British Journal of Cancer (2012) 106, 1068 – 1075

www.bjcancer.com

Variation in advanced stage at diagnosis of lung and female breast cancer in an English region 2006–2009

G Lyratzopoulos*[,1], GA Abel[1], JM Barbiere[1], CH Brown[2], BA Rous[2] and DC Greenberg[2]

[1]Department of Public Health and Primary Care, Cambridge Centre for Health Services Research, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK; [2]Eastern Cancer Registration and Information Centre, Unit C - Magog Court, Shelford Bottom, Hinton Way, Cambridge CB22 3AD, UK

BACKGROUND: Understanding variation in stage at diagnosis can inform interventions to improve the timeliness of diagnosis for patients with different cancers and characteristics.

METHODS: We analysed population-based data on 17 836 and 13 286 East of England residents diagnosed with (female) breast and lung cancer during 2006–2009, with stage information on 16 460 (92%) and 10 435 (79%) patients, respectively. Odds ratios (ORs) of advanced stage at diagnosis adjusted for patient and tumour characteristics were derived using logistic regression.

RESULTS: We present adjusted ORs of diagnosis in stages III/IV compared with diagnosis in stages I/II. For breast cancer, the frequency of advanced stage at diagnosis increased stepwise among old women (ORs: 1.21, 1.46, 1.68 and 1.78 for women aged 70–74, 75–79, 80–84 and ⩾85, respectively, compared with those aged 65–69 , $P<0.001$). In contrast, for lung cancer advanced stage at diagnosis was less frequent in old patients (ORs: 0.82, 0.74, 0.73 and 0.66, $P<0.001$). Advanced stage at diagnosis was more frequent in more deprived women with breast cancer (OR: 1.23 for most compared with least deprived, $P=0.002$), and in men with lung cancer (OR: 1.14, $P=0.011$). The observed patterns were robust to sensitivity analyses approaches for handling missing stage data under different assumptions.

CONCLUSION: Interventions to help improve the timeliness of diagnosis of different cancers should be targeted at specific age groups.
British Journal of Cancer (2012) **106,** 1068–1075. doi:10.1038/bjc.2012.30 www.bjcancer.com
Published online 1 March 2012

Keywords: stage; diagnosis; advanced; age; multiple; imputation

Increasing the proportion of cancer patients who are diagnosed in early stage could help decrease the number of cancer-related deaths (Abdel-Rahman *et al*, 2009). Therefore, national cancer control policies in several countries currently encompass initiatives supporting early detection and diagnosis (Olesen *et al*, 2009; Richards, 2009; Coleman *et al*, 2011).

The evidence base supporting these initiatives, however, is complex and heterogeneous (Richards, 2009). Markers and measures of the timeliness of diagnosis currently in use include short-term survival (NCIN (National Cancer Intelligence Network), 2008a; Møller *et al*, 2009; Rachet *et al*, 2009), diagnosis after an emergency hospital admission (NCIN (National Cancer Intelligence Network), 2010), and length of time intervals between symptom onset and diagnosis (Neal and Allgar, 2005; Macleod *et al*, 2009; Olesen *et al*, 2009). Stage at diagnosis is an excellent measure of early detection, but UK population-based data regarding this measure are limited. A recent National Audit Office report indicated that the completeness of stage information across English cancer registries is <40% (NAO (National Audit Office), 2010).

A better understanding of socio-demographic variation in stage at diagnosis could help stratify and tailor symptom awareness and early diagnosis interventions aimed at specific patient groups. We distinguish between 'stratification' that is, the targeting of an intervention to patient populations at a higher risk and 'tailoring', that is, the adaptation (or customising), a generic intervention to make its application more suitable for specific patient groups. An example of this concept relates to targeted interventions to increase breast cancer symptom awareness amongst older women (Forbes *et al*, 2011). It can also help focus early diagnosis audit efforts (RCGP (Royal College of General Practitioners), 2011) towards the cancers and patient groups with greatest potential for improvement.

Against this background, we have set out to examine socio-demographic variation in stage at diagnosis for female breast and lung cancers (two common cancers responsible for about 30% of all cancer diagnoses and cancer deaths in England (NCIN (National Cancer Intelligence Network), 2008b) during a recent period.

MATERIALS AND METHODS

Data

We analysed information on the stage at diagnosis of East of England patients diagnosed with female breast ('breast' hereafter) and lung cancer during the 4-year period 2006–2009 (International Classification of Diseases (ICD)-10 codes C50

*Correspondence: Dr G Lyratzopoulos; E-mail: gl290@medschl.cam.ac.uk
Received 1 November 2011; revised 23 January 2012; accepted 24 January 2012; published online 1 March 2012

and C34, respectively). The study period was chosen as the most recent for which data were available at the time of analysis. Anonymous data were extracted from the Eastern Cancer Registration and Information Centre (ECRIC), a population-based cancer registry covering a general population of ~5.7 million. The Registry has excellent performance as indicated by conventional measures of cancer registration quality such as death-certificate only registrations (~0%) and, uniquely at present among other English cancer registries, it holds information on stage at diagnosis for a particularly high proportion of patients (NAO (National Audit Office), 2010). Stage at diagnosis was classified using the 5th edition of the TNM classification, comprising stages I–IV (Sobin and Wittekind, 1997). Stage at diagnosis was assigned by CHB and BR, integrating clinical, imaging and pathological information. Patient socioeconomic status was ascribed using the income domain of the Index of Multiple Deprivation (IMD) 2004 deprivation score of the Lower Super Output Area (LSOA) of patients' residence in order to define quintile groups (1 = least deprived, or 'most affluent'; 5 = most deprived) (Office of the Deputy Prime Minister, 2004). The income domain of IMD 2004 incorporates information on the proportion of residents of a small area who live in households receiving state-funded support (for example, in the form of income support, unemployment benefit and tax credits). Tumour histological type was categorised into seven groups for breast (infiltrating ductal carcinoma, lobular carcinoma, mixed ductal lobular, other adenocarcinoma, other specified carcinoma, specified not carcinoma tumours and other unspecified) and eight for lung cancer (adenocarcinoma, squamous cell carcinoma, other non-small cell, small cell carcinoma, large cell carcinoma, carcinoid, other specified and other unspecified), using appropriate ICD-Oncology morphology codes (WHO (Word Health Organisation), 2000).

Analysis

We aimed to examine socio-demographic variation in advanced stage at diagnosis.

Initial analysis was confined to patients with known stage (complete case analysis). Binary logistic regression was used, defining advanced stage at diagnosis both as diagnosis in stages III/IV, or alternatively as diagnosis in stages II–IV (that is, diagnosis other than in stage I). For brevity, we present findings regarding variation in diagnosis in stages III/IV (*vs* I–II) in the main paper and append analysis relating to diagnosis at stage I (*vs* II–IV). We considered, but did not use, ordinal logistic regression because initial analysis provided evidence of violation of the proportional odds assumption.

Mixed-effects logistic regression models were used to predict advanced stage at diagnosis, adjusting for age group, deprivation quintile and tumour type (both cancers), sex (lung cancer) and screening detection status (breast cancer) as fixed effect categorical variables and including a random effect for Primary Care Trust. Although the UK government plans to abolish Primary Care Trusts in the future, they were responsible for planning, purchasing and quality assuring preventive services and primary or specialist health care for their residents during the study period (2006–2009). A model using only fixed effect variables for patient characteristics would assume that all observations are independent. In reality, patients within the same organisation may be more similar. Therefore, the models used recognise the hierarchical nature of the data, with patient-level observations being nested within Primary Care Trusts. Therefore, they provided information about patient-level variation (for example, between patients of different age, sex or deprivation status) without the risk of identifying spurious associations arising from potential clustering of different patient subgroups in Primary Care Trusts with higher or lower rates of advanced stage at diagnosis. To explore a potential interaction between age and sex for lung cancer, we have included in a subsequent model an interaction variable for age category (continuous) by sex.

Significance testing was principally based on joint log likelihood ratio tests. We specifically focused aspects of the analysis on patients aged >70 years of age because in recent decades improvements in cancer survival in this age group were smaller compared with those observed in younger patients, a finding thought to partially reflect relatively more advanced stage at diagnosis amongst older patients (Quaglia *et al*, 2009). Therefore, in addition to testing the overall effect of age, we also examined the significance of differences between patients ⩾70 years compared with patients in all other age groups. Further, tests for linear trend were used to examine the significance of deprivation group gradients by treating deprivation quintile as continuous rather than a categorical variable.

Sensitivity analysis Complete case analysis may be biased, depending on the mechanism responsible for missing data, that is, if data are not 'missing completely at random' (MCAR) (Appendix Table A1). (Sterne *et al*, 2009). Therefore, in addition, we have used two different sensitivity analysis approaches for handling potential bias arising from missing stage information, bearing in mind different assumptions about the potential mechanisms generating missing data.

First, we used multiple imputation to impute stage. Multiple imputation is a method increasingly used in the context cancer epidemiological studies (He *et al*, 2008; Nur *et al*, 2010; Ali *et al*, 2011). It assumes that data are 'missing at random' (MAR), that is, that any systematic differences between the missing and observed values can be estimated using information from the observed data (note: the MAR assumption does not mean that there are no systematic associations between missing data and specific variables) (Appendix Table A1). We included in imputation models survival, tumour histological grade, basis of diagnosis (that is, whether the diagnosis was verified with histology or not), Primary Care Trust and oestrogen receptor status (breast cancer imputation models only) in addition to all the variables used in the analysis models. All exposure variables used in either the analysis or imputation models were complete, except for grade and oestrogen receptor status (used in imputation models).

Second, as it is not possible to verify the MAR assumption empirically, we conducted sensitivity analysis with a more extreme imputation of missing stage that falls under the assumption of data 'missing not at random' (MNAR) (Appendix Table A1). To do this, we assigned all patients with unknown stage to the advanced stage category (III/IV), and repeated the analysis. This extreme case scenario approach is based on observations that the survival of patients with missing stage information is typically similar to that of patients diagnosed in advanced stage (ECRIC (Eastern Cancer Registration and Information Centre), 2011). We do not expect this extreme case scenario to represent a true situation, but we use it to illustrate how sensitive the complete case and multiple imputation analyses may be to the MCAR or MAR assumptions, respectively. All analysis was conducted in STATA 11 (StataCorp. 2009, College Station, TX, USA), including using the *ice* and *mim* commands used for multiple imputation (Royston, 2007). Further details are provided in Appendix Table A1.

RESULTS

Data relate to 17 836 and 13 286 patients with incident diagnosis of breast and lung cancer. Information on stage at diagnosis was complete for 16 460 (92%) and 10 435 (79%) patients. The completeness of stage information varied substantially between patients with different socio-demographic characteristics and tumour types – missing stage was more frequent in older patients in particular ($P<0.001$ for both cancers, Appendix Table A2). Among staged patients with breast and lung cancer, 41% and 15%

were diagnosed in stage I, and 86% and 21% in stages I/II, respectively (Table 1).

Multivariate complete case analysis

Breast cancer There was very strong evidence of an association between age and diagnosis in stages III/IV, (Table 2). Specifically for women aged ⩾70 years, the frequency of diagnosis in stages III/IV increased progressively with older age (odds ratios (ORs): 1.21, 1.46, 1.68 and 1.78 for women aged 70–74, 75–79, 80–84 and ⩾85 years, respectively, $P<0.001$). Increasing deprivation was associated with a greater frequency of stage III/IV diagnosis (joint log likelihood ratio $P=0.010$, p for trend = 0.002; Table 2).

Lung cancer There was very strong evidence of an association between age and advanced stage at diagnosis (Table 3). The frequency of stage III/IV diagnosis decreased progressively among patients aged ⩾70 years (ORs: of 0.82, 0.74, 0.73 and 0.66 for patients aged 70–74, 75–79, 80–84 and ⩾85 years, respectively, $P<0.001$). There was no evidence for deprivation group differences in lung cancer diagnosis at stages III/IV, in spite of an apparent trend towards lower frequency with increasing deprivation (p for trend = 0.236) (Table 3). There was strong evidence of a higher frequency of advanced stage at diagnosis in men (odds ratio of 1.14 for diagnosis in stages III/IV, $P=0.011$). There was no evidence for a differential effect of age in men and women (OR for men *vs* women per increase in age group category = 0.96, 95% CI 0.92–1.01, $P=0.100$). Although this may reflect lack of power, the size of the interaction indicates that a large synergistic effect is unlikely.

Examining variation in diagnosis in stage I *vs* II–IV produced overall similar findings for lung cancer. For breast cancer, the findings were similar in respect of variation in older age, but there was no evidence of deprivation differences (Appendix Tables A3 and A4).

Sensitivity analysis Repeating the analysis using multiple imputation of missing stage information produced highly similar values and patterns to those derived by the complete case analysis (Tables 4 and 5). Specifically, for both breast and lung cancer the same patterns of variation by age, deprivation and sex (for lung cancer only) were apparent. Repeating the analysis using the extreme case scenario approach (missing stage = advanced stage) produced similar patterns of variation for lung cancer. For breast cancer, in the extreme case scenario that the true stage at diagnosis of all women with missing information was either stage III or IV, deprivation differences in advanced stage at diagnosis would be smaller. The full output from all analysis models is provided in Appendix Table A5.

Table 1 Proportion of patients by stage, gender, age and deprivation group categories for breast and lung cancer (2006–2009)

	Breast			Lung		
	N	% among all patients	% among patients with known stage	*N*	% among all patients	% among patients with known stage
Stage						
Stage I	6788	38%	41%	1534	12%	15%
Stage II	7361	41%	45%	670	5%	6%
Stage III	1490	8%	9%	3483	26%	33%
Stage IV	821	5%	5%	4748	36%	46%
Unknown	1376	8%	n/a	2851	21%	n/a
Sex						
Men	n/a			7684	58%	
Women	17 836	100%		5602	42%	
Age group[a]						
15–39	770	4%				
40–44	1091	6%		n/a		
45–49	1539	9%				
15–49	n/a			380	3%	
50–54	2048	11%		443	3%	
55–59	1911	11%		903	7%	
60–64	2461	14%		1525	11%	
65–69	2152	12%		1762	13%	
70–74	1491	8%		2166	16%	
75–79	1590	9%		2384	18%	
80–84	1321	7%		2099	16%	
⩾85	1462	8%		1624	12%	
Deprivation group						
Affluent	4778	27%		2471	19%	
2	4658	26%		3072	23%	
3	4323	24%		3444	26%	
4	3081	17%		3072	23%	
Deprived	996	6%		1227	9%	

[a]Younger age groups were categorised differently for the two examined cancers because compared with breast cancer there were fewer patients with lung cancer in the younger age groups.

DISCUSSION

Summary of findings and comparisons with other literature

Using population-based data, we identified substantial socio-demographic variation in the stage at diagnosis of breast and lung cancer. Breast cancer patients who were ⩾70 years of age had a higher frequency of advanced stage at diagnosis. Conversely, age ⩾70 was associated with a lower frequency of advanced stage at diagnosis for lung cancer. Advanced stage at diagnosis was more frequent in more deprived patients with breast cancer. Men with lung cancer had a higher frequency of advanced stage at diagnosis.

Table 2 Breast cancer. Independent associations of age and deprivation with advanced stage at diagnosis (i.e., stage III/IV *vs* stage I/II)[a] ($n=16\,460$)

	Odds ratio	Lower 95% confidence interval	Higher 95% confidence interval	*P*
15–39	1.15	0.89	1.48	
40–44	1.02	0.81	1.28	
45–49	0.91	0.74	1.14	
50–54	0.92	0.74	1.14	
55–59	0.90	0.72	1.12	
60–64	0.91	0.74	1.12	
65–69	Reference			<0.001[b] (<0.001)[c]
70–74	1.21	0.98	1.49	
75–79	1.46	1.20	1.78	
80–84	1.68	1.37	2.07	
⩾85	1.78	1.45	2.18	
Most affluent	Reference			0.010[b] (0.002)[d]
2	1.16	1.02	1.32	
3	1.12	0.98	1.28	
4	1.29	1.12	1.49	
Deprived	1.23	1.00	1.52	

[a]From logistic regression models, with stage III/IV *vs* stage I/II diagnosis as the binary outcome variable. Models were adjusted for age, deprivation, tumour type and diagnosis through screening or symptomatically, and included a random effect for Primary Care Trust. [b]From joint log likelihood test for effect of age or deprivation as applicable. [c]From joint log likelihood ratio tests for significance of difference between patients aged ⩾70 years and patients in all other age groups. [d]From models with deprivation quintile group entered as a continuous variable.

Table 3 Lung cancer. Independent associations of age, deprivation and sex with advanced stage diagnosis (i.e., stage III/IV *vs* stage I/II)[a] ($n = 10\,435$)

	Odds ratio	Lower 95% confidence interval	Higher 95% confidence interval	*P*
Women	Reference			0.011[b]
Men	1.14	1.03	1.25	
15–49	1.33	0.93	1.90	<0.001[b] (<0.001)[c]
50–54	1.00	0.74	1.35	
55–59	1.26	0.99	1.61	
60–64	0.96	0.79	1.18	
65–69	Reference			
70–74	0.82	0.68	0.97	
75–79	0.74	0.62	0.88	
80–84	0.73	0.61	0.88	
⩾85	0.66	0.54	0.81	
Most affluent	Reference			0.290[b] (*0.236*)[d]
2	0.94	0.81	1.09	
3	0.97	0.83	1.12	
4	0.98	0.84	1.14	
Deprived	0.81	0.66	0.99	

[a]From logistic regression models, with stage II–IV *vs* stage I or stage III/IV *vs* stage I/II diagnosis as the binary outcome variable. Models were adjusted for age, sex, deprivation and tumour type, and included a random effect for Primary Care Trust. [b]From joint log likelihood test for effect of sex, age or deprivation as applicable. [c]From joint log likelihood ratio tests for significance of difference between patients aged ⩾70 years and patients in all other age groups. [d]From models with deprivation quintile group entered as a continuous variable.

Table 4 Breast cancer. Summary of outputs obtained by complete case analysis and sensitivity analyses (odds ratios for stage III/IV *vs* I/II).

	Complete case analysis[a]	Multiple imputation	Missing stage = II–IV
15–39	1.15	1.13	1.08
40–44	1.02	1.01	0.85
45–49	0.91	0.91	0.85
50–54	0.92	0.90	0.93
55–59	0.90	0.88	0.81
60–64	0.91	0.90	0.86
65–69	Reference		
70–74	1.21	1.23	1.08
75–79	1.46	1.49	1.30
80–84	1.68	1.74	1.77
⩾85	1.78	1.84	2.21
Most affluent	Reference		
2	1.16	1.20	1.12
3	1.12	1.16	1.07
4	1.29	1.32	1.21
Deprived	1.23	1.27	1.07

[a]This column replicates information included in Table 2 – presented here for ease of comparisons.

The findings were robust to multiple imputation of missing stage (under the MAR assumption). Similar patterns of variation were also observed for extreme case scenario analysis (under the MNAR assumption of missing stage = advanced stage), except that deprivation differences in advanced stage diagnosis for breast cancer were smaller.

Regarding age differences in stage at diagnosis, no apparent age patterns were apparent in a recent analysis of the US breast cancer data (CDC, 2010). For lung cancer, evidence from Denmark indicates a lower frequency of advanced stage at diagnosis with increasing age, as observed in our own study (Dalton *et al*, 2011).

For breast cancer, the observed socioeconomic differences concord with other evidence from the United Kingdom, United States and Canada, indicating a higher frequency of advanced stage at diagnosis among women of lower socioeconomic position. (Adams *et al*, 2004; Clegg *et al*, 2009; Cuthbertson *et al*, 2009; Booth *et al*, 2010). For lung cancer, studies from Canada, Denmark and Sweden have indicated only limited socioeconomic differences in advanced stage at diagnosis (Berglund *et al*, 2010; Booth *et al*, 2010; Dalton *et al*, 2011). A previous UK study reported lower frequency of advanced stage at diagnosis in more deprived patients (Brewster *et al*, 2001). The findings of our study are similar with previous UK research, although there was no independent evidence of an association (P for trend = 0.236) that may reflect the lack of power.

Table 5 Lung cancer. Summary of outputs obtained by complete case analysis and sensitivity analyses (odds ratios for stage III/IV *vs* I/II)

	Complete case analysis[a]	Multiple imputation	Missing stage = stage II–IV
Women	Reference		
Men	1.14	1.13	1.15
15–49	1.33	1.23	1.31
50–54	1.00	0.96	0.95
55–59	1.26	1.22	1.23
60–64	0.96	0.95	0.95
65–69	Reference		
70–74	0.82	0.80	0.82
75–79	0.74	0.72	0.75
80–84	0.73	0.73	0.78
⩾85	0.66	0.68	0.76
Most affluent	Reference		
2	0.94	0.97	0.95
3	0.97	1.01	0.97
4	0.98	1.04	0.99
Deprived	0.81	0.91	0.82

[a]This column replicates information included in Table 3 – presented here for ease of comparisons.

Strengths and limitations

The principal strengths of the study are its population-based design, and the high quality and completeness of information on stage at diagnosis and other tumour variables. Unlike previous studies in this field, we adjusted the analysis for tumour subtype and employed sensitivity analyses approaches using different assumptions about potential mechanisms responsible for missing stage data. Previous studies on stage at diagnosis of breast cancer did not encompass adjustment for screening or symptomatic detection status, and this factor complicated the interpretation of age and socioeconomic differences in stage at diagnosis (Macleod *et al*, 2000; Adams *et al*, 2004; Cuthbertson *et al*, 2009). In contrast, our findings indicate that substantial age and deprivation differences in stage at diagnosis of breast cancer exist independently of whether a woman was diagnosed by screening or after symptomatic presentation. A previous UK study on stage at diagnosis of lung cancer only reported on socioeconomic differences (not encompassing age and sex differences) in the mid-1990s (Brewster *et al*, 2001). Therefore, we believe the findings enrich substantially the currently available evidence on patterns of stage at diagnosis in patients with breast and lung cancer.

The study also has certain limitations. We could not adjust the analysis for ethnicity – a potential confounder of deprivation in particular. During the study period, the proportion of East of England residents belonging to ethnic minorities was relatively small, particularly among persons ⩾65 years (where the majority of cancer cases occur); ~97% of the East of England resident population in this age group were estimated as being British White in 2007 (ONS (Office for National Statistics), 2009). Given the demographic characteristics of the East of England population, the

findings can be considered to chiefly describe socio-demographic variation in stage at diagnosis among White British patients. Nevertheless, examination of patterns of stage at diagnosis by ethnic group is warranted in the future.

We examined data from a single region that includes about 10% of the total English population. Socioeconomic differences in short-term cancer survival, however, (a marker of early diagnosis) are relatively similar across different English regions (Rachet *et al*, 2009). Inequalities in cancer treatment patterns observed in East of England cancer patients are also similar to those observed nationwide (Wishart *et al*, 2010). These considerations indicate that the observed socio-demographic patterns of stage at diagnosis can be applicable to the rest of the English population. The size of the East of England population ($\sim$5.7 million) is similar to that of several European countries.

In common with previous authoritative UK research (Brewster *et al*, 2001; Adams *et al*, 2004; Rachet *et al*, 2010), we used an area-based measure of socioeconomic status in our study, relating to the population characteristics of highly homogeneous small areas (LSOA) (Woods *et al*, 2005). Socioeconomic status can be measured either directly (for example, by measuring a person's income, occupation or education) or indirectly (ecologically) by measuring the characteristics of the population of a small area (Liberatos *et al*, 1988). Both direct and area-based measures of socioeconomic status have limitations (Sloggett *et al*, 2007), and might be affected by lack of homogeneity within groups (for example, between patients of the same social class, income, education or neighbourhood) (Carstairs and Morris, 1989). Using an area-based measure of socioeconomic status may have either underestimated or overestimated socioeconomic gradients in stage at diagnosis compared with direct measures (Sloggett *et al*, 2007), and research examining such gradients using both area-based and direct measures would be useful.

Interpretation and research policy implications

A key consideration in interpreting the findings is whether the observed variation in advanced stage at diagnosis, particularly in relation to age, can be considered avoidable. In theory, the findings might in part reflect differences in the malignant potential of tumours between patients of different ages. The analysis was, however, adjusted for tumour subtype. This makes it less likely that age differences in tumour biology can be responsible for major part of the observed age differences in stage at diagnosis.

For breast cancer, it is possible that the observed variation in stage at diagnosis reflects differences in the awareness of cancer symptoms between different patient groups. Awareness of cancer symptoms and signs in the United Kingdom is socio-demographically patterned, and is lower among individuals aged >65 and of lower socioeconomic status (Robb *et al*, 2009). The findings of the study would support the targeting of breast cancer awareness interventions at older women (Forbes *et al*, 2011).

The lower frequency of advanced stage at diagnosis among older lung cancer patients could reflect more frequent use of chest *X* ray investigations in older patients (for example, in the context of investigating either a chest infection or other clinical presentations such as shortness of breath). A recent population study from Denmark indicated a lower frequency of advanced stage lung cancer diagnosis among patients with higher levels of comorbidity and also (as observed in our study) with increasing age (Dalton *et al*, 2011). Another potential explanation is that 'stage for stage' lung cancer is more symptomatic in older patients, for example, either because of a higher propensity to present with concomitant chest infection (prompting earlier investigation and leading to earlier diagnosis) or earlier presentation of dyspnoea because of physiologically declining lung capacity in older age. Further research in this area is clearly needed to explore the validity of these hypotheses, and to identify the mechanisms responsible for excess risk of advanced stage at diagnosis in relatively younger patients.

There was a substantial excess risk of advanced stage at diagnosis among breast cancer women $\geqslant$70 years of age. These differences should not be dismissed as clinically unimportant; in our study sample, one-third of women with breast cancer were aged $\geqslant$70 years. In the United Kingdom, life expectancy for women aged 70 and 80 year-old is 16.5 and 9.5 years, respectively (ONS (Office for National Statistics), 2011). Decreasing the frequency of advanced stage at diagnosis among women $\geqslant$70 years can therefore contribute substantially to reducing avoidable mortality in this age group. In contrast, the findings also identify opportunities for achieving earlier stage diagnosis of lung cancer in relatively young patients (for example, those aged 60–74 years).

CONCLUSION

There is substantial potential for improvements in early diagnosis in older patients with breast cancer and in relatively younger patients with lung cancer. The findings could help guide breast and lung cancer early diagnosis initiatives and research focused on individuals of different age groups at highest risk of advanced stage at diagnosis. These could, for example, encompass age stratified and tailored cancer symptoms awareness interventions, or educational interventions for physicians and healthcare professionals, targeted at patients of different age groups. We provide an exemplar of how population-based cancer registration information could help support national initiatives aimed at improving early diagnosis, and inform further policy and research.

ACKNOWLEDGEMENTS

We thank all staffs of the Eastern Cancer Registration and Information Centre and Dr David Gilligan for his insightful comments regarding older age advanced stage diagnosis patterns for lung cancer. GL is funded by a Post-Doctoral Research Fellowship award from the National Institute for Health Research.

REFERENCES

Abdel-Rahman M, Stockton D, Rachet B, Hakulinen T, Coleman MP (2009) What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *Br J Cancer* **101**(Suppl 2)**:** S115–S124

Adams J, White M, Forman D (2004) Are there socioeconomic gradients in stage and grade of breast cancer at diagnosis? Cross sectional analysis of UK cancer registry data. *BMJ* **329**(7458)**:** 142

Ali AM, Dawson SJ, Blows FM, Provenzano E, Ellis IO, Baglietto L, Huntsman D, Caldas C, Pharoah PD (2011) Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. *Br J Cancer* **104**(4)**:** 693–699

Berglund A, Holmberg L, Tishelman C, Wagenius G, Eaker S, Lambe M (2010) Social inequalities in non-small cell lung cancer management and survival: a population-based study in central Sweden. *Thorax* **65**(4)**:** 327–333

Booth CM, Li G, Zhang-Salomons J, Mackillop WJ (2010) The impact of socioeconomic status on stage of cancer at diagnosis and survival: a population-based study in Ontario, Canada. *Cancer* **116**(17)**:** 4160–4167

Brewster DH, Thomson CS, Hole DJ, Black RJ, Stroner PL, Gillis CR (2001) Relation between socioeconomic status and tumour stage in patients with breast, colorectal, ovarian, and lung cancer: results from four national, population based studies. *BMJ* **322:** 830 – 831

Carstairs V, Morris R (1989) Deprivation and mortality: an alternative to social class? *Community Med* **11**(3)**:** 210 – 219

CDC (Centres for Disease Control) (2010) Morbidity and Mortality Weekly Report. Surveillance of Screening-Detected Cancers (Colon and Rectum, Breast, and Cervix) - United States, 2004 – 2006. http://www.cdc.gov/mmwr/pdf/ss/ss.5909.pdf. Last accessed 23 January 2012

Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD, Goodman MT, Lynch CF, Schwartz SM, Chen VW, Bernstein L, Gomez SL, Graff JJ, Lin CC, Johnson NJ, Edwards BK (2009) Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer Causes Control* **20**(4)**:** 417 – 435

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, Adolfsson 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

Cuthbertson SA, Goyder EC, Poole J (2009) Inequalities in breast cancer stage at diagnosis in the trent region, and implications for the NHS Breast Screening Programme. *J Public Health (Oxf)* **31**(3)**:** 398 – 405

Dalton SO, Frederiksen BL, Jacobsen E, Steding-Jessen M, Osterlind K, Schüz J, Osler M, Johansen C (2011) Socioeconomic position, stage of lung cancer and time between referral and diagnosis in Denmark, 2001 – 2008. *Br J Cancer* **105**(7)**:** 1042 – 1048

ECRIC (Eastern Cancer Registration and Information Centre) (2011) Stage distribution of cancers diagnosed in 2009 in the East of England by cancer site and area of residence. http://www.ecric.nhs.uk/docs/ECRIC_incidenceXstage_2009.pdf. Last accessed 23 January 2012

Forbes LJ, Linsell L, Atkins L, Burgess C, Tucker L, Omar L, Ramirez AJ (2011) A promoting early presentation intervention increases breast cancer awareness in older women after 2 years: a randomised controlled trial. *Br J Cancer* **105**(1)**:** 18 – 21

He Y, Yucel R, Zaslavsky AM (2008) Misreporting, missing data, and multiple imputation: improving accuracy of Cancer Registry Databases. *Chance (N Y)* **21**(3)**:** 55 – 58

Liberatos P, Link BG, Kelsey JL (1988) The measurement of social class in epidemiology. *Epidemiol Rev* **10:** 87 – 121

Macleod U, Mitchell ED, Burgess C, Macdonald S, Ramirez AJ (2009) Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer* **101**(Suppl 2)**:** S92 – S101

Macleod U, Ross S, Gillis C, McConnachie A, Twelves C, Watt GC (2000) Socio-economic deprivation and stage of disease at presentation in women with breast cancer. *Ann Oncol* **11**(1)**:** 105 – 107

Møller H, Linklater KM, Robinson D (2009) A visual summary of the EUROCARE-4 results: a UK perspective. *Br J Cancer* **101**(Suppl 2)**:** S110 – S114

NAO (National Audit Office) (2010) Department of Health. Delivering the Cancer Reform Strategy. Report by the Comptroller and Auditor General. HC 568. Session 2010 – 2011. http://www.nao.org.uk/publications/1011/cancer_reform_strategy.aspx. Last accessed 23 January 2012

NCIN (National Cancer Intelligence Network) (2008a) One Year Cancer Survival Trends, England, 1985 – 2004; One Year Cancer Survival By Cancer Network, England, 2000 – 2004. http://www.ncin.org.uk/publications/reports/default.aspx. Last accessed 23 January 2012

NCIN (National Cancer Intelligence Network) (2008b) Cancer Incidence and Mortality By Cancer Network, UK, 2005. http://www.ncin.org.uk/publications/reports/default.aspx. Last accessed 23 January 2012

NCIN (National Cancer Intelligence Network) (2010) Routes to Diagnosis - NCIN Data Briefing. http://www.ncin.org.uk/publications/data_briefings/routes_to_diagnosis.aspx. Last accessed 23 January 2012

Neal RD, Allgar VL (2005) Sociodemographic factors and delays in the diagnosis of six cancers: analysis of data from the 'National Survey of NHS Patients: Cancer'. *Br J Cancer* **92**(11)**:** 1971 – 1975

Nur U, Shack LG, Rachet B, Carpenter JR, Coleman MP (2010) Modelling relative survival in the presence of incomplete data: a tutorial. *Int J Epidemiol* **39**(1)**:** 118 – 128

Office of the Deputy Prime Minister (2004) The English Indices of Deprivation 2004: Summary (revised). http://www.communities.gov.uk/documents/communities/pdf/131206.pdf. Last accessed 23 January 2012

Olesen F, Hansen RP, Vedsted P (2009) Delay in diagnosis: the experience in Denmark. *Br J Cancer* **101**(Suppl 2)**:** S5 – S8

ONS (Office for National Statistics) (2009) Population Estimates by Ethnic Group (experimental), Mid-2009. http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition = tcm%3A77 – 50029. Last accessed 23 January 2012

ONS (Office for National Statistics) (2011) England and Wales, Interim Life Tables, 1980 – 82 to 2008 – 10. http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition = tcm%3A77 – 22332. Last accessed 23 January 2012

Quaglia A, Tavilla A, Shack L, Brenner H, Janssen-Heijnen M, Allemani C, Colonna M, Grande E, Grosclaude P, Vercelli M (2009) The cancer survival gap between elderly and middle-aged patients in Europe is widening. *Eur J Cancer* **45**(6)**:** 1006 – 1016

Rachet B, Ellis L, Maringe C, Chu T, Nur U, Quaresma M, Shah A, Walters S, Woods L, Forman D, Coleman MP (2010) Socioeconomic inequalities in cancer survival in England after the NHS cancer plan. *Br J Cancer* **103**(4)**:** 446 – 453

Rachet B, Maringe C, Nur U, Quaresma M, Shah A, Woods LM, Ellis L, Walters S, Forman D, Steward J, Coleman MP (2009) Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *Lancet Oncol* **10**(4)**:** 351 – 369

RCGP (Royal College of General Practitioners) (2011) National Audit of Cancer Diagnosis in Primary Care. http://www.rcgp.org.uk/pdf/National_Audit_of_Cancer_Diagnosis_in_Primary-Care.pdf. Last accessed 23 January 2012

Richards MA (2009) The national awareness and early diagnosis initiative in England: assembling the evidence. *Br J Cancer* **101**(Suppl 2)**:** S1 – S4

Robb K, Stubbings S, Ramirez A, Macleod U, Austoker J, Waller J, Hiom S, Wardle J (2009) Public awareness of cancer in Britain: a population-based survey of adults. *Br J Cancer* **101**(Suppl 2)**:** S18 – S23

Royston P (2007) Multiple imputation of missing values: further update of ice, with an emphasis on interval censoring. *Stata J.* **7:** 445 – 464

Sloggett A, Young H, Grundy E (2007) The association of cancer survival with four socioeconomic indicators: a longitudinal study of the older population of England and Wales 1981 – 2000. *BMC Cancer* **7:** 20

Sobin LH, Wittekind CH (eds), International Union Against Cancer (UICC) (1997) *TNM Classification of Malignant Tumors*. 5th edition John Wiley & Sons Inc: New York

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338:** b2393

WHO (Word Health Organisation) (2000) International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). http://www.who.int/classifications/icd/adaptations/oncology/en/. Last accessed 23 January 2012

Wishart GC, Greenberg DC, Chou P, Brown CH, Duffy S, Purushotham AD (2010) Treatment and survival in breast cancer in the Eastern Region of England. *Ann Oncol* **21**(2)**:** 291 – 296

Woods LM, Rachet B, Coleman MP (2005) Choice of geographic unit influences socioeconomic inequalities in breast cancer survival. *Br J Cancer* **92:** 1279 – 1282

APPENDIX

Table A1 Additional details on methods of sensitivity analysis and imputation. Potential mechanisms responsible for missing stage data

Assumed mechanism	How each assumption relates to the analysis in this paper
'Missing completely at random' (MCAR): there are no systematic differences between the missing values and the observed values.	'Complete case analysis' will give unbiased (although less precise) estimates under the MCAR assumption. Said differently, complete case analysis implicitly assumes that data are 'missing completely at random'. Although this assumption does not hold (we know that stage is more likely to be missing in older patients), the potential for bias is minimised by the high level of stage data completeness.
'Missing at random' (MAR): any systematic difference between the missing and observed values can be explained by differences in observed data. Under this assumption, although patients with missing stage information may have a higher probability of being diagnosed in advanced stage compared with patients with observed stage, this probability can be estimated from the associations of stage with age, sex, tumour type and so on among patients with observed stage.	The assumption that stage data are 'missing at random' underpins sensitivity analysis using multiple imputation. This assumption becomes more reasonable by also including in imputation models variables other than those used in the analysis models (e.g., survival, grade and basis of diagnosis).[a]
'Missing not at random' (MNAR): even after information from patients with observed stage and its associations with other variables are taken into account, systematic differences remain between patients with missing and observed stage. For example, because more advanced stage at diagnosis is more likely to remain unobserved.	The assumption that stage data are 'missing not at random' underpins sensitivity analysis using substitution of unknown stage values with advanced stage. We do not expect this extreme case scenario to be true, but it illustrates how sensitive the complete case and multiple imputation analyses may be to the MCAR or the MAR assumptions, respectively.

[a]When only outcome data are missing (e.g., on patient stage), complete case analysis will give unbiased estimates under the assumption that data are 'missing at random' when the missing outcome is dependent only on variables included in the analysis model. This assumption is more reasonable than the 'missing completely at random' one, but may still not hold; however, it can become even more reasonable by including additional variables in the imputation models, as applied in this study.

Further details on imputation

We used the STATA *ice* command for multiple imputation (see reference by Royston *et al* (2007) of main paper). For (female) breast cancer, imputation models included information on age, deprivation, tumour type, screening detection status, tumour grade, oestrogen receptor status, histological verification status, Primary Care Trust and survival. For lung cancer, imputation models included information on age, sex, deprivation, tumour type, tumour grade, histological verification status and Primary Care Trust and survival.

Stage was treated as a binary variable (stage III/IV *vs* I/II). Grade was treated as an ordinal variable with four levels for colorectal and lung cancer, and three levels for breast cancer. All other variables except for survival were treated as categorical. For survival we used the Nelson–Aalen estimate of the cumulative hazard function along with an indicator variable describing vital status at end of follow-up.

Informed by considerations of the proportion of missing data for different variables, 50 imputed data sets were generated for either cancer. All imputed data sets were each analysed separately and then combined using Rubin's rules, using the STATA *mim* command.

Analysis models on the imputed data sets included the same variables as those used in the analysis models using the complete case analysis approach, that is, age, deprivation, tumour type and screening detection status for breast cancer, and age, sex, deprivation and tumour type for lung cancer, including a random effect for Primary Care Trust.

Table A2 Predictors of missing stage

	Total	Staged	% staged	p (χ^2)
(a) Breast cancer				
Affluent	4778	4385	92	0.490[a]
2	4658	4321	93	
3	4323	4007	93	
4	3081	2809	91	
Deprived	996	938	94	
15–39	770	709	92	<0.001[b]
40–44	1091	1036	95	
45–49	1539	1437	93	
50–54	2048	1930	94	
55–59	1911	1832	96	
60–64	2461	2350	95	
65–69	2152	2036	95	
70–74	1491	1393	93	
75–79	1590	1458	92	
80–84	1321	1133	86	<0.001[c]
⩾85	1462	1146	78	
Infiltrating ductal carcinoma	12826	12030	94	<0.001[b]
Lobular carcinoma	2099	1922	92	
Mixed ductal lobular	1211	1164	96	
Other adenocarcinoma	709	653	92	
Other specified carcinoma	89	79	89	
Other unspecified	863	609	71	
Specified not carcinoma	39	3	8	
All patients	17836	16460	92	
(b) Lung cancer				
Men	5602	4392	78	0.736[a]
Women	7684	6043	79	
Affluent	2471	1900	77	0.009[b]
2	3072	2402	78	
3	3444	2734	79	
4	3072	2397	78	
Deprived	1227	1002	82	
15–49	380	287	76	<0.001[a]
50–54	443	359	81	
55–59	903	743	82	
60–64	1525	1248	82	
65–69	1762	1416	80	
70–74	2166	1759	81	
75–79	2384	1899	80	
80–84	2099	1597	76	<0.001[c]
⩾85	1624	1127	69	
Adenocarcinoma	2366	1901	80	<0.001[a]
Carcinoid	100	16	16	
Large cell carcinoma	145	128	88	
Other non-small cell	2475	2117	86	
Small cell carcinoma	1464	1150	79	
Specified other	10	2	20	
Squamous cell carcinoma	2351	2040	87	
Unspecified other	4375	3081	70	
All patients	13286	10435	79	

(a) [a]From univariate logistic regression for stage completeness, with deprivation quintile group entered as a continuous exposure variable. [b]From χ^2-test. [c]From log likelihood ratio tests for significance of difference between the 'older' age groups (i.e. age groups ⩾70 years) and other age groups. (b) [a]From χ^2-test. [b]From univariate logistic regression for stage completeness, with deprivation quintile group entered as a continuous variable. [c]From log likelihood ratio tests for significance of difference between the 'older' age groups (i.e., age groups ⩾70 years) and other age groups.

Table A3 Findings in relation to variation in breast cancer diagnosis at stage I *vs* stages II–IV

	Stage II-IV *vs* stage I			
	Odds ratio	**Lower 95% confidence interval**	**Higher 95% confidence interval**	***P***
Breast cancer				
15–39	2.04	1.69	2.47	<0.001[a]
40–44	1.67	1.42	1.96	
45–49	1.57	1.35	1.81	
50–54	1.30	1.14	1.48	
55–59	1.23	1.08	1.41	
60–64	1.06	0.93	1.20	
65–69	Reference			
70–74	1.45	1.25	1.67	(<0.001[b])
75–79	1.70	1.47	1.97	
80–84	1.99	1.69	2.34	
⩾85	2.41	2.04	2.86	
Most affluent	Reference			0.335[a]
2	1.03	0.94	1.13	(*0.172*[c])
3	1.03	0.94	1.13	
4	1.11	1.00	1.24	
Deprived	1.00	0.86	1.17	

Independent associations of age and deprivation with diagnosis in stage I *vs* II–IV[d] ($n = 16460$) [a]From joint log likelihood test for effect of age or deprivation as applicable. [b]From joint log likelihood ratio tests for significance of difference between patients aged ⩾70 years and patients in all other age groups. [c]From models with deprivation quintile group entered as a continuous variable. [d]From logistic regression models, with diagnosis in stage II–IV *vs* stage I as the binary outcome variable. Models were adjusted for age, deprivation, tumour type and diagnosis through screening or symptomatic presentation, and included a random effect for Primary Care Trust.

Table A4 Findings in relation to variation in lung cancer diagnosis at stage I *vs* stages II–IV

	Complete case analysis			Multiple imputation				Missing stage = stage III/IV		
	OR	**95% LCI**	**95% UCI**	**OR**	**95% LCI**	**95% UCI**	**FMI**	**OR**	**95% LCI**	**95% UCI**
Women	Ref.	—	—	Ref.	—	—	—	Ref.	—	—
Men	1.14	1.03	1.25	1.13	1.03	1.25	0.170	1.15	1.05	1.27
15–49	1.33	0.93	1.90	1.23	0.86	1.75	0.197	1.31	—	—
50–54	1.00	0.74	1.35	0.96	0.71	1.30	0.157	0.95	0.93	1.84
55–59	1.26	0.99	1.61	1.22	0.96	1.55	0.119	1.23	0.71	1.27
60–64	0.96	0.79	1.18	0.95	0.78	1.16	0.150	0.95	0.97	1.56
65–69	Ref.	—	—	Ref.	—	—	—	—	0.78	1.15
70–74	0.82	0.68	0.97	0.80	0.68	0.96	0.109	0.82	0.69	0.98
75–79	0.74	0.62	0.88	0.72	0.60	0.86	0.171	0.75	0.64	0.89
80–84	0.73	0.61	0.88	0.73	0.61	0.87	0.16	0.78	0.65	0.93
⩾85	0.66	0.54	0.81	0.68	0.55	0.83	0.23	0.76	0.62	0.92
Most affluent	Ref.	—	—	Ref.	—	—	—	Ref.	—	—
2	0.94	0.81	1.09	0.97	0.84	1.12	0.138	0.95	0.82	1.10
3	0.97	0.83	1.12	1.01	0.87	1.17	0.184	0.97	0.84	1.12
4	0.98	0.84	1.14	1.04	0.89	1.21	0.209	0.99	0.86	1.15
Deprived	0.81	0.66	0.99	0.91	0.75	1.10	0.186	0.82	0.67	0.99
Adenocarcinoma	Ref.	—	—	Ref.	—	—	—	Ref.	—	—
Squamous cell carcinoma	0.91	0.79	1.05	0.89	0.77	1.02	0.116	0.83	0.72	0.95
Other non-small cell types	2.07	1.77	2.42	1.97	1.70	2.29	0.099	1.87	1.61	2.18
Small cell carcinoma	4.06	3.23	5.12	3.90	3.10	4.92	0.207	3.94	3.14	4.94
Large cell carcinoma	1.51	0.97	2.36	1.44	0.93	2.22	0.065	1.29	0.83	1.99
Carcinoid	0.02	0.00	0.18	0.02	0.00	0.15	0.764	1.53	0.87	2.70
Specified other	0.41	0.03	6.63	0.74	0.06	9.27	0.627	2.30	0.29	18.35
Unspecified other	1.94	1.67	2.24	1.85	1.59	2.15	0.289	2.07	1.80	2.37

Abbreviations: OR = odds ratio; Ref = reference; LCI = lower confidence interval; UCI = upper confidence interval; FMI = fraction of missing information (for each respective variable category, it denotes the proportion of the estimation that used imputed missing information).

Table A5 Full outputs of all analysis models presented in main paper (for stage III/IV *vs* I/II comparisons)

	Complete case analysis			Multiple imputation				Missing stage = stage III/IV		
	OR	**95% LCI**	**95% UCI**	**OR**	**95% LCI**	**95% UCI**	**FMI**	**OR**	**95% LCI**	**95% UCI**
(a) Breast cancer, stage III/IV vs stage I–II										
15–39	1.15	0.89	1.48	1.13	0.88	1.46	0.062	1.08	0.87	1.34
40–44	1.02	0.81	1.28	1.01	0.80	1.27	0.069	0.85	0.70	1.05
45–49	0.91	0.74	1.14	0.91	0.73	1.13	0.088	0.85	0.70	1.02
50–54	0.92	0.74	1.14	0.90	0.72	1.11	0.066	0.93	0.78	1.11
55–59	0.90	0.72	1.12	0.88	0.71	1.09	0.058	0.81	0.67	0.98
60–64	0.91	0.74	1.12	0.89	0.73	1.10	0.076	0.86	0.72	1.02
65–69	Ref.			Ref.				Ref.		
70–74	1.21	0.98	1.49	1.22	0.99	1.50	0.064	1.08	0.90	1.29
75–79	1.46	1.20	1.78	1.50	1.23	1.82	0.062	1.30	1.09	1.54
80–84	1.68	1.37	2.07	1.75	1.43	2.15	0.115	1.77	1.49	2.10
⩾85	1.78	1.45	2.18	1.86	1.52	2.27	0.128	2.21	1.87	2.62
Most affluent	Ref.			Ref.				Ref.		
2	1.16	1.02	1.32	1.20	1.05	1.36	0.107	1.12	1.00	1.25
3	1.12	0.98	1.28	1.16	1.02	1.32	0.091	1.07	0.95	1.19
4	1.29	1.12	1.49	1.32	1.15	1.52	0.116	1.21	1.07	1.36
Deprived	1.23	1.00	1.52	1.27	1.03	1.57	0.149	1.07	0.89	1.29
Infiltrating ductal carcinoma	Ref.			Ref.				Ref.		
Lobular carcinoma	1.59	1.39	1.81	1.62	1.42	1.84	0.061	1.54	1.38	1.73
Mixed ductal lobular	1.09	0.90	1.32	1.10	0.91	1.33	0.051	0.97	0.82	1.15
Other adenocarcinoma	0.99	0.79	1.25	0.98	0.78	1.23	0.058	0.99	0.82	1.20
Other specified carcinoma	0.58	0.26	1.26	0.58	0.27	1.27	0.118	0.90	0.52	1.54
Other unspecified	3.90	3.26	4.66	4.01	3.37	4.77	0.254	4.57	3.93	5.32
Specified not carcinoma	3.57	0.28	46.04	1.78	0.14	22.25	0.870	81.48	19.36	342.93
Screening detection status- no	Ref.									
Screening detection status-yes	0.26	0.22	0.31	0.27	0.22	0.32	0.030	0.20	0.17	0.24
(b) Lung cancer, odds ratios of stage III/IV vs stage I/II										
Women	Ref.			Ref.				Ref.		
Men	1.14	1.03	1.25	1.13	1.03	1.25	0.170	1.15	1.05	1.27
15–49	1.33	0.93	1.90	1.23	0.86	1.75	0.197	1.31		
50–54	1.00	0.74	1.35	0.96	0.71	1.30	0.157	0.95	0.93	1.84
55–59	1.26	0.99	1.61	1.22	0.96	1.55	0.119	1.23	0.71	1.27
60–64	0.96	0.79	1.18	0.95	0.78	1.16	0.150	0.95	0.97	1.56
65–69	Ref.			Ref.					0.78	1.15
70–74	0.82	0.68	0.97	0.80	0.68	0.96	0.109	0.82	0.69	0.98
75–79	0.74	0.62	0.88	0.72	0.60	0.86	0.171	0.75	0.64	0.89
80–84	0.73	0.61	0.88	0.73	0.61	0.87	0.16	0.78	0.65	0.93
⩾85	0.66	0.54	0.81	0.68	0.55	0.83	0.23	0.76	0.62	0.92
Most affluent	Ref.			Ref.				Ref.		
2	0.94	0.81	1.09	0.97	0.84	1.12	0.138	0.95	0.82	1.10
3	0.97	0.83	1.12	1.01	0.87	1.17	0.184	0.97	0.84	1.12
4	0.98	0.84	1.14	1.04	0.89	1.21	0.209	0.99	0.86	1.15
Deprived	0.81	0.66	0.99	0.91	0.75	1.10	0.186	0.82	0.67	0.99
Adenocarcinoma	Ref.			Ref.				Ref.		
Squamous cell carcinoma	0.91	0.79	1.05	0.89	0.77	1.02	0.116	0.83	0.72	0.95
Other non-small cell types	2.07	1.77	2.42	1.97	1.70	2.29	0.099	1.87	1.61	2.18
Small cell carcinoma	4.06	3.23	5.12	3.90	3.10	4.92	0.207	3.94	3.14	4.94
Large cell carcinoma	1.51	0.97	2.36	1.44	0.93	2.22	0.065	1.29	0.83	1.99
Carcinoid[a]	0.02	0.00	0.18	0.02	0.00	0.15	0.764	1.53	0.87	2.70
Specified other[a]	0.41	0.03	6.63	0.74	0.06	9.27	0.627	2.30	0.29	18.35
Unspecified other	1.94	1.67	2.24	1.85	1.59	2.15	0.289	2.07	1.80	2.37

Abbreviations: FMI = fraction of missing information (for each respective variable category, it denotes the proportion of the estimation that used imputed missing information); LCI = lower confidence interval; OR = odds ratio; Ref = reference; UCI = upper confidence interval. [a]For these two groups, large differences are apparent between the analysis under the missing stage = stage IV analysis and either complete case analysis or multiple imputation. Both these groups were small and had a particularly small proportion of patients with observed stage (<20%), most of whom were in stage I/II. The above indicate that the missing stage = stage IV assumption for patients with missing stage in these two groups is unlikely to be reasonable; we nevertheless present findings for consistency.