

**Keywords:** breast cancer survival; ethnicity; stage; grade; receptors

# Short-term breast cancer survival in relation to ethnicity, stage, grade and receptor status: national cohort study in England

Henrik Møller<sup>\*1,2,3</sup>, Katherine Henson<sup>2</sup>, Margreet Lüchtenborg<sup>1,2</sup>, John Broggio<sup>2</sup>, Jackie Charman<sup>2</sup>, Victoria H Coupland<sup>2</sup>, Elizabeth Davies<sup>1</sup>, Ruth H Jack<sup>2</sup>, Richard Sullivan<sup>1</sup>, Peter Vedsted<sup>3</sup>, Kieran Horgan<sup>4,5</sup>, Neil Pearce<sup>6</sup> and Arnie Purushotham<sup>1</sup>

<sup>1</sup>Cancer Epidemiology, Population and Global Health, King's College London, Bermondsey Wing, 3rd Floor, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK; <sup>2</sup>National Cancer Registration and Analysis Service, Public Health England, Skipton House, 80 London Road, London SE1 6LH, UK; <sup>3</sup>Research Centre for Cancer Diagnosis in Primary Care, Department of Public Health, Aarhus University, Nordre Ringgade 1, Aarhus C 8000, Denmark; <sup>4</sup>School of Medicine, University of Leeds, Leeds, West Yorkshire, LS9 7TF, UK; <sup>5</sup>Department of Breast Surgery, Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Beckett Street, Leeds, West Yorkshire, LS9 7TF, UK and <sup>6</sup>Departments of Medical Statistics and Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

**Background:** In the re-organisation of cancer registration in England in 2012, a high priority was given to the recording of cancer stage and other prognostic clinical data items.

**Methods:** We extracted 86 852 breast cancer records for women resident in England and diagnosed during 2012–2013. Information on age, ethnicity, socio-economic status, comorbidity, tumour stage, grade, morphology and oestrogen, progesterone and HER2 receptor status was included. The two-year cumulative risk of death from any cause was estimated with the Kaplan-Meier method, and univariate and multivariate Cox proportional hazards regressions were used to estimate hazard ratios (HR) and their 95% confidence intervals (95% CI). The follow-up ended on 31 December 2014.

**Results:** The completeness of registration for prognostic variables was generally high (around 80% or higher), but it was low for progesterone receptor status (41%). Women with negative receptor status for each of the oestrogen, progesterone and HER2 receptors (triple-negative cancers) had an adjusted HR for death of 2.00 (95%CI 1.84–2.17). Black women had an age-adjusted HR of 1.77 (1.48–2.13) compared with White women.

**Conclusions:** The excess mortality of Black women with breast cancer has contributions from socio-economic factors, stage distribution and tumour biology. The study illustrates the richness of detail in the national cancer registration data. This allows for analysis of cancer outcomes at a high level of resolution, and may form the basis for risk stratification.

Breast cancer is the most common cancer in the UK, with approximately 54 000 new cases diagnosed each year (Cancer Research UK, 2016). International comparisons have shown that breast cancer survival in England is lower than in comparable, industrialised countries (Møller *et al*, 2010; Coleman *et al*, 2011). The survival deficit is particularly manifest in the first months after

diagnosis, consistent with delayed diagnosis and the presence of a subset of patients with very advanced and rapidly fatal disease (Møller *et al*, 2010).

Black women with breast cancer have a particularly high risk of death. In California, Black and Hispanic women with breast cancer had a higher risk of death compared with non-Hispanic White

\*Correspondence: Professor H Møller; E-mail: henrik.moller@kcl.ac.uk

Received 15 July 2016; revised 7 September 2016; accepted 25 September 2016; published online 25 October 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16

women (Boyer-Chammard *et al*, 1999). A recent study from USA showed that Black women with breast cancer had a more advanced stage distribution and lower survival than White patients (DeSantis *et al*, 2016). In South East England, Black Caribbean and Black African patients had higher breast cancer-specific mortality than White patients (Jack *et al*, 2009). Data from a prognostic study of outcomes in sporadic and hereditary breast cancer in the UK showed that young Black women with breast cancer had poorer outcomes, compared with White and Asian women (Copson *et al*, 2014). It is not yet clear if these variations are due to socio-economic, biological, or other factors. Some studies have suggested that social, personal and biological factors may each contribute to a part of the excess mortality in Black women with breast cancer (Jack *et al*, 2013; Iqbal *et al*, 2015; Warner *et al*, 2015).

Breast tumours may express receptors of which the three most important are the oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor 2 receptor (HER2). These biological markers are considered when determining the most suitable treatment, alongside patient age, morphology, tumour size, tumour grade, lymph node status and lymphovascular invasion (Lakhani, 2012; Dawson *et al*, 2013). Survival outcomes were better in ER-positive and PR-positive patients (Fisher *et al*, 1988). HER2 is part of the epidermal growth factor receptor (EGFR) family and is over-expressed in 18–20% of breast cancers (Onitilo *et al*, 2009; Parise and Caggiano, 2014). There is a significant association between HER2 over-expression and poor prognosis, with decreased disease-free survival and overall survival in node-positive patients (Piccart *et al*, 2001). Some studies characterised breast cancers on the basis of combinations of ER, PR and HER2 and showed wide variation in survival from good prognosis in ER+, PR+ subtypes to low survival in the ER-, PR-, HER2- subtype (Onitilo *et al*, 2009; Parise and Caggiano, 2014). Breast cancer that are negative for all of ER, PR and HER2 (triple negative cancers) tends to exhibit a more aggressive pattern of disease and a proportion show greater resistance to conventional systemic chemotherapy (Dent *et al*, 2007; Foulkes *et al*, 2010).

Traditionally, tumour characteristics were poorly recorded in population-based cancer registries. In recent years, a high priority was given to the recording of cancer stage and other clinical data items by the National Cancer Registration and Analysis Service, who register all cancers diagnosed in England (The National Cancer Registration Service, 2016). The present paper describes the availability of data on important biological factors in the recently introduced English cancer registration data set. It provides an analysis of the survival of breast cancer patients from Black, Asian and White ethnic groups, with consideration of underlying differences in socio-economic factors, tumour stage and tumour biology.

## DATA AND METHODS

All breast cancer registrations for women in England diagnosed during 2012–2013 were extracted from the National Cancer Registration and Analysis Service. There were 87 538 records. We excluded 25 records with unknown vital status and 661 records that were based entirely on information from a death certificate, leaving 86 852 records for the analysis. Of these remaining cases, 106 had a record of diagnosis on the same date as the date of death. In order to retain these records in the survival analysis we added one day to their survival time.

We assigned women to broad ethnicity groupings: White, Black, Asian, and Other and unknown, the latter included mixed groups. The information on ethnicity is based on the self-reported ethnicity that persons provide when admitted to a hospital. The Black group was subsequently divided into subgroups Black Caribbean, Black African and Other Black.

ER was assessed according to national guidelines using immunohistochemical assessment with mandatory participation in a national quality assurance scheme. ER is reported semi-quantitatively with recording of both the proportion and intensity of nuclear cell reactivity; most histopathology laboratories categorise ER according to Allred score, with a tumour scoring 3 or more defined as ER positive. PR assessment is not mandatory but laboratories that do perform PR assays typically report according to the Allred scoring system with the same cut-off to define PR positivity. HER2 is assessed according to national guidelines and algorithm with immunohistochemistry as first line technique and *in situ* hybridisation as second line. Scores of 0 or 1+ are considered as HER2 negative and scores of 3+ as HER2 positive. Cases with borderline membrane reactivity on immunohistochemistry were categorised according to the ratio of number of copies of the Her2 gene to a chromosome 17 centromeric probe. Cases with a ratio of 2.00 or more were regarded as HER2 positive.

Covariates in the survival analysis were age (tabulated in 10-year groups; for age adjustment a second-order polynomial of age was used in the regression models in order to accommodate a non-linear association between age and mortality), socio-economic status (SES) (quintiles based on the income domain of IMD 2010 (Department for Communities and Local Government, 2011), geographical area of residence (nine Government Office Regions in England), Charlson comorbidity score (based on diagnoses in inpatient and day-case hospital discharge episodes in the 3-year period prior to breast cancer diagnosis, and excluding cancer itself as a comorbidity) (Charlson *et al*, 1987; Quan *et al*, 2005), breast cancer morphology (ductal, lobular, other and unspecified), tumour-nodes-metastasis (TNM) stage, grade, and ER, PR and HER2 receptor status of the breast cancer. The information on and ER, PR and HER2 receptor status was combined to define a group of triple-negative tumours where all three receptors were known and negative.

The 2-year cumulative risk of death from any cause was estimated with the Kaplan-Meier method, and univariate and multivariate Cox proportional hazards regressions were used to estimate hazard ratios (HR) and their 95% confidence intervals (95% CI). The follow-up ended on 31 December 2014. The total number of deaths was 8761 and breast cancer was recorded as the underlying cause of death in 70% of the cases with an available cause of death.

We initially analysed the entire breast cancer cohort, with sequential adjustment for age, socio-economic status, region, comorbidity, stage, grade, ER, PR and HER2 receptor status and morphology. We evaluated the assumption of proportional hazards using Schoenfeld residuals, and we explored the internal consistency and sensitivity using stratified analyses by period of follow-up, broad age group, subgroups of Black women, broad geography and stage. We finally did a focused analysis of the subpopulation where the mortality HR was particularly high for Black women compared with White women: premenopausal women in London who had stage IV breast cancer.

## RESULTS

Table 1 shows the distributions of age, socio-economic status, geography of residence, comorbidity score, morphology, stage, grade and receptor status.

Data on key biological prognostic factors were available in most cases, that is, stage (83%), grade (94%), ER (77%), HER2 (74%), but it was lower for PR status (41%). The missing information was not a random subset, and cases with missing data on stage, grade or HER2 had high HRs. Figure 1 demonstrates that cases with missing data on tumour grade had very high mortality in the short term after diagnosis.

**Table 1. Distribution of covariates in survival analysis of 86 852 breast cancer cases in England 2012–2013 with follow-up to 2014**

Characteristic	Persons	%	Deaths	2Y cum. risk (%)	Univariate		Mutually adjusted	
					HR	95% CI	HR	95% CI
<b>Age at diagnosis</b>								
0–39	3606	4	212	5.7	1.55	1.33–1.81	1.18	1.01–1.38
40–49	13 767	16	534	3.9	1.02	0.91–1.14	0.96	0.85–1.07
50–59	18 436	21	692	3.8	1.00		1.00	
60–69	22 639	26	1221	5.4	1.44	1.32–1.59	1.50	1.37–1.65
70–79	14 981	17	1906	12.9	3.59	3.29–3.92	2.70	2.47–2.95
80–89	10 799	12	2933	27.4	8.35	7.69–9.08	5.09	4.67–5.54
90+	2624	3	1263	49.5	18.00	16.40–19.75	10.12	9.19–11.15
Trend					$\chi^2(1) = 5401.1; P < 0.001$		$\chi^2(1) = 2702.7; P < 0.001$	
<b>Socio-economic deprivation</b>								
1 Least deprived	19 828	23	1521	7.8	1.00		1.00	
2	19 733	23	1911	9.6	1.27	1.19–1.36	1.18	1.10–1.26
3	18 409	21	1850	10.2	1.32	1.23–1.41	1.24	1.16–1.32
4	16 007	18	1871	11.8	1.56	1.45–1.67	1.36	1.27–1.46
5 Most deprived	12 875	15	1608	12.6	1.67	1.56–1.79	1.44	1.34–1.55
Trend					$\chi^2(1) = 247.1; P < 0.001$		$\chi^2(1) = 122.6; P < 0.001$	
<b>Region of residence</b>								
London	9925	11	970	9.9	1.00		1.00	
North East	4315	5	498	11.4	1.18	1.06–1.31	1.65	1.47–1.84
North West	11 698	13	1239	10.6	1.08	0.99–1.17	1.33	1.22–1.45
Yorkshire and the Humber	8334	10	859	10.8	1.05	0.95–1.15	1.33	1.21–1.46
East Midlands	7685	9	751	9.6	0.98	0.89–1.08	1.23	1.11–1.35
West Midlands	9171	11	937	10.4	1.04	0.95–1.13	1.37	1.25–1.50
East of England	10 317	12	1024	9.9	1.01	0.92–1.10	1.45	1.33–1.60
South East	15 199	17	1471	9.8	0.99	0.91–1.07	1.16	1.06–1.26
South West	10 208	12	1012	10.0	1.00	0.92–1.09	1.21	1.11–1.33
Heterogeneity					$\chi^2(8) = 17.9; P = 0.02$		$\chi^2(8) = 110.4; P < 0.001$	
<b>Charlson comorbidity</b>								
0	69 582	80	4942	7.1	1.00		1.00	
1	7552	9	1363	18.6	2.76	2.60–2.93	1.61	1.51–1.71
2+	9718	11	2456	25.4	4.06	3.87–4.26	2.22	2.11–2.34
Trend					$\chi^2(1) = 3593.6; P < 0.001$		$\chi^2(1) = 963.4; P < 0.001$	
<b>Morphology</b>								
Ductal	65 730	76	5821	8.9	1.00		1.00	
Lobular	10 653	12	1012	9.8	1.08	1.01–1.16	0.95	0.89–1.02
Other and unspecified	10 469	12	1928	18.4	2.23	2.12–2.35	1.31	1.23–1.39
Heterogeneity					$\chi^2(2) = 949.6; P < 0.001$		$\chi^2(2) = 89.0; P < 0.001$	
<b>TNM stage</b>								
I	32 218	37	961	3.0	1.00		1.00	
II	27 858	32	1807	6.5	2.23	2.06–2.41	1.68	1.55–1.82
III	7170	8	893	12.5	4.35	3.97–4.76	3.32	3.02–3.65
IV	4523	5	2238	50.6	23.67	21.95–25.53	13.91	12.86–15.05
NA	15 083	17	2862	18.7	6.89	6.40–7.41	3.64	3.37–3.93
Trend (excluding NA)					$\chi^2(1) = 7444.8; P < 0.001$		$\chi^2(1) = 5090.8; P < 0.001$	
<b>Grade</b>								
1	12 921	15	651	5.0	1.00		1.00	
2	42 090	48	3306	7.9	1.58	1.45–1.72	1.19	1.09–1.30
3	26 416	30	3166	12.0	2.43	2.23–2.65	1.83	1.67–2.00
NA	5425	6	1638	30.7	7.12	6.50–7.80	2.08	1.88–2.29
Trend (excluding NA)					$\chi^2(1) = 570.6; P < 0.001$		$\chi^2(1) = 275.5; P < 0.001$	
<b>Oestrogen receptor</b>								
Negative	9355	11	1511	16.8	1.00		1.00	
Positive	57 354	66	4303	7.5	0.44	0.42–0.47	0.60	0.56–0.65
NA	20 143	23	2947	14.6	0.90	0.84–0.96	0.81	0.74–0.89
Heterogeneity (excluding NA)					$\chi^2(1) = 755.2; P < 0.001$		$\chi^2(1) = 209.4; P < 0.001$	
<b>Progesterone receptor</b>								
Negative	11 464	13	1679	15.4	1.00		1.00	
Positive	24 422	28	1679	7.1	0.46	0.43–0.49	0.70	0.65–0.76
NA	50 966	59	5403	10.5	0.68	0.64–0.72	0.69	0.64–0.75
Heterogeneity (excluding NA)					$\chi^2(1) = 522.5; P < 0.001$		$\chi^2(1) = 111.9; P < 0.001$	
<b>HER2 receptor</b>								
Negative	54 874	63	4271	7.9	1.00		1.00	
Positive	9292	11	786	8.5	1.09	1.01–1.18	0.82	0.75–0.88
NA	22 686	26	3704	16.2	2.16	2.06–2.25	1.11	1.05–1.18
Heterogeneity (excluding NA)					$\chi^2(1) = 5.2; P = 0.02$		$\chi^2(1) = 30.4; P < 0.001$	
<b>Triple negative</b>								
No or NA	82 874	95	8087	9.8	1.00		1.00	
Yes	3978	5	674	18.2	1.86	1.72–2.01	2.00	1.84–2.17
Heterogeneity					$\chi^2(1) = 239.0; P < 0.001$		$\chi^2(1) = 271.5; P < 0.001$	

Abbreviations: CI = confidence interval, ER = oestrogen receptor, HER2 = human epidermal growth factor 2 receptor, HR = hazard ratio, NA = not available, PR = progesterone receptor. There were 8761 deaths from any cause. Two-year cumulative risks of death estimated with the Kaplan-Meier method, and hazard ratios from univariate and mutually adjusted Cox proportional hazards regression analyses. ER, PR and HER2 status were included in the same model, but triple-negative status was analysed in a separate model.

Each of the covariates in Table 1 was associated with mortality among breast cancer patients in univariate analysis. Mutual adjustment attenuated the estimated effects of age, socio-economic status, comorbidity, morphology, stage, grade, and oestrogen and progesterone receptor status. The mutual adjustment changed the association between region of residence and mortality. In the adjusted analysis, London residents had lower mortality than residents in other regions; this was mainly the result of adjustment for socio-economic status and stage. The effect of the derived triple-negative characteristic was robust to statistical adjustment and changed very little in the adjusted model.

The variables in Table 1 were subsequently used in sequential statistical adjustment in the analyses of ethnicity.

Table 2 shows the principal analysis of the broad ethnic groups. In the age-adjusted analysis, Black breast cancer patients had higher mortality than White patients (HR: 1.77; 95% CI: 1.48–2.13). Asian and Other and unknown groups of women had mortality rates very similar to White women (0.98 [0.81–1.18] and 0.99 [0.95–1.04], respectively).

The excess mortality of Black women was much reduced by adjustment for socio-economic status, geography and comorbidity (1.52 [1.26–1.84]). Adjustment for stage, grade, receptor status and morphology separately showed that stage provided the largest further attenuation of the effect (1.35 [1.12–1.63]), but each of the other tumour characteristics had its own independent effect on the Black vs White difference. In the fully adjusted model the excess mortality of Black women was reduced to 1.24 (1.03–1.50). From these estimates, the excess mortality in Black women was attributable in sequence to recorded social and person-level characteristics (socio-economic status and comorbidity) (32% [23–45%]), then to recorded tumour stage (22% [14–34%]) and then to recorded biological characteristics (grade, morphology, receptors) (14% [8–25%]), leaving 31% (22–43%) unexplained.

Table 3 shows the HRs of Black women vs White women in different stratified analyses. Stratification of the follow-up period into three 1-year intervals did not materially change the estimate. The excess mortality of Black women was much stronger in young women (0–49 years) than in middle-aged and older women (50+ years), and the effect in the younger group was less sensitive (more resilient) to statistical adjustment for person and tumour characteristics. The effect was different in subgroups of Black women with the highest excess consistently in the Other Black subgroup and the lowest excess in the Black Caribbean subgroup.

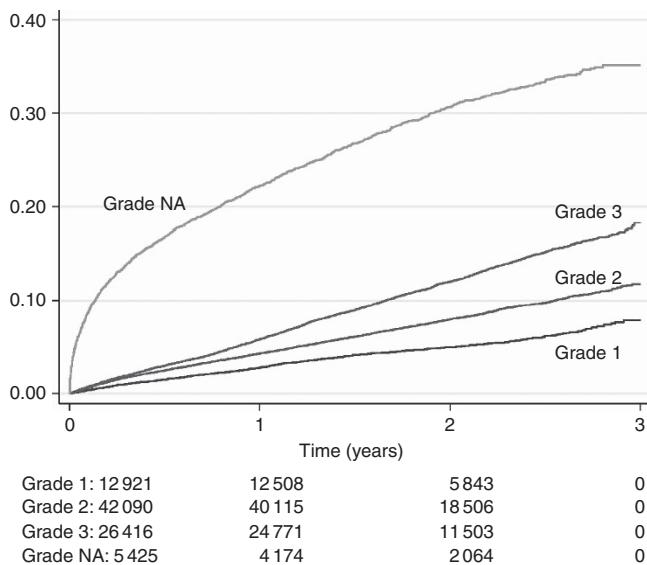


Figure 1. Kaplan-Meier failure estimates of death from any cause in breast cancer patients, in relation to tumour grade.

Table 2. Survival analysis of 86 852 breast cancer cases in England 2012–2013 with follow-up to 2014

Adjustments	Model 1		Model 2		Model 3a		Model 3b		Model 3c		Model 3d		Model 4	
	Age		Age SES Region Como		Age SES Region Como Stage		Age SES Region Como Grade		Age SES Region Como Receptors		Age SES Region Como Morphology		All covariates	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
White	1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Black	1.77	1.48–2.13	1.52	1.26–1.84	1.35	1.12–1.63	1.38	1.14–1.66	1.44	1.19–1.74	1.49	1.24–1.80	1.24	1.03–1.50
Asian	0.98	0.81–1.18	0.87	0.72–1.05	0.88	0.73–1.06	0.82	0.68–0.99	0.85	0.71–1.03	0.88	0.73–1.06	0.86	0.71–1.03
Other and unknown	0.99	0.95–1.04	1.03	0.98–1.08	1.03	0.98–1.08	1.05	1.00–1.10	1.01	0.96–1.06	1.03	0.98–1.08	1.05	1.00–1.11
Heterogeneity	$\chi^2(3) = 38.4; P < 0.001$		$\chi^2(3) = 23.2; P < 0.001$		$\chi^2(3) = 13.3; P = 0.004$		$\chi^2(3) = 20.0; P < 0.001$		$\chi^2(3) = 18.3; P < 0.001$		$\chi^2(3) = 20.8; P < 0.001$		$\chi^2(3) = 12.5; P = 0.006$	

Abbreviations: CI = confidence interval, ER = oestrogen receptor, HER2 = human epidermal growth factor 2 receptor, HR = hazard ratio, PR = progesterone receptor, SES = socio-economic status. Cox proportional regression analyses of ethnicity. Model 1 adjusted for age, socioeconomic status, region of residence and comorbidity. Model 2 adjusted for age, socioeconomic status, region of residence, comorbidity and stage (missing stage as a category). Model 3a adjusted for age, socioeconomic status, region of residence, comorbidity and grade. Model 3b adjusted for age, socioeconomic status, region of residence, comorbidity, and ER, PR and HER2 status. Model 3c adjusted for age, socioeconomic status, region of residence, comorbidity and morphology. Model 4 adjusted for age, socioeconomic status, region of residence, comorbidity, stage, grade, ER, PR and HER2 status, and morphology.

**Table 3. Survival analysis of Black and White breast cancer women cases in England 2012–2013 with follow-up to 2014**

Adjustments	Model 1		Model 2		Model 3		Model 4	
	Age		Age SES Region Como		Age SES Region Como Stage		All covariates	
Stratification variable	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
All (from Table 2)	1.77	1.48–2.13	1.52	1.26–1.84	1.35	1.12–1.63	1.24	1.03–1.50
Period of follow-up								
0–1	1.71	1.32–2.22	1.46	1.12–1.91	1.28	0.98–1.67	1.15	0.88–1.50
1–2	1.81	1.36–2.43	1.59	1.18–2.14	1.43	1.06–1.93	1.34	0.99–1.81
2–3	1.89	1.11–3.22	1.54	0.89–2.66	1.37	0.79–2.37	1.33	0.77–2.31
Age								
0–49	2.43	1.83–3.23	2.15	1.58–2.93	1.84	1.35–2.51	1.71	1.25–2.33
50+	1.40	1.10–1.79	1.19	0.93–1.53	1.07	0.84–1.38	0.98	0.76–1.25
Ethnic subgroup								
Black Caribbean	1.54	1.19–2.01	1.31	1.00–1.70	1.20	0.92–1.57	1.10	0.84–1.43
Black African	1.89	1.37–2.61	1.70	1.22–2.35	1.43	1.03–1.98	1.25	0.90–1.74
Other Black	2.36	1.60–3.46	2.05	1.39–3.02	1.76	1.19–2.59	1.81	1.23–2.67
Geography								
London	1.74	1.37–2.21	1.59	1.25–2.02	1.51	1.19–1.93	1.43	1.12–1.83
Other	1.55	1.12–2.16	1.28	0.92–1.78	1.07	0.77–1.49	0.92	0.66–1.27
Stage								
I	1.45	0.69–3.06	1.28	0.60–2.73			1.07	0.50–2.29
II	1.14	0.71–1.81	1.05	0.65–1.69			0.98	0.61–1.58
III	1.26	0.81–1.95	1.22	0.77–1.93			1.09	0.69–1.74
IV	1.51	1.15–1.97	1.63	1.24–2.15			1.47	1.11–1.93
NA	1.16	0.71–1.90	1.01	0.61–1.66			0.93	0.57–1.53

Abbreviations: CI = confidence interval, ER = oestrogen receptor, HER2 = human epidermal growth factor 2 receptor, HR = hazard ratio, NA = not available, PR = progesterone receptor; SES = socio-economic status. Cox proportional regression analyses comparing Black women with White women. Model 1 adjusted for age. Model 2 adjusted for age, socioeconomic status, region of residence and comorbidity. Model 3 adjusted for age, socioeconomic status, region of residence, comorbidity and stage (missing stage as a category). Model 4 adjusted for age, socioeconomic status, region of residence, comorbidity, stage, grade, ER, PR and HER2 status, and morphology.

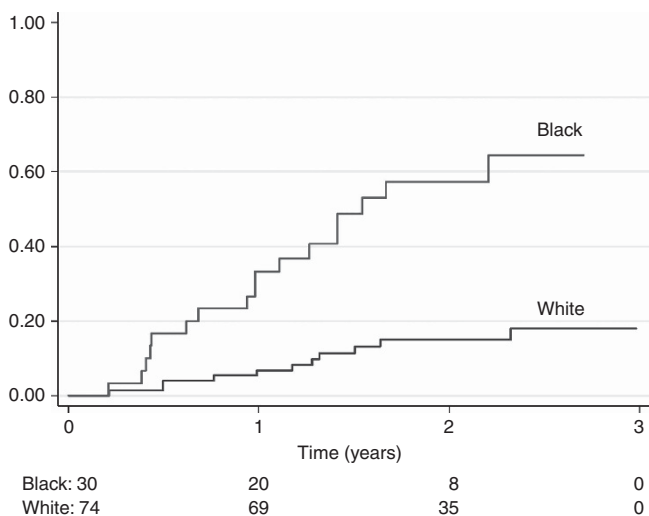


Figure 2. Kaplan-Meier failure estimates of death from any cause in Black and White breast cancer patients, 0–49 years of age, resident in London, with stage IV cancer.

The excess mortality of Black women was highest and least sensitive to adjustments in the London population. Finally, analysis stratified by stage showed that the excess mortality of Black breast cancer patients was strongest and least sensitive to adjustments among women with stage IV cancer. The age-adjusted HR within stage IV was 1.51 (1.15–1.97) and the fully adjusted HR was 1.47 (1.11–1.93).

The subsequent analysis focused on the population of young breast cancer patients in London who were diagnosed with

stage IV breast cancer. This is the subgroup where the difference between Black and White women is largest. Figure 2 shows the Kaplan-Meier analysis of the young, stage IV breast cancer patients in London, comparing White patients (74 patients; 11 deaths) and Black patients (30 patients; 17 deaths). The difference was highly statistically significant (log-rank test,  $P < 0.001$ ). There was no indication of a difference between Black Caribbean, Black African and Other Black groups ( $P = 0.89$ ). Comparing Black and White women, the age-adjusted HR was 5.21 (2.42–11.22) (data not shown). The estimate was not sensitive to further statistical adjustment for socio-economic status, grade, receptor status or morphology, but adjustment for co-morbidity reduced the estimate to 3.70 (1.62–8.44). A higher proportion of the Black patients had missing data for size of the primary tumour (73 vs 62%) and for nodal status (80 vs 58%). Among women with non-missing values, Black women had larger median tumour size (37 mm vs 28 mm) and a larger number of positive nodes (3 vs 2).

## DISCUSSION

This is a very large study of current mortality outcomes in unselected breast cancer patients in a large, national setting. The modernisation of cancer registration in England has led to a much improved data set with high completeness of collection of several important prognostic factors. The present analysis confirms known or expected associations of mortality with age at diagnosis, socio-economic status, co-morbidity, morphology, stage, grade and receptor status. The study shows higher mortality in Black breast cancer patients, compared with White patients, and demonstrates independent contributions to this excess from personal and social factors, from tumour stage, and from biological characteristics of the cancer.

The analyses and findings presented here have been possible because data on prognostic factors (especially stage, grade and tumour receptor status) are now available for breast cancer cases in the English national cancer registration data set at a level that permits detailed, adjusted analyses. The known excess mortality of Black women with breast cancer has contributions from an adverse case-mix distribution, including socio-economic factors, stage distribution and tumour biology.

Population-based cancer registration aims to register all cases of cancer in the geographically defined population, regardless of route to diagnosis, basis of diagnosis, stage of disease or receipt of specialised oncology care. Compared with groups of cancer patients accrued from a histopathology case series or oncology case series, the population-based register will inevitably include a sub-set of cancer patients that were only diagnosed shortly before death and patients that for other reasons received no active cancer care. It is, therefore, to be expected that the population-based data set will be partially completed with data items that require diagnostic procedures and pathology. Registered cases with no information on, for example, stage, grade or hormone receptor status indicate situations where clinical or pathological investigations were not possible or were considered to be of little relevance to the care of the patient. The missing data on these items are, therefore, likely to be selective. This is evident from the Kaplan-Meier analysis for tumour grade (Figure 1) and was also observed for patients with missing data on stage and hormone receptor status. When the missing data are selective we consider it misleading to impute the missing values, and we decided to represent the missing data as a separate category of each variable (Galati and Seaton, 2016).

The mutual statistical adjustment in the Cox regression model greatly attenuated the high HRs for missing stage (the HR of 6.89 was reduced to 3.64), missing grade (7.12 to 2.08) and missing HER2 receptor (2.16 to 1.11). This indicates a degree of correlation between these characteristics in a sub-set of advanced and rapidly fatal cases where staging and histopathology analysis were perhaps not relevant to the care of the patient.

Breast cancer cases with missing data on ER and PR status had HRs that were intermediate between the receptor-negative and the receptor-positive subgroups, but we note that the proportion of cases with no information on PR status was high (59%). The importance of ER status as a prognostic factor has been established firmly since the 1970s (Fisher *et al*, 1988), but the role of PR status has been less certain. The prevailing attitude in the UK has been that PR status was less important than ER and HER2 status, and the 2009 NICE guidelines advised against routine assessment of PR status (National Institute for Health and Care Excellence, 2009). This may be the reason for the current poor recording of PR status. The present results should re-enforce the view that PR status may be a relevant indicator of mortality outcomes, independently of other prognostic factors, including ER and HER2 status.

We were particularly interested in the category of 'triple-negative' cancers (Dent *et al*, 2007; Foulkes *et al*, 2010) and we attempted to derive this entity from the recorded ER, PR and HER2 data. We identified a sub-set of 5% of the total cohort of breast cancer cases who were known to be negative for each of the three hormone receptors, and found that these had an HR of 2.00 compared with the remaining 95% of patients. However, due to the missing information on some receptor data, the 5% is less than the expected value of 11–17% (Dent *et al*, 2007; Foulkes *et al*, 2010), and the estimated HR probably lower than the true value due to the misclassification.

The prognosis of Black women with breast cancer is known to be worse than in White women, most likely for a variety of reasons, such as socio-economic deprivation, lower uptake of mammographic screening, more advanced stage and a more adverse

biological case-mix (Jack *et al*, 2009, 2014). By means of sequential adjusted regression models, it was attempted to establish the origins of the difference in prognosis of Black and White patients, and to attribute the excess mortality in Black cancer patients to factors at different levels, ranging from the social and personal, through stage of disease, and to the biological characteristics of the cancer. This sequence was chosen because the social and personal characteristics are at least in principle modifiable, and could be subject to intervention such as facilitating the awareness of symptoms of cancer or the uptake of mammographic screening. Stage of the cancer is intermediate because this is a mixture of effects from the social and personal level (e.g., through early diagnosis) and the biological level (more aggressive biology giving rise to a more advanced stage distribution). The factors at the biological level are most likely not amenable to intervention. We found that the excess mortality of Black breast cancer patients had contributions from all three levels (social/personal; stage; biology).

Our final analysis focused on the sub-set of the data where the Black/White difference was strongest (i.e., age group 0–49 years; London residents; stage IV cancer). This reduced the analysis population from 86 852 to 104. We had hoped that analysis of this niche group would reveal something about the underlying cause of the ethnic difference, but we were not able to attribute any of this marked difference to the available characteristics. Most plausibly, a larger sub-set of the Black patients is diagnosed very late with the disease in a potentially untreatable, incurable state.

The factors investigated in this study partially explain the substantial variation in breast cancer survival by ethnicity; however, 31% could not be explained by the available variables. This may in part be due to imperfect classification of the available covariates, or other factors (including treatment) that may influence the variation in breast cancer survival, and which were not investigated. Potentially important is the use of treatment, as it has been shown that ethnicity can be associated with delivery of treatment (Fedewa *et al*, 2010; Sail *et al*, 2012; Reeder-Hayes *et al*, 2013; Silber *et al*, 2013). One American study found that African American breast cancer patients had a higher risk of both a 60-day and 90-day delay of chemotherapy following surgery (Fedewa *et al*, 2010), which may influence variation in short-term survival. In addition, a number of other factors, for example genetics (Pal *et al*, 2015), family history, comorbidities, alcohol, smoking and education (Tannenbaum *et al*, 2013; Wu *et al*, 2013; Shariff-Marco *et al*, 2014) have been found to be associated with breast cancer survival, and it is not yet understood how the factors interact. Previous work in South East England showed that Black African women were less likely to receive surgery, chemotherapy and hormone therapy than White women (Jack *et al*, 2009). While this may contribute to the differences in survival, it would also point to possible worrying inequity in treatment uptake between ethnic groups. It remains an important limitation of this analysis that detailed treatment information is not yet available in the analysis data set. Work is in progress to include radiotherapy, systemic therapy and surgery information in the data set.

The principal strength and relevance of this analysis is the use of national, population-based data on breast cancer in England. The recency of the present data are both a strength and a limitation. Data and outcomes of patients diagnosed during 2012–2013 are likely to be relevant to the clinical management of breast cancer today, but the duration of follow-up for mortality is short and this analysis addresses short-term survival only. The ultimate outcome in breast cancer care is long-term cure and good quality of life. In defence of our short-term outcomes analysis, we consider, firstly, that short-term survival is a necessary condition for long-term survival and quality of life, and, secondly, international comparisons have made it clear that

short-term survival is a particular concern in breast cancer care in England (Møller *et al*, 2010).

## ACKNOWLEDGEMENTS

The study was supported by a Cancer Research UK Early Diagnosis Policy Research Grant and by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. This work was funded by Cancer Research UK and the National Institute for Health Research.

## CONFLICT OF INTEREST

Katherine Henson has in the past been an employee of BUPA. Ruth Jack and Elizabeth Davies have received research funding for an unrelated project from Hospital Corporation of America International. All other authors declare no conflict of interest. The views expressed are those of the authors and not necessarily those of CRUK, the NHS, the NIHR or the Department of Health. The researchers were independent of the funders, and the researchers alone interpreted the data and decided to publish.

## CONTRIBUTORS

HM, ML, and AP conceived and designed the study. JB and JC assembled the data set. H.M. analysed the data. All authors contributed to the interpretation of the results. All authors contributed to the manuscript and approved the final version.

## REFERENCES

- Boyer-Chammond A, Taylor TH, Anton-Culver H (1999) Survival differences in breast cancer among racial/ethnic groups: a population-based study. *Cancer Detect Prev* **23**(6): 463–473.
- Cancer Research UK (2016) Breast cancer incidence (invasive). Available at <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero>.
- Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* **40**(5): 373–383.
- Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan CE, Turner D, Marrett L, Gjerstorff ML, Johannesen TB, Adolffson J, Lambe M, Lawrence G, Meechan D, Morris EJ, Middleton R, Steward J, Richards MA. ICBP Module 1 Working Group (2011) Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* **377**(9760): 127–138.
- 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 (2014) Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study. *Br J Cancer* **110**(1): 230–241.
- Dawson SJ, Rueda OM, Aparicio S, Caldas C (2013) A new genome-driven integrated classification of breast cancer and its implications. *Embo J* **32**(5): 617–628.
- DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A (2016) Breast cancer statistics, 2015: convergence of incidence rates between black and white women. *CA Cancer J Clin* **66**(1): 31–42.
- Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clin Cancer Res* **13**(15): 4429–4434.
- Department for Communities and Local Government. *English Indices of Deprivation 2010* (2011) Available at <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>.
- Fedewa SA, Ward EM, Stewart AK, Edge SB (2010) Delays in adjuvant chemotherapy treatment among patients with breast cancer are more likely in African American and Hispanic populations: a national cohort study 2004–2006. *J Clin Oncol* **28**(27): 4135–4141.
- Fisher B, Redmond C, Fisher ER, Caplan R (1988) Relative worth of oestrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncol* **6**(7): 1076–1087.
- Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-negative breast cancer. *N Engl J Med* **363**(20): 1938–1948.
- Galati JC, Seaton KA (2016) MCAR is not necessary for the complete cases to constitute a simple random subsample of the target sample. *Stat Methods Med Res* **25**: 1527–1534.
- Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA (2015) Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *JAMA* **313**(2): 165–173.
- Jack RH, Davies EA, Møller H (2009) Breast cancer incidence, stage, treatment and survival in ethnic groups in South East England. *Br J Cancer* **100**(3): 545–550.
- Jack RH, Davies EA, Renshaw C, Tutt A, Grocock MJ, Coupland VH, Møller H (2013) Differences in breast cancer hormone receptor status in ethnic groups: a London population. *Eur J Cancer* **49**(3): 696–702.
- Jack RH, Møller H, Robson T, Davies EA (2014) Breast cancer screening uptake among women from different ethnic groups in London: a population-based cohort study. *BMJ Open* **4**(10): e005586.
- Lakhani SR (2012) *WHO classification of tumours of the breast; International Agency for Research on Cancer; World Health Organization*. International Agency for Research on Cancer: Lyon.
- Møller H, Sandin F, Bray F, Klint A, Linklater KM, Purushotham A, Robinson D, Holmberg L (2010) Breast cancer survival in England, Norway and Sweden: a population-based comparison. *Int J Cancer* **127**(11): 2630–2638.
- National Institute for Health and Care Excellence (2009) Early and locally advanced breast cancer: diagnosis and treatment. NICE. Available at <https://www.nice.org.uk/guidance/cg80/chapter/1-Guidance#endocrine-therapy>.
- Onitilo AA, Engel JM, Greenlee RT, Mukesh BN (2009) Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res* **7**(1–2): 4–13.
- Pal T, Bonner D, Cragun D, Monteiro AN, Phelan C, Servais L, Kim J, Narod SA, Akbari MR, Vadaparqampil ST (2015) A high frequency of BRCA mutations in young black women with breast cancer residing in Florida. *Cancer* **121**(23): 4173–4180.
- Parise CA, Caggiano V (2014) Breast cancer survival defined by the ER/PR/HER2 subtypes and a surrogate classification according to tumor grade and immunohistochemical biomarkers. *J Cancer Epidemiol* **2014**: 469251.
- Piccari M, Lohrisch C, Di Leo A, Larsimont D (2001) The predictive value of HER2 in breast cancer. *Oncology* **61**(Suppl 2): 73–82.
- Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA (2005) Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* **43**(11): 1130–1139.
- Reeder-Hayes KE, Meyer AM, Dusetzina S, Liu H, Wheeler SB (2013) Racial disparities in initiation of endocrine therapy for early-stage breast cancer. *J Clin Oncol* **31**(15 suppl): 6572.
- Sail K, Franzini L, Lairson D, Du X (2012) Differences in treatment and survival among African-American and Caucasian women with early stage operable breast cancer. *Ethn Health* **17**(3): 309–323.
- Shariff-Marco S, Yang J, John EM, Sangaramoorthy M, Hertz A, Koo J, Nelson DO, Schupp CW, Shema SJ, Cockburn M, Satariano WA, Yen IH, Ponce NA, Winkleby M, Keegan TH, Gomez SL (2014) Impact of neighbourhood and individual socioeconomic status on survival after breast cancer varies by race/ethnicity: the Neighbourhood and Breast Cancer Study. *Cancer Epidemiol Biomarkers Prev* **23**(5): 793–811.
- Silber JH, Rosenbaum PR, Clark AS, Giantonio BJ, Ross RN, Teng Y, Wang M, Niknam BA, Ludwig JM, Wang W, Even-Shoshan O, Fox KR

- (2013) Characteristics associated with differences in survival among black and white women with breast cancer. *JAMA* **310**(4): 389–397.
- Tannenbaum SL, Koru-Sengul T, Miao F, Byrne MM (2013) Disparities in survival after female breast cancer diagnosis: a population-based study. *Cancer Causes Control* **24**(9): 1705–1715.
- The National Cancer Registration Service. Available at <http://www.ncr.nhs.uk/> (2016) .
- Warner ET, Tamimi RM, Hughes ME, Ottesen RA, Wong YN, Edge SB, Theriault RL, Blayney DW, Niland JC, Winer EP, Weeks JC, Partridge AH (2015) Racial and ethnic differences in breast cancer survival: mediating effect of tumor characteristics and sociodemographic and treatment factors. *J Clin Oncol* **33**(20): 2254–2261.
- Wu AH, Gomez SL, Vigen C, Kwan ML, Keegan TH, Lu Y, Shariff-Marco S, Monroe KR, Kurian AW, Cheng I, Caan BJ, Lee VS, Roh JM, Sullivan-Halley J, Henderson BE, Bernstein L, John EM, Spoto R (2013) The California Breast Cancer Survivorship Consortium (CBCSC): prognostic factors associated with racial/ethnic differences in breast cancer survival. *Cancer Causes Control* **24**(10): 1821–1836.

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 4.0 Unported License.