



This is a repository copy of *Bridging the Age Gap: a prognostic model that predicts survival and aids in primary treatment decisions for older women with oestrogen receptor-positive early breast cancer.*

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/163043/>

Version: Published Version

---

**Article:**

Ward, S.E. [orcid.org/0000-0002-3344-0913](https://orcid.org/0000-0002-3344-0913), Holmes, G.R. [orcid.org/0000-0002-8861-9638](https://orcid.org/0000-0002-8861-9638), Morgan, J.L. et al. (5 more authors) (2020) Bridging the Age Gap: a prognostic model that predicts survival and aids in primary treatment decisions for older women with oestrogen receptor-positive early breast cancer. *British Journal of Surgery*, 107 (12). pp. 1625-1632. ISSN 0007-1323

<https://doi.org/10.1002/bjs.11748>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



# Bridging the Age Gap: a prognostic model that predicts survival and aids in primary treatment decisions for older women with oestrogen receptor-positive early breast cancer

S. E. Ward<sup>1</sup>, G. R. Holmes<sup>1</sup>, J. L. Morgan<sup>2</sup>, J. W. Broggio<sup>4,5</sup>, K. Collins<sup>3</sup>, P. D. Richards<sup>1</sup>, M. W. R. Reed<sup>6</sup> and L. Wyld<sup>2</sup>

<sup>1</sup>Department of Health Economics and Decision Science, School for Health and Related Research, and <sup>2</sup>Department of Oncology and Metabolism, Medical School, University of Sheffield, and <sup>3</sup>Centre for Health and Social Care Research, Sheffield Hallam University, Sheffield, <sup>4</sup>National Cancer Registration and Analysis Service, Public Health England, Birmingham, <sup>5</sup>Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford, and <sup>6</sup>Brighton and Sussex Medical School, University of Sussex, Brighton, UK

Correspondence to: Dr Geoffrey R. Holmes, Department of Health Economics and Decision Science, School for Health and Related Research, University of Sheffield, Sheffield S1 4DA, UK (e-mail: geoff.r.holmes@sheffield.ac.uk)

**Background:** A prognostic model was developed and validated using cancer registry data. This underpins an online decision support tool, informing primary treatment choice for women aged 70 years or older with hormone receptor-positive early breast cancer.

**Methods:** Data from women diagnosed between 2002 and 2010 in the English Northern and Yorkshire and West Midlands regions were used to develop the model. Primary treatment options of surgery with adjuvant endocrine therapy or primary endocrine therapy were compared. Models predicting the hazard of breast cancer-specific mortality and hazard of other-cause mortality were combined to derive survival probabilities. The model was validated externally using data from the Eastern Cancer Registration and Information Centre.

**Results:** The model was developed using data from 23 842 women, and validated externally on a data set from 14 526 patients. The overall model calibration was good. At 2 and 5 years, predicted mortality from breast cancer and other causes differed from the observed rate by less than 1 per cent. At 5 years, there were slight overpredictions in breast cancer mortality (2629 predicted *versus* 2556 observed deaths;  $P = 0.142$ ) and mortality from all causes (6399 *versus* 6320 respectively;  $P = 0.583$ ). The discrepancy varied between subgroups. Model discrimination was 0.75 or above for all mortality measures.

**Conclusion:** A prognostic model for older women with oestrogen receptor-positive early breast cancer was developed and validated in the present study. This forms a basis for an online decision support tool (<https://agegap.shef.ac.uk/>).

Paper accepted 3 May 2020

Published online in Wiley Online Library (www.bjs.co.uk). DOI: 10.1002/bjs.11748

## Introduction

Breast cancer is the most common cancer to affect women, with 54 896 patients diagnosed in the UK in 2016<sup>1</sup>. Around one-third occur in women aged over 70 years. The standard of care for early breast cancer is surgical removal of the primary cancer, axillary surgery, and adjuvant therapies that may include chemotherapy, antihuman epidermal growth factor receptor (HER2) therapy, bisphosphonates, antioestrogens and radiotherapy. However, age-related practice varies widely; older women are less likely to receive adjuvant therapy than their younger counterparts<sup>2</sup>. The National Institute for Health and Care Excellence<sup>3</sup>

recommends that women with early breast cancer, irrespective of age, are treated with surgery and systemic therapy rather than endocrine therapy alone, unless significant co-morbidity precludes surgery. However, up to 40 per cent of older women with oestrogen receptor (ER)-positive cancer in the UK have historically received primary endocrine therapy.

Treatment with primary endocrine therapy is justified in some instances. It was shown to be effective in several trials in the 1980s, with no survival disadvantage relative to surgery, although rates of local control were suboptimal<sup>4</sup>. For some older women, surgery is associated

**Table 1** Patient characteristics used as co-variables in the two components of the prognostic model

	Description	Values	In BCM model	In OCM model
Treatment	Primary treatment	Surgery or primary endocrine therapy	✓	
Age	Age at diagnosis	Age above 70 years	✓	✓
Co-morbidity	CCI minus age component	0, 1, 2, ≥ 3	✓	✓
Frailty	ADL-related functional status	1, 2, 3, 4, 5		✓
Grade	Tumour grade	I, II, III	✓	
Nodal status	Positive nodes present	Yes, no	✓	
Size	Tumour size	Size in mm	✓	
Detection	Detection method	Screened or symptomatic	✓	
Deprivation	Average deprivation of patient postcode	Low (1), medium (2–4), high (5)	✓	

BCM, breast cancer mortality; OCM, other-cause mortality; CCI, Charlson Co-morbidity Index; ADL, activities of daily living.

with significant risks<sup>5</sup> and others may prefer minimal treatment<sup>6</sup>. Tang and colleagues<sup>5</sup> reported unacceptable rates of morbidity and mortality in surgically treated US nursing home residents. Older and less fit women are more likely to die from competing risks and may experience a significant decline in quality of life after surgery. Selecting the best treatment for an individual woman is complex and currently no tools exist to support the decision.

Models to inform clinician and patient decisions about adjuvant therapy after surgery already exist. They are based on clinical prognosticators for recurrence risk and breast cancer death. These include the Nottingham Prognostic Index<sup>7</sup>, PREDICT<sup>8</sup>, Adjuvant! Online<sup>9</sup>, OPTIONS and CancerMaths<sup>10</sup>. PREDICT is widely used in the UK, and is based on clinicopathological factors including tumour size, tumour grade, lymph node status, ER status, HER2 status and mode of detection. These models, however, consider the impact of adjuvant treatment after surgery rather than the initial decision regarding surgery itself. They might give an estimate of the expected outcome for an older woman having surgery, but not on the alternative of primary endocrine therapy. There is also evidence that PREDICT may be less accurate in 10-year outcome prediction in women aged over 75 years<sup>11</sup>. Existing prognostic models do not explicitly consider age-related factors such as co-morbidities and frailty. Co-morbidity is a strong predictor of competing mortality<sup>12</sup> and should be included when modelling an elderly population. Variations in functional and cognitive status and physiological reserves in older populations should also be considered<sup>12</sup>. In addition, deprivation level is usually linked to a higher burden of co-morbidity and frailty, and possibly also to undertreatment.

The aim of this study was to develop and validate a new prognostic model to inform the primary treatment choice (surgery or primary endocrine therapy) for

women aged at least 70 years with ER-positive early breast cancer.

## Methods

A prognostic model was developed comprising two sub-models: breast cancer mortality and other-cause mortality. Model parameters (*Table 1*) were selected based on literature review, exploratory investigation and expert advice. The hazard of breast cancer mortality was modelled as a function of patient age, co-morbidity score (Charlson Co-morbidity Index (CCI) without the age component) and deprivation level; tumour detection pathway; tumour diameter, grade and nodal status; and treatment choice. Registry data do not contain any information about concomitant co-morbidities and data on these were derived from linked Hospital Episode Statistics (HES) data collected during inpatient spells for patients with cancer. The hazard of other-cause mortality was modelled as a function of patient age, co-morbidity score and frailty. The subsequent model development procedure, including technical details of how these co-variables were incorporated into the model, is detailed in *Appendix S1* (supporting information).

The resulting breast cancer mortality model was a Royston–Parmar (RP) restricted cubic spline model with eight co-variables. RP models allow relaxation of the proportional hazards assumption associated with well known Cox models<sup>13</sup> and, being parametric, facilitate extrapolation of survival predictions as required for prognostic modelling. During the model building process, evidence was found for non-proportional hazards for some co-variables (*Table 1*). Interaction terms were also found to be statistically significant between treatment and three other co-variables: tumour grade, size and nodal status. The coefficients of the breast cancer submodel were estimated using the flexsurv library in the open

**Table 2 Patient characteristics in the West Midlands and Northern and Yorkshire training data set, and the Eastern Cancer Registration and Information Centre validation data set**

		Training data set (n = 23 842)	External validation data set (n = 14 526)
Region	NYCRIS	12 908 (54.1)	0 (0)
	WMCIU	10 934 (45.9)	0 (0)
	ECRIC	0 (0)	14 526 (100)
Age (years)	70–74	6399 (26.8)	3776 (26.0)
	75–79	6325 (26.5)	3813 (26.2)
	80–84	5513 (23.1)	3330 (22.9)
	85–89	3661 (15.4)	2234 (15.4)
	90–94	1482 (6.2)	1040 (7.2)
	95–99	408 (1.7)	307 (2.1)
	≥ 100	54 (0.2)	26 (0.2)
Deprivation category	1 (low)	3857 (16.2)	2882 (19.8)
	2	4867 (20.4)	3634 (25.0)
	3	4753 (19.9)	4025 (27.7)
	4	5060 (21.2)	2851 (19.6)
	5 (high)	5298 (22.2)	1134 (7.8)
	Missing	7 (0.0)	0 (0)
Detection	Screened	1345 (5.6)	1330 (9.2)
	Symptomatic	22 497 (94.4)	13 196 (90.8)

Values in parentheses are percentages unless indicated otherwise. NYCRIS, Northern and Yorkshire Cancer Registry and Information Service; WMCIU, West Midlands Cancer Intelligence Unit; ECRIC, Eastern Cancer Registration and Information Centre.

source software package R (R Foundation for Statistical Computing, Vienna, Austria)<sup>14</sup>.

The three co-variables in the other-cause mortality sub-model (age, co-morbidity and frailty) were modelled using the proportional hazards assumption. Frailty was approximated by a version of the activities of daily living (ADL) score<sup>15</sup>, represented by an integer value ranging from 0 to 5; 0 means no difficulties, and 5 means complete difficulty in the components eat, toilet, dress, transfer, bathe and walk. This variable was not recorded in the registry data, so a Markov chain Monte Carlo approach was adapted from Koissi and Högnäs<sup>16</sup>. This process inferred frailty weights, that is the probability of being at each ADL level for each patient, as well as model parameters that are the hazard ratios of other-cause mortality for each level (*Appendix S1*, supporting information). This estimation was carried out using the open source WinBUGS package<sup>17</sup> and the R2WinBUGS interface<sup>18</sup>.

Hazards predicted by the submodels were combined and transformed appropriately (*Appendix S1*, supporting information) to derive probabilities of death by 2 and 5 years (breast cancer-specific, other cause and all cause).

**Table 3 Tumour characteristics in the West Midlands and Northern and Yorkshire training data set, and the Eastern Cancer Registration and Information Centre validation data set**

		Training data set (n = 23 842)	External validation data set (n = 14 526)
Tumour stage	I	5140 (21.6)	3568 (24.6)
	II	8444 (35.4)	5360 (36.9)
	III	1871 (7.8)	1412 (9.7)
	IV	1511 (6.3)	1041 (7.2)
	Missing	6876 (28.8)	3145 (21.7)
Tumour size	< 10 mm	971 (4.1)	722 (5.0)
	10–19 mm	4584 (19.2)	3153 (21.7)
	20–49 mm	8486 (35.6)	5835 (40.2)
	≥ 50 mm	1369 (5.7)	1062 (7.3)
	Missing	8432 (35.4)	3754 (25.8)
Tumour grade	I	3118 (13.1)	1516 (10.4)
	II	10 479 (44.0)	6072 (41.8)
	III	5746 (24.1)	3586 (24.7)
	Missing	4499 (18.9)	3352 (23.1)
Node-positive	No	6548 (27.5)	4122 (28.4)
	Yes	5579 (23.4)	2807 (19.3)
	Missing	11 715 (49.1)	7597 (52.3)
ER-positive	No	5115 (21.5)	6525 (44.9)
	Yes	18 727 (78.5)	8001 (55.1)
HER2-positive	No	4758 (20.0)	4053 (27.9)
	Yes	850 (3.6)	656 (4.5)
	Missing	18 234 (76.5)	9817 (67.6)
Co-morbidity score	0	16 686 (70.0)	11 964 (82.4)
	1	1883 (7.9)	1451 (10.0)
	2	979 (4.1)	616 (4.2)
	≥ 3	550 (2.3)	397 (2.7)
	Missing	3744 (15.7)	98 (0.7)
Treatment	Primary endocrine therapy	10 471 (43.9)	6037 (41.6)
	Surgery	13 371 (56.1)	8489 (58.4)
Vital status	Censored	7076 (29.7)	5520 (38.0)
	Breast cancer death	7312 (30.7)	4008 (27.6)
	Other death	9454 (39.7)	4998 (34.4)
Mean survival (days)		1906	1768

Values in parentheses are percentages unless indicated otherwise. ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2.

Model training data, obtained from the West Midlands (West Midlands Cancer Intelligence Unit, WMCIU) and the Northern and Yorkshire (Northern and Yorkshire Cancer Registry and Information Service, NYCRIS) cancer registry regions, included all first diagnoses of invasive breast cancer in women aged 70 years and over between 2002 and 2010. The patient and disease variables were representative of wider UK national data in terms of age

**Table 4 Internal model validation: calibration and discrimination results for breast cancer and all-cause mortality at 2 and 5 years across the whole training data set**

	No. of patients	Observed mortality	Predicted mortality	Absolute difference, predicted – observed (%)	Relative difference, predicted versus observed (%)	AUC
2-year BCM	16 201	1016 (6.3)	1022 (6.3)	0	0.5	0.77
5-year BCM	16 194	2556 (15.8)	2629 (16.2)	0.4	2.8	0.75
2-year ACM	16 201	2477 (15.3)	2546 (15.7)	0.4	2.8	0.79
5-year ACM	16 194	6320 (39.0)	6399 (39.5)	0.5	1.2	0.80

Values in parentheses are percentages. Discrimination, calculated as area under the receiver operating characteristic (ROC) curve (AUC), indicates concordance between predicted and observed risk; a value of 0.5 represents no concordance and 1 represents perfect concordance. BCM, breast cancer mortality; ACM, all-cause mortality.

distribution, deprivation pattern, tumour stage and biological subtype distributions (Tables 2 and 3). The data from these regions were also of the highest available quality in terms of accuracy and completeness in comparison with wider UK data. Survival data were derived from death certificates from the Office for National Statistics, with a mean follow up of 5.2 years and a censoring date of 17 January 2017. Details of data preprocessing have been reported elsewhere<sup>19</sup> and are summarized in Appendix S1 and Table S1 (supporting information). For external validation, Eastern Cancer Registration and Information Centre (ECRIC) data were obtained on all first diagnoses of invasive breast cancer in women aged 70 years or more between 2002 and 2012, with a mean follow-up of 4.8 years. The majority of administrative censoring for the validation data occurred in January 2016 (Appendix S1, supporting information). Cancer registry data in the UK do not record co-morbidities or frailty. Co-morbidity was derived from linked records in the HES data set. HES records were searched from 18 months before the date of diagnosis and linkage was made using National Health Service (NHS) number, date of birth, sex and postcode.

Both data sets contained variables with a non-negligible number of missing values (Tables 2 and 3). Multiple imputation was used to create complete versions to avoid excluding patients with any missing data. Missing a variable is often dependent on patient characteristics, leading to potential exclusion bias. The distribution of these variables in patients with similar characteristics was used to impute values for the missing variables. To account for the uncertainty in this process, 15 imputations of both data sets were created. The analysis was done on each imputation, and results combined using Rubin's rules<sup>20</sup>. Details of the imputation process have been published previously<sup>19</sup> and are summarized in Appendix S1 and Table S2 (supporting information).

Validation comprised assessment of calibration and discrimination. Calibration tests whether the model predicts

the correct number of deaths over a given interval. The time points chosen for validation were 2 and 5 years, owing to the age and typical frailty of the women who will use the tool. For 2-year calibration, for instance, all women with complete 2-year follow-up (excluding those lost to follow-up before 2 years) were selected and the sum of 2-year all-cause mortality predictions was compared with the number of observed deaths. Discrimination measures whether patients with a higher predicted probability of death are, on average, those who die more frequently. Discrimination is calculated as the area under the receiver operating characteristic (ROC) curve (AUC); an AUC of 0.5 represents the equivalent of randomly allocated mortality probabilities, and an area of 1 represents perfect concordance between probabilities and outcomes<sup>8,21</sup>. All validation results were produced by averaging over the individual results for the 15 imputed data sets and also by weighted averaging of mortality predictions using the ADL level frailty weights inferred for each patient.

## Results

Patient and tumour characteristics for the training and external data sets are shown in Tables 2 and 3. The cohorts were similar, except for a higher proportion of screened patients (9.2 versus 5.6 per cent) and a lower proportion of patients in the high-deprivation group (7.8 versus 22.2 per cent) in the external versus training data sets. After preprocessing, 18 727 (78.5 per cent) patients in the training data set were classified as ER-positive; of these, 10 085 women (53.9 per cent) had surgery and 8642 (46.1 per cent) received primary endocrine therapy.

## Hazard ratios

Hazard ratios for surgery versus primary endocrine therapy groups were patient-dependent. Values for a range of subgroups are shown in Table S3 (supporting information).

**Table 5** Internal calibration check of the model: comparison of observed and predicted all-cause mortality at 2 and 5 years for subgroups of patients

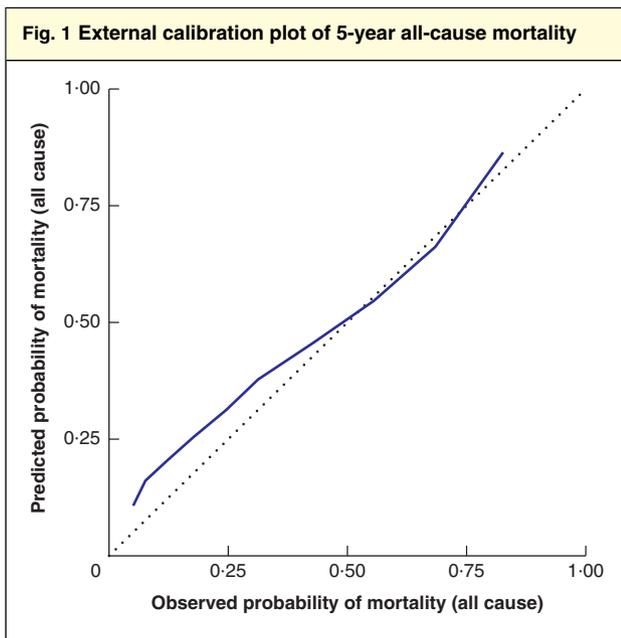
	2 years					5 years				
	No. of patients	Observed mortality	Predicted mortality	Absolute mortality difference (%) <sup>*</sup>	Relative mortality difference (%) <sup>†</sup>	No. of patients	Observed mortality	Predicted mortality	Absolute mortality difference (%) <sup>*</sup>	Relative mortality difference (%) <sup>†</sup>
<b>All patients</b>	16 201	2477 (15.3)	2546	0.4	2.8	16 194	6320 (39.0)	6399	0.5	1.2
<b>Treatment</b>										
Surgery	9921	712 (7.2)	913	2.0	28.1	9915	2274 (22.9)	2675	4.0	17.6
Primary endocrine therapy	6280	1765 (28.1)	1633	-2.1	-7.5	6279	4046 (64.4)	3724	-5.1	-8.0
<b>Age (years)</b>										
70-74	4445	284 (6.4)	275	-0.2	-3.4	4443	901 (20.3)	870	-0.7	-3.4
75-79	4391	441 (10.0)	463	0.5	5.2	4389	1307 (29.8)	1351	1.0	3.4
80-84	3703	596 (16.1)	632	1.0	6.1	3700	1581 (42.8)	1668	2.3	5.4
85-89	2413	652 (27.1)	655	0.0	0.2	2413	1524 (63.2)	1511	-0.5	-0.8
≥ 90	1249	504 (40.4)	521	1.4	3.5	1249	1007 (80.6)	999	-0.6	-0.8
<b>Node-positive</b>										
No	9404	1083 (11.5)	1208	1.3	11.5	9402	2981 (31.7)	3215	2.5	7.8
Yes	6797	1394 (20.5)	1338	-0.8	-4.0	6792	3339 (49.2)	3184	-2.3	-4.6
<b>Co-morbidity score</b>										
0	13 862	1793 (12.9)	1877	0.6	4.7	13 856	4883 (35.2)	5005	0.9	2.5
1	1560	416 (26.7)	377	-2.5	-9.3	1560	881 (56.5)	835	-3.0	-5.3
2	465	137 (29.5)	147	2.2	7.6	465	316 (68.0)	304	-2.7	-3.9
≥ 3	314	131 (41.7)	145	4.5	10.9	313	240 (76.7)	255	4.5	6.0
<b>Detection</b>										
Screen-detected	1110	28 (2.5)	44	1.4	53.9	1110	88 (7.9)	147	5.3	66.3
Symptomatic	15 091	2449 (16.2)	2502	0.4	2.2	15 084	6232 (41.3)	6252	0.1	0.3
<b>Tumour grade</b>										
I	3137	414 (13.2)	437	0.7	5.5	3137	1067 (34.0)	1106	1.2	3.6
II	9596	1417 (14.8)	1491	0.8	5.2	9592	3671 (38.3)	3753	0.9	2.2
III	3468	646 (18.6)	618	-0.8	-4.3	3465	1582 (45.7)	1540	-1.2	-2.6
<b>Tumour size (mm)</b>										
< 10	977	78 (8.0)	92	1.4	18.4	976	217 (22.2)	256	4.1	18.4
10-19	4874	509 (10.4)	576	1.4	13.2	4873	1365 (28.0)	1555	3.9	13.9
20-29	5128	734 (14.3)	799	1.2	8.8	5126	1975 (38.5)	2024	0.9	2.5
30-49	3689	720 (19.5)	694	-0.7	-3.5	3687	1807 (49.0)	1690	-3.2	-6.5
≥ 50	1533	436 (28.4)	385	-3.3	-11.7	1532	956 (62.4)	874	-5.3	-8.6
<b>Deprivation category</b>										
Low	2673	313 (11.7)	365	2.0	16.7	2672	867 (32.4)	947	3.0	9.3
Medium	9971	1538 (15.4)	1578	0.4	2.6	9966	3855 (38.7)	3952	1.0	2.5
High	3557	626 (17.6)	603	-0.7	-3.7	3556	1598 (44.9)	1500	-2.7	-6.1

Values in parentheses are percentages. <sup>\*</sup>Predicted - observed; <sup>†</sup>predicted *versus* observed.

### Internal validation

Overall calibration of the model was good (Table 4). At 2 and 5 years, predicted breast cancer and other-cause mortality differed from observed rates by less than 1 per cent in all instances. At 5 years, there was a slight overprediction of breast cancer mortality (2629 predicted *versus* 2556 observed deaths;  $P=0.142$ ) and

all-cause mortality (6399 *versus* 6320;  $P=0.583$ ). The AUC representing discrimination was 0.75 or above for all mortality measures (Table 4). Performance was similar to that of PREDICT (AUC 0.76-0.78)<sup>8</sup>. Fig. S1 (supporting information) shows predicted and observed 5-year all-cause mortality by deciles of observed mortality. Calibration was good for the intermediate deciles, but



Patients are divided into ten groups using deciles of mortality predictions. For each group, the average 5-year all-cause mortality prediction is plotted against the average observed mortality. A straight line of unit gradient represents perfect calibration.

less good for low-risk (overprediction) and moderately high-risk (under-prediction) deciles.

The degree of underestimation or overestimation of all-cause mortality varied between subgroups (*Table 5*). The 5-year relative mortality difference exceeded  $\pm 10$  per cent for the surgery subgroup, the screened population subgroup and two of the tumour size subgroups (less than 10 mm and 10–19 mm). The 2-year relative mortality difference exceeded  $\pm 10$  per cent for the surgery, node-negative, co-morbidity score  $\geq 3$  and screen-detected subgroups. All-cause mortality was also overpredicted for patients with smaller tumours and underpredicted for those with the largest tumours (over 50 mm). At both 2 and 5 years, mortality was overpredicted in the surgery subgroup and underpredicted in the primary endocrine therapy subgroup. These subgroup differences were driven by the other-cause mortality estimates (*Table S4*, supporting information), with 13.4 and 17.3 per cent underprediction for patients who had primary endocrine therapy at 2 and 5 years respectively.

### External validation

Overall calibration was also good in the external data set (*Table S5*, supporting information). Five-year predicted breast cancer mortality exceeded the observed rate in the training data by 0.5 per cent, and the difference for

all-cause mortality was 0.1 per cent. Two-year predicted all-cause mortality exceeded the observed rate by 1.2 per cent, whereas at 5 years there was a small underprediction ( $-0.9$  per cent). *Fig. 1* shows predicted and observed 5-year all-cause mortality by deciles of observed mortality. Calibration was good for the higher-risk deciles, but there was some overprediction for the low-risk deciles. Discrimination results for 2- and 5-year breast cancer mortality and all-cause mortality in the external validation data set had AUC values in the range 0.75–0.80 (*Table S5*, supporting information). The results of external subgroup validation are shown in *Table S6* (supporting information).

### Discussion

A prognostic model for women aged 70 years or older with ER-positive early breast cancer was developed and validated. This model was targeted specifically at supporting the decision of whether to undergo surgery or opt for primary endocrine therapy. The model was developed using a large cohort of patients from the West Midlands (WMCIU) and the Northern and Yorkshire (NYCRIS) cancer registries. These together cover around 25 per cent of the population in England, and are regarded as being broadly representative of the UK demographic distribution. The model was validated using data from the Eastern Cancer Registry (ECRIC) in England. The model was shown to be well calibrated for all-cause mortality and to have good discrimination, with similar performance to other prognostic models in early breast cancer, such as PREDICT. The model additionally performed well on the external validation data set. The Age Gap prognostic model provides outputs for both breast cancer mortality and other-cause mortality. This is an important issue in older patients, who have an increased risk of dying from other causes. By age 85 years, around three-quarters of deaths are from causes other than breast cancer.

The findings of this analysis are not directly comparable to those of other predictive models in breast cancer. PREDICT and other commonly used models focus on decisions around adjuvant therapy following breast cancer surgery, rather than whether or not a patient will benefit from surgery itself. These models are typically trained on a mixed-age population, and not designed to deal with the greater other-cause mortality rates among patients aged 70 years or more. In contrast, the Age Gap model is targeted at older women for whom the benefits *versus* harms of surgery are more complex, owing to co-morbidities and frailty.

In this model, other-cause mortality is underpredicted for patients treated with primary endocrine therapy (absolute mortality difference -6.8 per cent) (Table S4, supporting information). The other-cause mortality model has taken the first step to incorporating the impact of co-morbidities and frailty on individual-patient predictions of mortality, but the prediction is currently subject to limitations. Cancer registry data in the UK do not record co-morbidities or frailty. Here, co-morbidity was derived from linked records in the HES data set. Data were available from HES only if a patient had a hospital inpatient or day-case admission preceding their cancer diagnosis. This methodology relies heavily on the accuracy of coding within HES. HES data are likely to under-record co-morbidities in patients who have chronic co-morbidities managed in the community or outpatient setting, such as diabetes or dementia. On average, women are more likely to receive primary endocrine therapy if they are older and have chronic co-morbidities<sup>22,23</sup>. Under-reporting of co-morbidities in patients undergoing primary endocrine therapy may, at least in part, contribute to the underprediction of the other-cause mortality for such patients.

Validation of the model was conducted at 2 and 5 years after diagnosis as sufficient 10-year uncensored follow-up data were not available. This also reflects the decision that these prognostic timescales are the most appropriate for the majority of women in this older age group, for many of whom a 10-year prediction may not give a positive message. Further validation will be carried out when full data are available, and presentation of 10-year predictions will be reviewed in the light of user and patient input.

No measure of frailty is included in cancer registry data sets. Frailty was approximated by a version of the ADL score as used by Stineman and colleagues<sup>15</sup>. Values from this US study were used to form priors for the parameters in the developed model. The mean age of participants, 77.4 years, was similar to that in the UK population. However, the distributions of ADL level in these study participants may not fully reflect the UK population of older women with breast cancer. Further validation of the current methodology will be possible using data collected in the prospective Bridging the Age Gap cohort study, which includes co-morbidities and frailty data, once sufficiently mature survival data become available.

The present analysis benefits from the large retrospective data set of patients with breast cancer. This reduces the biases from exclusion criteria of many RCTs, which often include both older age and complex co-morbidity. Use of routinely collected data, however, resulted in a high proportion of missing values, especially for patients in the

non-surgical group. For instance, tumour size and clinical node status based on imaging was not clearly recorded in these patients. No perfect method exists for obtaining these missing data. Multiple imputation, however, is less prone to bias than complete-case analysis and/or treating 'missing' as a category. It also allows the propagation of uncertainty owing to missing data into the estimates of co-variable effects.

The prognostic model is part of a decision support intervention (DESI) including an online tool (<https://agegap.shef.ac.uk/>) with two other patient-facing decision aids<sup>24</sup>, tailored to the information needs and preferences of older women<sup>25,26</sup> in decision-making between surgery and primary endocrine therapy. This DESI has been developed for women whose choice between surgery and primary endocrine therapy is not clear cut. The online tool predictions are generated by the model described in the present study. The co-morbidity score is determined by entering a history of patient co-morbidities via a tick-box list. The frailty score is similarly generated by entering the level of difficulty experience in six functional domains. In the current version, for reasons of patient sensitivity, the deprivation level is not entered and a medium level is assumed for all patients. The other model inputs are straightforward.

A prospective cohort study, part of the Bridging the Age Gap in Breast Cancer project, will finally offer a more detailed data set to further assess the effects of surgery and primary endocrine therapy on survival in older women.

## Acknowledgements

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (grant reference number RP-PG-1209-10071). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. S.E. Ward and G.R. Holmes are joint first authors of this article. *Disclosure:* The authors declare no conflict of interest.

## References

- 1 Cancer Research UK. *Breast Cancer Incidence (Invasive) Statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-Zero> [accessed 24 July 2019].
- 2 National Audit of Breast Cancer in Older Patients. *National Audit of Breast Cancer in Older Patients: 2019 Annual Report. Results of the Prospective Audit in England and Wales for Women Diagnosed Between January 2014 and December 2017*; 2019. <https://www.hqip.org.uk/resource/national-audit-of-breast->

- cancer-in-older-patients-2019-annual-report/#.Xtkb8jpKhPY [accessed 24 July 2019].
- 3 National Institute for Health and Care Excellence (NICE). *Early and Locally Advanced Breast Cancer: Diagnosis and Management*. NICE guideline NG101. <https://www.nice.org.uk/guidance/ng101> [accessed 6 October 2019].
  - 4 Morgan JL, Reed MW, Wyld L. Primary endocrine therapy as a treatment for older women with operable breast cancer – a comparison of randomised controlled trial and cohort study findings. *Eur J Surg Oncol* 2014; **40**: 676–684.
  - 5 Tang V, Zhao SJ, Boscardin J, Sudore R, Covinsky K, Walter LC *et al*. Functional status and survival after breast cancer surgery in nursing home residents. *JAMA Surg* 2018; **153**: 1090–1096.
  - 6 Husain LS, Collins K, Reed M, Wyld L. Choices in cancer treatment: a qualitative study of the older women's (> 70 years) perspective. *Psychooncology* 2008; **17**: 410–416.
  - 7 Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC *et al*. A prognostic index in primary breast cancer. *Br J Cancer* 1982; **45**: 361–366.
  - 8 Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G *et al*. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010; **12**: R1.
  - 9 Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N *et al*. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001; **19**: 980–991.
  - 10 Engelhardt EG, Garvelink MM, de Haes JCJM, van der Hoeven JJM, Smets EMA, Pieterse AH *et al*. 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* 2014; **32**: 238–250.
  - 11 Candido Dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK *et al*. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res* 2017; **19**: 58.
  - 12 Kiderlen M, de Glas NA, Bastiaannet E, van de Water W, de Craen AJM, Guicherit OR *et al*. Impact of comorbidity on outcome of older breast cancer patients: a FOCUS cohort study. *Breast Cancer Res Treat* 2014; **145**: 185–192.
  - 13 Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002; **21**: 2175–2197.
  - 14 Jackson CH. Flexsurv: a platform for parametric survival modeling in R. *J Stat Softw* 2016; **70**: 1–33.
  - 15 Stineman MG, Xie DW, Pan Q, Kurichi JE, Zhang Z, Saliba D *et al*. All-cause 1-, 5-, and 10-year mortality in elderly people according to activities of daily living stage. *J Am Geriatr Soc* 2012; **60**: 485–492.
  - 16 Koissi M-C, Högnäs G. Using WinBUGS to study family frailty in child mortality, with an application to child survival in Ivory Coast. *Afr Populat Stud* 2005; **20**: 1–8.
  - 17 Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000; **10**: 325–337.
  - 18 Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw* 2005; **12**: 1–16.
  - 19 Ward SE, Richards PD, Morgan JL, Holmes GR, Broggio JW, Collins K *et al*. Omission of surgery in older women with early breast cancer has an adverse impact on breast cancer-specific survival. *Br J Surg* 2018; **105**: 1454–1463.
  - 20 Rubin DB. *Multiple Imputation for Survey Nonresponse*. Wiley: New York, 1987.
  - 21 Gray E, Marti J, Brewster DH, Wyatt JC, Hall PS; SATURNE Advisory Group. Independent validation of the PREDICT breast cancer prognosis prediction tool in 45 789 patients using Scottish Cancer Registry data. *Br J Cancer* 2018; **119**: 808–814.
  - 22 Richards P, Ward S, Morgan J, Lagord C, Reed M, Collins K *et al*. The use of surgery in the treatment of ER+ early stage breast cancer in England: variation by time, age and patient characteristics. *Eur J Surg Oncol* 2016; **42**: 489–496.
  - 23 Bates T, Evans T, Lagord C, Monypenny I, Kearins O, Lawrence G. A population based study of variations in operation rates for breast cancer, of comorbidity and prognosis at diagnosis: failure to operate for early breast cancer in older women. *Eur J Surg Oncol* 2014; **40**: 1230–1236.
  - 24 Lifford KJ, Edwards A, Burton M, Harder H, Armitage F, Morgan EL *et al*. Efficient development and usability testing of decision support interventions for older women with breast cancer. *Patient Prefer Adherence* 2019; **13**: 131–143.
  - 25 Burton M, Collins KA, Lifford KJ, Brain K, Wyld L, Caldron L *et al*. The information and decision support needs of older women (>75 yrs) facing treatment choices for breast cancer: a qualitative study. *Psychooncology* 2015; **24**: 878–884.
  - 26 Burton M, Kilner K, Wyld L, Lifford KJ, Gordon F, Allison A *et al*. Information needs and decision-making preferences of older women offered a choice between surgery and primary endocrine therapy for early breast cancer. *Psychooncology* 2017; **26**: 2094–2100.

### Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.