroup (AUC = 0.592 for both 10-year OS and BCSS), although the predicted number of deaths for OS and BCSS were within 6% of those observed. Fewer patients with HER2+ tumours were available in the 10-year evaluation ( $n = 140$  for both OS and BCSS) compared with HER2- tumours ( $n = 327$  and  $n = 325$ , respectively), which might have contributed to the reduction in accuracy of PREDICT for patients with HER2+ tumours.

In terms of OS and BCSS overall, the findings of our study demonstrated that PREDICT performs equally well, whereas the

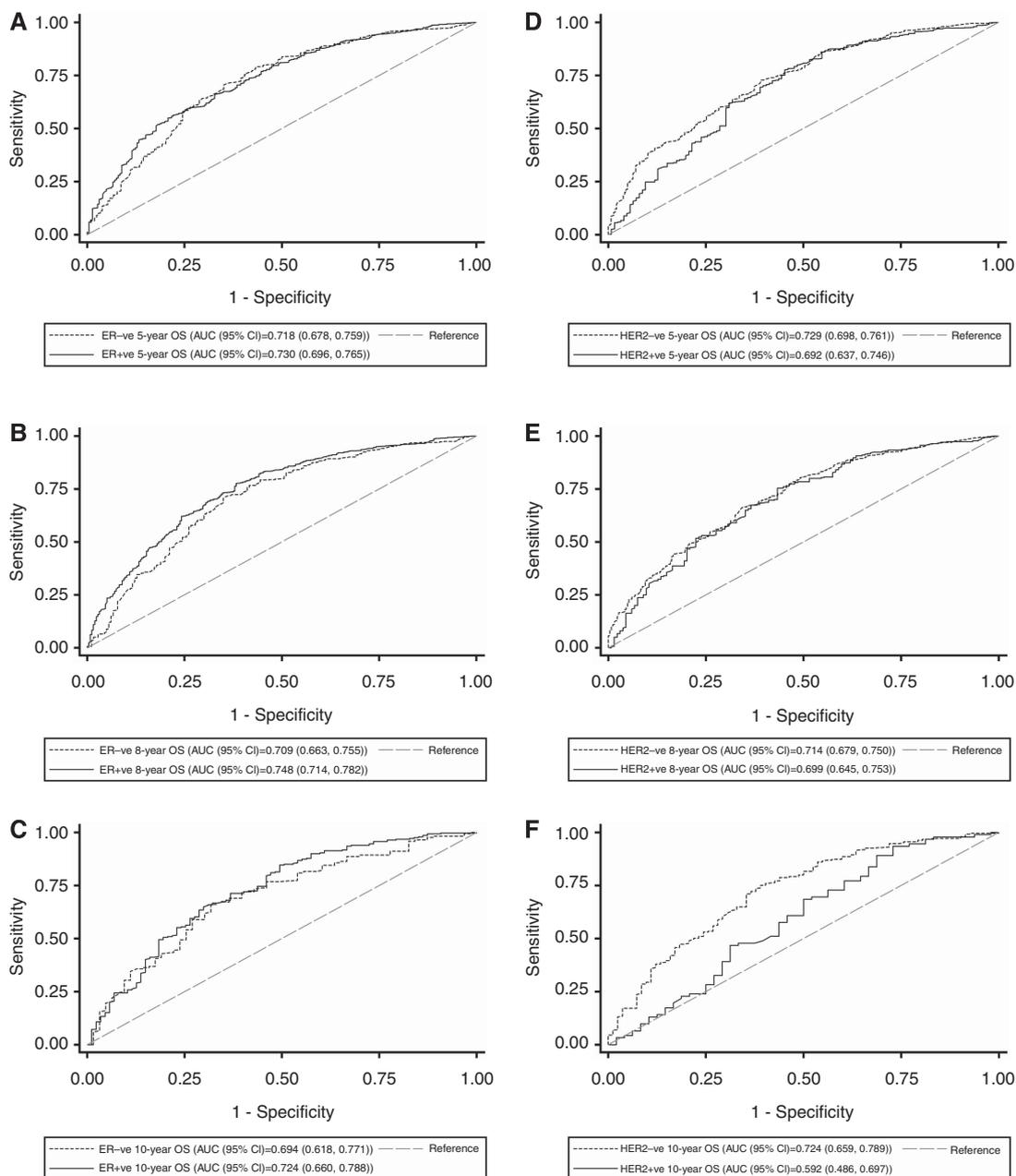


Figure 4. ROC curves for OS: split by ER status at (A) 5 years, (B) 8 years, and (C) 10 years; and split by HER2 status at (D) 5 years, (E) 8 years, and (F) 10 years.

Wishart *et al* validation showed that PREDICT is slightly better at providing BCSS estimates compared with OS estimates (Wishart *et al*, 2012). The validation also found that PREDICT underestimated the number of deaths at 10 years by 32% for patients aged 20–35 years (Wishart *et al*, 2012). In contrast to this, our study of outcomes for patients aged 18–40 years at diagnosis showed an overestimation in 10-year all-cause mortality of 8%. We explored this a little further by dichotomising age at diagnosis into two groups (data not shown), <35 and 35–40 years. Surprisingly, for the group aged <35 years, the number of deaths predicted (OS) was within 2% of those observed (60 vs 61,  $P=0.898$ ) and within 14% (104 vs 91,  $P=0.173$ ) for patients aged 35–40 years, indicating that differences in response to treatment between these two age groups could lead to some disparity between the performance of PREDICT between these two age groups.

There could be several reasons for the differences found in this study compared with other studies. The similarity of the OS and

BCSS results of our study is likely due to the fact that competing mortality does not play as important a role within this age group; excluding five patients with a missing/unknown cause of death, over 96% of deaths were due to breast cancer. In addition, PREDICT was developed to provide long-term (10-year) outcomes, not short-term, which might explain the reduced accuracy of the 5-year estimates. A possible reason for poor performance in patients aged 20–35 years is the low numbers on which the model was based; only 401 women aged  $\leq 40$  years were used in the development of PREDICT (131 patients had ER – and 270 ER + tumours). Furthermore, the number of patients aged  $\leq 40$  years evaluated in the validation of the enhanced PREDICT tool was relatively small ( $n = 159$ ; Wishart *et al*, 2012). It should be noted that poorer prediction in younger patients is a common finding across a number of other prognostic tools, which were found to overestimate OS by up to 30% in younger women (Engelhardt *et al*, 2014). As PREDICT did not perform well in a number of subgroups in our study, including the ER – subgroup, further

modifications to PREDICT using this larger data set could improve survival estimates for younger breast cancer patients. Furthermore, inclusion of the proliferation marker KI67 in the PREDICT model has led to a statistically significant improvement in function of the PREDICT model for ER+ patients (Wishart *et al*, 2014), which may also improve prognostication for younger patients.

There are some limitations to our study, which should be taken into account when interpreting the results. Less than half the patients from the POSH cohort had reached 10 years from diagnosis at the time of this analysis and so only a relatively small number could be included in the 10-year comparison ( $n = 607$ ). This is, however, still considerably larger than the number of women aged  $\leq 40$  years in the validation of PREDICT ( $n = 159$ ; Wishart *et al*, 2012). Our study also demonstrated an over-estimation in OS at 10 years so arguably this longer follow-up may not improve the estimates. Our data confirm a need for caution in extrapolating data from older cohorts to inform management in young patients with breast cancer. It also confirms the need to investigate treatment approaches in trials involving sufficiently large numbers of younger women, which would allow independent analysis to determine whether there are major outcome differences and to understand why. Trials of more treatment approaches specifically directed to younger patients with breast cancer should be investigated (Aebi *et al*, 2000; Colleoni *et al*, 2006; Narod, 2012). An additional limitation is that adjuvant Herceptin has been used in the United Kingdom routinely since 2005 so patients diagnosed with HER2+ disease before this date will only have received Herceptin in the metastatic setting. It is therefore possible that the outcome of HER2 patients in the POSH cohort as a whole is inferior to HER2 patients diagnosed and treated in the United Kingdom since 2005. However, the impact of this is likely to be minimal as the number of patients with HER2+ tumours diagnosed before 2005 ( $n = 298$ ) and since 2005 ( $n = 381$ ) included in this analysis is relatively small.

In conclusion, this study has demonstrated that PREDICT, a web-based tool that is easy to navigate for both patients and users, is a valuable resource in providing accurate and reliable long-term outcomes for younger patients. Although caution should currently be used when interpreting the short-term survival estimates in younger patients and the long-term estimates of younger patients with ER- tumours, it is intended that future modifications of PREDICT will include the incorporation of the POSH data set to allow for more robust estimates for younger women with breast cancer. Accurate prediction of outcome at both short- and long-term time points may be particularly important to women trying to determine the optimal timing of risk-reducing mastectomies.

## ACKNOWLEDGEMENTS

Funding for data collection and analysis for the POSH study was provided by Cancer Research UK (grants A7572, A11699, and C1275/A15956) and the study was sponsored by University Hospital Southampton NHS Foundation Trust. We thank the NIHR NCRN for supporting patient recruitment, and all patients participating in the POSH study.

## CONFLICT OF INTEREST

EC has received honoraria from Roche. The remaining authors declare no conflict of interest.

## REFERENCES

- Adami HO, Malke B, Holmberg L, Persson I, Stone B (1986) The relation between survival and age at diagnosis in breast cancer. *N Engl J Med* **315**: 559–563.
- Aebi S, Gelber S, Castiglione-Gertsch M, Gelber RD, Collins J, Thürlimann B, Rudenstam CM, Lindtner J, Crivellari D, Cortes-Funes H, Simoncini E, Werner ID, Coates AS, Goldhirsch A (2000) Is chemotherapy alone adequate for young women with oestrogen receptor-positive breast cancer? *Lancet* **355**: 1869–1874.
- Ahn SH, Son BH, Kim SW, Kim SI, Jeong J, Ko SS, Han W (2007) Poor outcome of hormone receptor-positive breast cancer at very young age is due to tamoxifen resistance: nationwide survival data in Korea—a report from the Korean Breast Cancer Society. *J Clin Oncol* **25**: 2360–2368.
- Blamey RW, Ellis IO, Pinder SE, Lee AH, Macmillan RD, Morgan DA, Robertson JF, Mitchell MJ, Ball GR, Haybittle JL, Elston CW (2007a) Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999. *Eur J Cancer* **43**: 1548–1555.
- Blamey RW, Pinder SE, Ball GR, Ellis IO, Elston CW, Mitchell MJ, Haybittle JL (2007b) Reading the prognosis of the individual with breast cancer. *Eur J Cancer* **43**: 1545–1547.
- Campbell HE, Taylor MA, Harris AL, Gray AM (2009) An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. *Br J Cancer* **101**: 1074–1084.
- Cancer Research UK (2014) <http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/commoncancers/#Top3> Accessed on March 2014.
- Colleoni M, Rotmensz N, Peruzzotti G, Maisonneuve P, Orlando L, Ghisini R, Viale G, Pruneri G, Veronesi P, Luini A, Intra M, Cardillo A, Torrisi R, Rocca A, Goldhirsch A (2006) Role of endocrine responsiveness and adjuvant therapy in very young women (below 35 years) with operable breast cancer and node negative disease. *Ann Oncol* **17**: 1497–1503.
- Copson E, Eccles B, Maishman T, Gerty S, Stanton L, Cutress RI, Altman DG, Durcan L, Simmonds P, Lawrence G, Jones L, Bliss J, Eccles D. POSH study steering group (2013a) Prospective Observational Study of Breast Cancer Treatment Outcomes for UK Women Aged 18–40 Years at Diagnosis: The POSH Study. *J Natl Cancer Inst* **105**: 978–988.
- Copson E, Maishman T, Gerty S, Eccles B, Stanton L, Cutress RI, Altman DG, Durcan L, Simmonds P, Jones L, Tapper W. POSH study steering group Eccles D (2013a) Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study. *Br J Cancer* **110**: 230–241.
- de la Rochefordiere A, Asselain B, Campana F, Scholl SM, Fenton J, Vilcoq JR, Durand JC, Pouillart P, Magdelenat H, Fourquet A (1993) Age as prognostic factor in premenopausal breast carcinoma. *Lancet* **341**: 1039–1043.
- D'Eredita G, Giardina C, Martellotta M, Natale T, Ferrarese F (2001) Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution. *Eur J Cancer* **37**: 591–596.
- Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG. POSH study steering group (2007) Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. *BMC Cancer* **7**: 160.
- El Saghir NS, Seoud M, Khalil MK, Charafeddine M, Salem ZK, Geara FB, Shamseddine AI (2006) Effects of young age at presentation on survival in breast cancer. *BMC Cancer* **6**: 194.
- Engelhardt EG, Garvelink MM, de Haes JH, van der Hoeven JJ, Smets EM, Pieterse AH, Stiggelbout AM (2014) Predicting and communicating the risk of recurrence and death in women with early-stage breast cancer: a systematic review of risk prediction models. *J Clin Oncol* **32**: 328–350.
- Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, Nicholson RI, Griffiths K (1982) A prognostic index in primary breast cancer. *Br J Cancer* **45**: 361–366.
- Narod SA (2012) Breast cancer in young women. *Nat Rev Clin Oncol* **9**: 460–470.
- Nixon AJ, Neuberger D, Hayes DF, Gelman R, Connolly JL, Schnitt S, Abner A, Recht A, Vicini F, Harris JR (1994) Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. *J Clin Oncol* **12**: 888–894.
- Olivetto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BD, Davis GJ, Chia SK, Gelmon KA (2005) Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol* **23**: 2716–2725.

- Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, Parker HL (2001) Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* **19**: 980–991.
- Todd JH, Dowle C, Williams MR, Elston CW, Ellis IO, Hinton CP, Blamey RW, Haybittle JL (1987) Confirmation of a prognostic index in primary breast cancer. *Br J Cancer* **56**: 489–492.
- Wishart G, Rakha E, Green E, Ellis I, Pharoah P (2014) Inclusion of KI67 significantly improves performance of the Predict prognostication and prediction model for early breast cancer (abstract). *Eur J Surg Oncol* **40**: 607.
- Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, Caldas C, Pharoah PD (2010) PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* **12**: R1.
- Wishart GC, Bajdik CD, Azzato EM, Dicks E, Greenberg DC, Rashbass J, Caldas C, Pharoah PD (2011) A population-based validation of the prognostic model PREDICT for early breast cancer. *Eur J Surg Oncol* **37**: 411–417.
- Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, Greenberg DC, Green AR, Gelmon KA, Kosma W-M, Olson JE, Beckmann MW, Winqvist R, Cross SS, Severi G, Huntsman D, Pylka K, Ellis I, Nielsen TO, Giles G, Blomqvist C, Fasching PA, Couch FJ, Rakha E, Foulkes WD, Blows FM, Bégin LR, van't Veer LJ, Southey M, Nevanlinna H, Mannermaa A, Cox A, Cheang M, Baglietto L, Caldas C, Garcia-Closas M, Pharoah PDP (2012) PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer* **107**: 800–807.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

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

## APPENDIX

POSH Steering group: Professor Diana Eccles, Dr Peter Simmonds, Professor Douglas G Altman, Dr Paul Pharoah, Professor Louise Jones, Professor Ros Eeles, Professor D Gareth Evans,

Professor Andrew Hanby, Professor Alistair M Thompson, Professor Shirley Hodgson, Mr Hisham Hamed, Dr Ruth Warren, and Professor Sunil Lakhani.