View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

British Journal of Cancer (2012) 106, 1910 – 1916
© 2012 Cancer Research UK All rights reserved 0007 – 0920/12
www.bjcancer.com

Histological confirmation of breast cancer registration and self-reporting in England and Wales: a cohort study within the UK Collaborative Trial of Ovarian Cancer Screening

EO Fourkala[1], A Gentry-Maharaj[1], M Burnell[1], A Ryan[1], R Manchanda[1], A Dawnay[1], I Jacobs[1], M Widschwendter[1] and U Menon*[,1]

[1]Gynaecological Cancer Research Centre, UCL EGA Institute for Women's Health, University College London, 149 Tottenham Road, London W1T 7DN, UK

BACKGROUND: In research studies, accurate information of cancer diagnosis is crucial. In women with breast cancer (BC), we compare cancer registration (CR) in England/Wales and self-reporting with independent confirmation.

METHODS: In the UK Collaborative Trial of Ovarian Cancer Screening, notification of BC diagnosed between randomisation and 31 December 2009 was obtained through (1) CR (17 October 2011) and (2) self-reporting using postal-questionnaire. Breast cancer was confirmed using a detailed questionnaire (BC questionnaire BCQ) completed by the treating clinician (gold standard). Apparent sensitivity and positive-predictive value of CR/self-reporting *vs* BCQ were calculated.

RESULTS: Of 1065 women with possible BC notification, diagnosis was confirmed in 932 (87.5%). A total of 3.1% (28 out of 918) of BC CR and 12.4% (128 out of 1032) of women with self-reported BC only had *in-situ* carcinoma on BCQ. Another 4.6% (43 out of 932) of BCQ-confirmed cancer did not have a BC registration, and 3.6% (34 out of 932) did not self-report BC. Apparent sensitivity of CR and self-reporting *vs* BCQ were 95.4 and 96.4%, respectively. Positive-predictive value of self-reporting (87.1%) was significantly lower than that of CR (96.8%). Women aged < 65 were more likely to over report *in-situ* carcinoma as BC. Overall, 73 (6.8%) women would have been misclassified/missed if CR, and 167 (15.6%) if self-reporting data alone was used.

CONCLUSION: This study confirms the reliability of BC registration in England/Wales and highlights the fact that 1 in 10 women self-reporting BC might only have *in-situ* breast carcinoma.

British Journal of Cancer (2012) **106,** 1910–1916. doi:10.1038/bjc.2012.155 www.bjcancer.com
Published online 17 May 2012
© 2012 Cancer Research UK

Keywords: breast cancer; cancer registry; self-reporting; UKCTOCS

National cancer registries collect comprehensive cancer information for the whole population, which enables documentation of historical trends in cancer incidence/survival. In the UK, there are three registries covering England and Wales, Scotland and Northern Ireland, respectively. New diagnosis of cancer occurring in the populations is acquired from a variety of sources, and checked for validity and completeness through a complex process of clinical data linkage and consolidation. Overall, the data on cancer registrations (CRs) has been shown for the most part to be reliable (Gulliford *et al*, 1993). Few major errors in the International Classification of Diseases (ICD) coding have been reported (Lapham and Waugh, 1992), with the data regarding cancer stage, grade and date of treatment being less consistent. Brewster *et al* (2002) concluded that even though the quality of cancer registry data may be good, improvements are needed in standardising the recording of information by clinicians.

Cancer registries are often used in research studies where cancer diagnosis and mortality are the key outcome measures or eligibility criteria. To compensate for delays or lack of completeness in CR, researchers often use additional sources such as self-reporting using questionnaires or assessment of medical records. Questionnaires are regarded as the most cost-effective way in obtaining these data (Abraham *et al*, 2009). However, the validity of self-reporting seems dependent on the site of cancer, with breast cancer (BC) having the best sensitivity (Paganini-Hill and Chao, 1993; Schrijvers *et al*, 1994; Bergmann *et al*, 1998; Desai *et al*, 2001; Parikh-Patel *et al*, 2003; Manjer *et al*, 2004; Dominguez *et al*, 2007; Brewster and Stockton, 2008), and on a variety of factors such as age, education, previous family history and race (Schrijvers *et al*, 1994; Bergmann *et al*, 1998; Desai *et al*, 2001; Parikh-Patel *et al*, 2003; Manjer *et al*, 2004; Dominguez *et al*, 2007; Abraham *et al*, 2009). The abstraction of clinical information from medical reports obtained directly from the clinicians treating the patient is the most accurate means of collecting cancer data (Schootman *et al*, 2005). However, it can be extremely time-consuming and expensive, especially across multiple hospitals (Phillips *et al*, 2005).

Most reports on accuracy of cancer data have compared two information sources (self-reported data, cancer registry records or medical notes). Only two studies have looked at all three sources of cancer data. Bergmann *et al* (1998) compared accuracy of

*Correspondence: Professor U Menon, E-mail: u.menon@ucl.ac.uk
Received 5 January 2012; revised 20 March 2012; accepted 22 March 2012; published online 17 May 2012

self-reported cancer diagnoses with population-based cancer registry data and investigated self-reports of cancer that were not confirmed by the registries using medical records. In a more recent large study of women attending the US mammography facilities between 1996 and 2006, Abraham *et al* (2009) compared self-reporting with data from cancer registries and pathology databases. Both studies were in the US populations, and there is no data we are aware of from England that has explored all three sources together.

In the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), all women were flagged through the cancer registry and were sent questionnaires that included specific questions regarding BC. Clinicians were contacted to confirm any BC notification. Given the wide implications of reliability of cancer diagnosis, we compare each (registration and self-reporting of BC) with histopathological confirmation by the clinician. In addition, we elucidate possible causes of errors and discrepancies; investigate the effect of time on BC registration delays, and examine the association between BC self-reporting and previously described factors such as age, education and family history of BC.

METHODS

Subjects

The UKCTOCS is a multi-centre, randomised controlled trial for ovarian cancer screening in postmenopausal women from the general population in England, Wales and Northern Ireland (Menon *et al*, 2008). A total of 189 063 women were recruited into the trial from England and Wales between April 2001 and September 2005.

BC cases

The BC cases were women residing in England and Wales, who were diagnosed with possible invasive BC in the period between randomisation and December 2009. Breast cancers diagnosed after women had completed their follow-up questionnaire were excluded. Women recruited from Northern Ireland were excluded as data from the Northern Ireland cancer registry was incomplete at start of study.

Identification of possible BC cases

Cancer registry All participants are being followed up through a 'flagging study' with the National Hospital Service (NHS) Information Centre for Health and Social Care (formerly Office for National Statistics) in England and Wales. The linkage was performed using the NHS number, the surname, the address and date of birth of the volunteer. The UKCTOCS co-ordinating centre receives electronic data that includes cancer registration on each subject from the NHS Information Centre for Health and Social Care every 6 months. The relevant ICD codes for malignant neoplasm of breast used for this analysis were C50*- ICD-Code 10 and 174*- ICD-Code 9. Women with ICD codes for benign conditions or *in-situ* carcinomas of the breast were not included. When women had two cancer registry records with different dates of diagnosis, the earliest date of diagnosis was used.

Self-report through the UKCTOCS follow-up questionnaire The UKCTOCS protocol included an 11-item follow-up questionnaire (FUQ) sent 3–5 years after randomisation. It included a specific question related to BC diagnosis since randomisation (Figure 1). The overall UKCTOCS FUQ response rate was 71.2% (134 602 out of 189 063). For the self-reported BCs, it was not possible to distinguish between a recurrence and second primary.

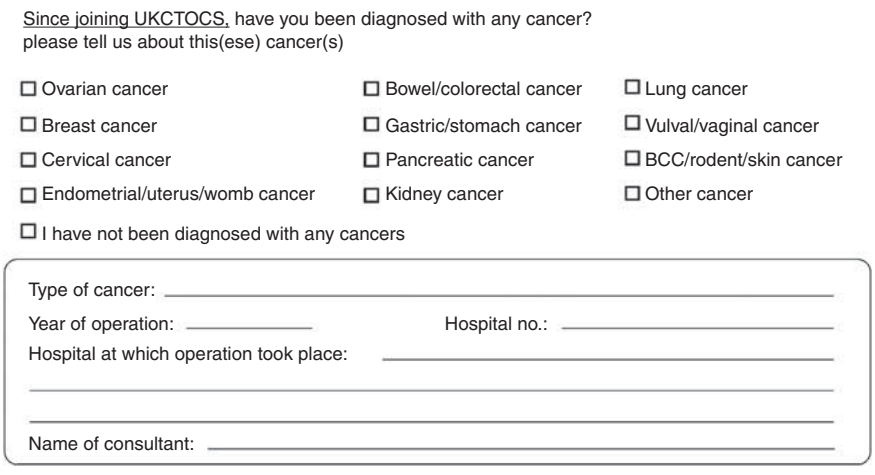
Report from the UKCTOCS trial centres A few notifications were also received directly from the participating trial centres of women, identified as a result of ovarian cancer screening using annual serum CA125 in the trial.

Confirmation of BC diagnosis

The treating physician of all women who were identified to have developed BC after randomisation, and who had provided contact details of their clinician were sent BC Questionnaire (BCQ, Supplementary Figure 1). This 15-item BCQ included questions on diagnosis, histopathology and treatment, and was used to confirm diagnosis.

Data analysis

Performance characteristics All women for whom it was possible to obtain data from the three sources were included in the analysis. If both BC and *in-situ* carcinoma of breast was reported in the same woman, BC diagnosis was used for comparisons. Analysis was undertaken comparing cancer registry and self-reported data with the gold standard-BCQ. Misclassifications were identified for cancer registry and self-reporting individually. The true-positive, false-positive, false-negative and true-negative (TN) were assessed, and the apparent sensitivity and positive-predictive value (PPV) of each data source was calculated as shown in Figure 2. Apparent sensitivity was used, as it was not possible to contact physicians of


Since joining UKCTOCS, have you been diagnosed with any cancer?
please tell us about this(ese) cancer(s)

☐ Ovarian cancer ☐ Bowel/colorectal cancer ☐ Lung cancer
☐ Breast cancer ☐ Gastric/stomach cancer ☐ Vulval/vaginal cancer
☐ Cervical cancer ☐ Pancreatic cancer ☐ BCC/rodent/skin cancer
☐ Endometrial/uterus/womb cancer ☐ Kidney cancer ☐ Other cancer
☐ I have not been diagnosed with any cancers

Type of cancer: _____
Year of operation: _____ Hospital no.: _____
Hospital at which operation took place: _____
_____
_____
Name of consultant: _____

Figure 1 Question on BC diagnosis included in the follow-up questionnaire sent to the women participating in the trial.

© 2012 Cancer Research UK

all women to identify those with BC, who had not self-reported and had no CR (TN). Fisher's test was used to compare sensitivities and PPVs of cancer registry and self-report. Combining cancer registry and self-reported data was explored using the following rules: (a) BC case is correctly reported if both sources concurred for BC diagnosis and (b) BC case is correctly reported if either source (cancer registry or self-report) reported BC diagnosis.

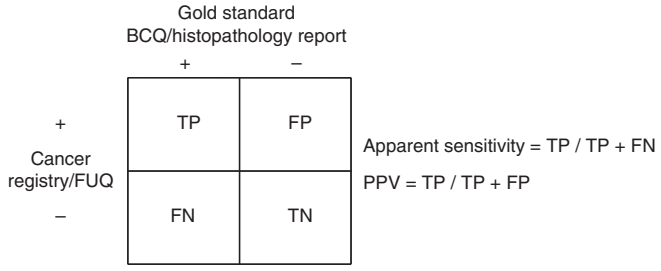


Figure 2 Description of performance characteristics used for data analysis. Calculation of apparent sensitivity and positive predictive value (PPV) for cancer registry and follow-up questionnaire (FUQ) data using the BC questionnaire (BCQ)/ histopathology report as the gold standard. Abbreviations: FN = false negative; FP = false positive; TN = true negative; TP = true positive.

CR delays The effect of time on cancer-registration delays was assessed by looking at the completeness of relevant CRs according to the year of diagnosis and time from diagnosis to cancer-registry notification.

Performance characteristics in relation to baseline characteristics The baseline characteristics of the study women were calculated using descriptive statistics. The following factors were analysed to investigate whether self-reporting is dependent on them: ethnicity (White, Black, others) age at follow-up, BC family history (no and yes, including first- and second-degree relatives), education (high: college/university degree or nursing/teaching qualification, low: A/O-level, vocational qualifications such as clerical and commercial/ or equivalent, other: none of the above), smoking (have you ever been a smoker – yes, no), alcohol consumption each week (units per week are provided), pill use (yes, no), sterilisation (yes, no), hysterectomy (yes, no) and HRT use at recruitment (yes, no). Apparent sensitivity was modelled using logistic regression, with the above characteristics as the independent variables, and using only those cases where the BCQ confirmed BC. All four variables were suitably categorised before modelling, and from the regression, the respective odds ratio and significance levels were estimated, given the other variables' presence in the equation. Positive-predictive value was also modelled in exactly the same way. Analysis was carried out using a computer-assisted programme SPSS version 12.0.1 (Chicago, IL, USA).

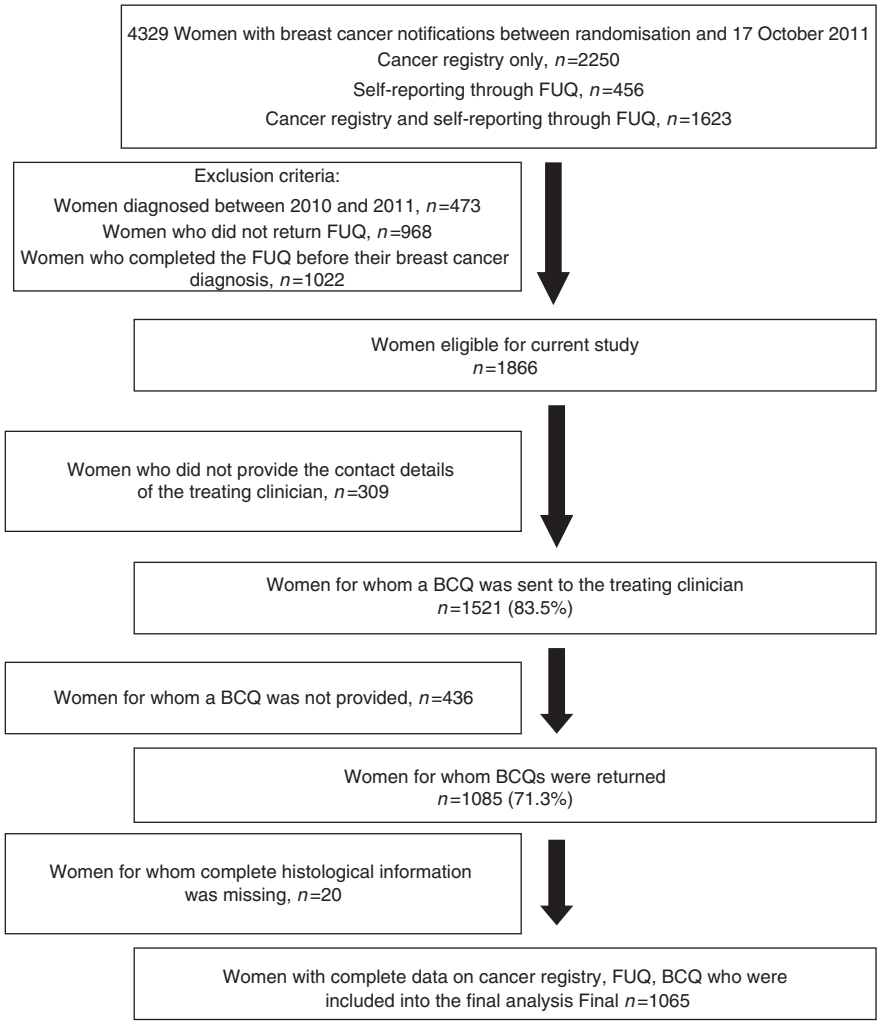


Figure 3 Diagram showing how the study subjects were identified. Abbreviations: BCQ = BC questionnaire; FUQ = follow-up questionnaire.

© 2012 Cancer Research UK

RESULTS

In the trial centre, 4329 notifications of possible BC diagnoses were received between April 2001 and October 2011. A total of 2463 women were excluded from the study, as they were diagnosed between 2010 and 2011 (473), completed the FUQ before their BC diagnosis (1022) or had not returned their FUQ questionnaires (968). Of the 1866 eligible women, it was possible to obtain a complete set of data, including self-reporting, CR and clinician confirmation using BCQ/histopathology for 1065 women (Figure 3). The baseline characteristics of these women are detailed in Table 1. Of these 1065 women, BCQ confirmed invasive BC in 932 (87.6%), *in-situ* ductal carcinoma of the breast in 126, *in-situ* lobular carcinoma of the breast in 2, benign changes in 3, and no breast pathology in 2. The median age at diagnosis of the 932 women with BCQ-confirmed BC was 63 years (95% CI: 62.77–63.53 years).

Table 1 Distribution and frequency of sociodemographic characteristics of all the study subjects ($N = 1065$)

	Distribution	
Characteristics	**No of women**	**%**
Age at follow-up		
>65	523	49.1
<65	536	50.3
Missing	6	0.6
Age at diagnosis[a]		
>65	538	57.3
<65	393	42.2
Missing	1	0.5
Ethnicity		
White	1046	98.2
Black	7	0.7
Other	7	0.7
Missing	5	0.5
Education		
Other	281	26.4
Low	385	36.1
High	352	33.1
Missing	47	4.4
Alcohol (units per week)		
0	205	19
1–3′	570	54
4–7′	261	25
Missing	29	3
Smoking		
Yes	469	44
Breast cancer family history at recruitment		
Yes	290	27.2
Pill use		
Yes	646	60.7
Sterilisation		
Yes	180	16.9
Hysterectomy		
Yes	185	17.4
HRT use at recruitment		
Yes	295	27.7
Year of diagnosis[a]		
2001	2	0.21
2002	32	3.43
2003	89	9.55
2004	187	20.1
2005	250	26.8
2006	230	24.7
2007	99	10.6
2008	40	4.29
2009	2	0.21
Not known	1	0.11

[a]For the 932 women with confirmed breast cancer diagnosis.

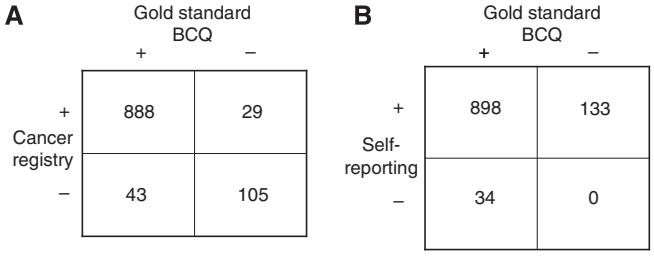
A total of 918 women had BC registration on 17 October 2011. In all, 3.1% (28 out of 918) of women with BC registrations had only evidence of *in-situ* carcinoma of the breast, and 0.1% (1/918) had only evidence of benign breast pathology. Of the 932 confirmed BCs on BCQ, 43 (4.6%) did not have a BC registration, despite having a confirmed diagnosis a median of 6.8 years (inter-quartile range: 4.9 years) before the last cancer registry follow-up (Figure 4A). The apparent sensitivity of cancer registry was 95.4% (93.4 to 96.2) and PPV was 96.8% (95.3 to 97.8; Figure 4C).

A total of 1032 women self-reported BC. In all, 12.4% (128 out of 1032) women self-reporting BC only had *in-situ* carcinoma of the breast on BCQ; 3.6% (34/932) of women did not report BC (Figure 4B), despite having a confirmed diagnosis a median of 1.6 years (inter-quartile range: 2.3 years) before completion of FUQ. The apparent sensitivity of self-reporting on the FUQ was 96.4% (95.3 to 97.3) and PPV was 87.1% (86.5 to 90.5; Figure 4C).

Out of 932 women with confirmed BC diagnosis on BCQ, both cancer registry and self-reporting concurred in 819 (87.9%) women diagnosed with BC. Overall, in the cohort of 1065 women, 73 (6.8%) women would have been misclassified/missed if cancer registry data alone was used, and 167 (15.6%) if self-reporting data alone was used. When the rule that BC case is confirmed if either source (cancer registry or self-reporting through the FUQ)-reported BC diagnosis was applied, there were 210 women (19.7%) who would have been misclassified or missed. The rule that both sources (cancer registry and self-reporting) need to concur for BC diagnosis gave the lowest rate of misclassifications with 30 (2.7%) women who would have been falsely identified as BC cases (Table 2).

Cancer registry delays

We investigated whether delays in cancer registry might account for the lack of CR on 17 October 2011. Between 2001 and 2002, there were no women diagnosed with BC with a missing registration code. The highest percentage of women 30% (12/40) with no CR were those diagnosed in the year 2008. Only three women diagnosed in 2009 were included in this analysis, and all of them had a CR (Figure 5).



C

Performance characteristics	Cancer registry	Self-reporting	*P*-value
	BCQ (Gold standard)		
% Apparent sensitivity (95% CI)	95.4 (93.4 to 96.2)	96.4 (94.9 to 97.3)	0.21
% PPV (95% CI)	96.8 (95.3 to 97.8)	87.1 (86.5 to 90.5)	<0.0001

Figure 4 Performance characteristics for cancer registry and UKCTOCS follow-up questionnaire (FUQ). Numbers of true positives (TP), false positives (FP), false negatives (FN) and true positives (TP) for BC cases identified within UKCTOCS. Comparison with gold standard (BQC/histopathology) of (**A**) Cancer Registry and (**B**) self-reporting through FUQ. (**C**) Calculation of sensitivity and positive predictive value (PPV).

© 2012 Cancer Research UK

Table 2 Identified misclassifications/errors on comparison to BCQ and their causes. The error is dependent on the data source and how it is interpreted

	Data source and interpretation			
Cause of misclassification	**Self-report**	**Cancer registry**	**Cancer registry and self-reporting (both need to concur for breast cancer diagnosis)**	**Cancer registry and self-reporting (breast cancer diagnosis, if either report the diagnosis)**
In-situ carcinoma of the breast or benign condition or no cancer misclassified as breast cancer or other cancer	133	30	29	133
Breast cancer misclassified as *in-situ* carcinoma of the breast or benign condition or other cancer	0	14[a]	0	14[a]
Breast cancer not reported/registered	34	29	1	63
Total number (%) of errors (either misclassified or not reported/registered)	167	73	30	210

[a]A total of 11 women (1.1%) with *in-situ* ductal carcinoma of the breast registration of whom 2 had a second registration for another cancer and 3 (0.3%) other cancer alone.

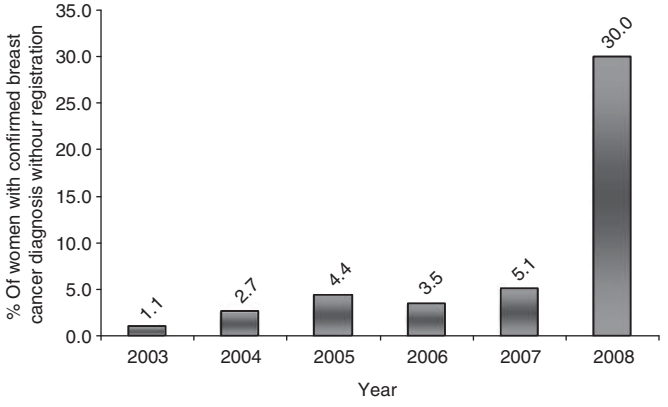


Figure 5 Proportion of women with confirmed BC without a registration code per year, based on the last cancer registry follow-up (17 October 2011).

Apparent sensitivity and PPV of BC by women self-reporting based on characteristics

Age at completion of FUQ was a significant determinant for PPV but not for apparent sensitivity, with women <65 in comparison with women >65 over-reporting their BC diagnosis. All the other factors investigated were not shown to be associated with self-reporting (Table 3).

DISCUSSION

This is the first study in England and Wales to examine the performance characteristics of both self-reporting and CR for BC against the gold standard of confirmation by the treating physician. It confirms the reliability of BC registration in England and Wales. Our data also suggests that although self-reporting using postal questionnaires is a good source of cancer data, 1 in 10 women self-reporting BC may only have *in-situ carcinoma*. This finding highlights the need for better patient information regarding the differences between *in-situ* carcinomas and invasive disease. It also points out that on surveys when women are questioned about their BC diagnoses, a separate question on *in-situ* carcinomas should be included.

One of the key strengths of this study is that CR data was available for 99.99% of the cohort. Eight women refused consent to flagging. We were able to 'flag' all women recruited from England and Wales who gave consent with the NHS Information Centre for Health and Social Care. This was due to having accurate NHS numbers as a result of electronic transfer of the information from Primary Care age–sex registers (Menon *et al*, 2008). Additional strengths include over 70% of women in the cohort completing the self-reporting questionnaire (FUQ) and confirmation of BC diagnosis by contacting the consultants responsible for treatment with high response rates (70%). There is a time lag between diagnosis and CR, and this was taken into account by including CR records till October 2011, but limiting the analysis to cancers diagnosed till December 2009. A limitation was that at the time of analysis, in women with a BC notification where the woman had not provided the name of the breast specialist, the GP had not been contacted for confirmation. This analysis does not distinguish between women who have two primary cancers or a primary cancer with recurrence. It has been done per woman and not per primary BC diagnosis. Additionally, it was not possible to contact the GP of all the 189 063 women in the cohort to identify any women with BC, who had neither self-reported nor had a CR. Hence, we are unable to estimate the number of cases with no CR where women too had not self-reported. Recent data suggests this effect may be small. On linking CR with information from the Hospital Episode Statistics (HES) database for BC patients in England for the period 2001–2007, Moller *et al* (2011) found that HES data only added 2.0% to the number of BC registrations. We have recently obtained HES data for the women in the cohort from England and hope to explore this in due course.

Cancer registry has high sensitivity (95.4%) and PPV (96.8%), for BC and researchers can rely on this data. The apparent sensitivity reported in this study is comparable to the most recent report of 98.0% by Brewster *et al* (2008) investigating the Scottish cancer registry. The rates are similar to Gathani *et al* (2005) who reported a sensitivity of 96% for BC diagnosis on comparing cancer registry data with the NHS Breast Screening Programme in the largest study so far in England, including more than 5000 BC cases. Overall, though there are few studies in England on validation of CR data and most of them include only a small number of BC cases, the data suggests that sensitivity of cancer registry for BC has improved over the last decades (from 72% in 1987 to 95% in this report), and it is likely that further improvements in the recording of cancer data by the regional cancer registries will result in complete data as similar to the Scandinavian countries (Jensen *et al*, 2002).

Sensitivity of self-reporting is 96.4% with significantly lower PPV (87.1%) when compared with that of cancer registry. The observed high sensitivity of self-reporting for BC diagnosis is comparable to that reported in the US studies by Parikh-Patel *et al* (2003; 98.1%) and Abraham *et al* (2009; 96.9%). A key finding is that almost 12% of women self-reporting BC were only found to have *in-situ* carcinoma of the breast and this was significantly

© 2012 Cancer Research UK

Table 3 Characteristics of the study women with known data as determinants of apparent sensitivity and PPV. The respective OR and significance levels were estimated

Variable	Apparent sensitivity	*N*	OR	95% CI	*P*-value	PPV	*N*	OR	95% CI	*P*-value
Ethnicity										
White	98.5	872	1			87.2	985	1		
Black		5					7			
Other		7	(Not measurable)				7	(Not measurable)		
Age groups (years)										
50–64			1			85.2	499	1		
65–80			0.7		0.6	89.3	505	1.6	1.1–2.5	0.02
Education										
Other	98.3	239	1			84.8	277	1		
Low	99.1	338	1.7	0.3–8.2	0.5	87.7	382	1.4	0.9–2.2	0.2
High	98.1	312	0.2	0.2–2.7	0.6	88.7	345	1.5	0.9–2.4	0.1
Breast cancer family history										
No	98.8	645	1			86.8	734	1		
Yes	98	244	0.6	0.2–2.5	0.4	88.6	270	1.6	0.8–1.9	0.3
Alcohol										
None	98.3	178	1			85.4	205	1		
1–2 Units per week	97.7	352	0.7	0.2–2.8	0.6	88.7	388	1.3	0.8–2.2	0.2
⩾3 Units per week	99.4	359	3.6	0.6–23.3	0.2	86.9	411	1.1	0.6–1.8	0.7
Smoking										
No	98.8	490	1			87.1	556	1		
Yes	98.2	399	0.8	0.2–2.0	0.4	87.5	448	1.1	0.8–1.6	0.6
Oral contraceptive pill										
No	98.6	351	1			87.1	394	1		
Yes	98.5	538	1.1	0.3–3.8	0.9	87.5	610	0.9	0.6–1.5	0.9
Sterilisation										
No	98.8	735	1			87	832	1		
Yes	98.2	154	0.6	0.2–4.0	0.8	88.4	172	1.2	0.7–2.1	0.4
Hysterectomy										
No	98.8	732	1			87	831	1		
Yes	97.5	157	0.5	0.1–1.7	0.2	88.5	173	1.2	0.7–1.9	0.6
HRT use										
No	98.8	635	1			86.9	722	1		
Yes	97.5	254	0.6	0.2–2.0	0.4	88.3	282	1.2	0.8–1.9	0.4
TOTAL	98.5					87.3				

more in women aged <65. On reviewing the literature, it is clear that although we have considered *in-situ* breast carcinoma, which was self-reported as BC under false positives (over-reporting), Abrahams *et al* (2009) in their analysis have included this group under true positives. It suggests that while a diagnosis of *in-situ* carcinoma may be a difficult concept for the women, researchers too are not in agreement on how it should be grouped.

A total of 3.6% of women did not report BC despite having a confirmed diagnosis (under-reporting). A variety of factors have been suggested to affect BC self-reporting, including age, sex, education and family history of BC (Bergmann *et al*, 1998; Abraham *et al*, 2009). The only factor in our analysis that was associated with self-reporting was age. Although age at completion of FUQ did not make any difference to under-reporting (apparent sensitivity), it was significantly correlated with over-reporting (PPV). Younger women were more likely to over-report *in-situ* breast carcinoma as BC compared with older respondents. Previous studies have shown age to affect both under- and over-reporting (Schrijvers *et al*, 1994; Bergmann *et al*, 1998; Parikh-Patel *et al*, 2003; Manjer *et al*, 2004; Dominguez *et al*, 2007). In addition, it has been suggested that women who are less educated are more likely to over-report cancer diagnosis (Schrijvers *et al*, 1994; Bergmann *et al*, 1998; Desai *et al*, 2001; Abraham *et al*, 2009) and those with a family history of BC are better responders when asked about their BC diagnosis (Abraham *et al*, 2009). We found no significant effect on reporting of either education or family history of BC. In studies that rely entirely on cancer data self-reported by participants, factors that have been shown to affect reporting may need to be taken into account.

Studies often use cancer registry or self-reported data. If only one source was to be used, then in England and Wales, cancer registry would be more complete with only 6.8% of women with BC missed/misclassified compared with 15.6% using self-reported data alone. In studies or trials where both sources are available, the lowest rate of misclassification of BC (2.7%) would be obtained by applying the rule that both sources need to concur for BC diagnosis. Nearly one in five women with BC notification would be missed/misclassified if the rule used was that BC is confirmed if reported on either source (cancer registry or self-reporting).

In conclusion, this study confirms that BC registration in England and Wales is highly reliable. Self-reporting, using postal questionnaires is another good source of cancer data. However, one in ten women self-reporting BC might only have *in-situ* breast carcinoma and women aged <65 are more likely to over report. It suggests the need for standardisation of information that patients receive.

© 2012 Cancer Research UK

ACKNOWLEDGEMENTS

We are grateful to all UKCTOCS trialists, staff, and especially participants who consented for use of their medical notes in secondary studies, and the physicians who completed the BCQ. The UKCTOCS was core-funded by MRC, Cancer Research UK and the Department of Health, with additional support from the Eve Appeal. This project was supported by a grant from the UCLH/UCL Comprehensive Biomedical Research Centre (project no 152), and most of this work has been undertaken at the UCLH/UCL, which received a proportion of its funding from the Department of Health NIHR Biomedical Research Centers funding scheme.

Conflict of interest

Ian Jacobs has consultancy arrangements with Becton Dickinson, who have an interest in tumour markers and ovarian cancer. Ian Jacobs and Usha Menon have a financial interest through UCL Business and Abcodia Ltd in the third party exploitation of clinical trials biobanks, which have been developed through the research at UCL. No other financial disclosures.

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

REFERENCES

Abraham L, Geller BM, Yankaskas BC, Bowles EJ, Karliner LS, Taplin SH, Miglioretti DL (2009) Accuracy of self-reported breast cancer among women undergoing mammography. *Breast Cancer Res Treat* **118**(3): 583–592

Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW (1998) Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries *Am J Epidemiol* **147**(6): 556–562

Brewster DH, Stockton D, Harvey J, Mackay M (2002) Reliability of cancer registration data in Scotland, 1997. *Eur J Cancer* **38**(3): 414–417

Brewster DH, Stockton DL (2008) Ascertainment of breast cancer by the Scottish Cancer Registry: an assessment based on comparison with five independent breast cancer trials databases. *Breast* **17**(1): 104–106

Desai MM, Bruce ML, Desai RA, Druss BG (2001) Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. *Am J Epidemiol* **153**(3): 299–306

Dominguez FJ, Lawrence C, Halpern EF, Drohan B, Grinstein G, Black DM, Smith BL, Gadd MA, Specht M, Kopans DB, Moore RH, Hughes SS, Roche CA, Hughes KS (2007) Accuracy of self-reported personal history of cancer in an outpatient breast center. *J Genet Couns* **16**(3): 341–345

Gathani T, Bull D, Green J, Reeves G, Beral V (2005) Breast cancer histological classification: agreement between the Office for National Statistics and the National Health Service Breast Screening Programme. *Breast Cancer Res* **7**(6): R1090–R1096

Gulliford MC, Bell J, Bourne HM, Petruckevitch A (1993) The reliability of Cancer Registry records. *Br J Cancer* **67**(4): 819–821

Jensen AR, Overgaard J, Storm HH (2002) Validity of breast cancer in the Danish Cancer Registry. A study based on clinical records from one county in Denmark. *Eur J Cancer Prev* **11**(4): 359–364

Lapham R, Waugh NR (1992) An audit of the quality of cancer registration data. *Br J Cancer* **66**(3): 552–554

Moller H, Richards S, Hanchett N, Riaz SP, Luchtenborg M, Holmberg L, Robinson D (2011) Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. *Br J Cancer* **105**(1): 170–176

Manjer J, Merlo J, Berglund G (2004) Validity of self-reported information on cancer: determinants of under- and over-reporting. *Eur J Epidemiol* **19**(3): 239–247

Menon U, Gentry-Maharaj A, Ryan A, Sharma A, Burnell M, Hallett R, Lewis S, Lopez A, Godfrey K, Oram D, Herod J, Williamson K, Seif M, Scott I, Mould T, Woolas R, Murdoch J, Dobbs S, Amso N, Leeson S, Cruickshank D, McGuire A, Campbell S, Fallowfield L, Skates S, Parmar M, Jacobs I (2008) Recruitment to multicentre trials–lessons from UKCTOCS: descriptive study. *Bmj* **337**: a2079

Paganini-Hill A, Chao A (1993) Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. *Am J Epidemiol* **138**(2): 101–106

Parikh-Patel A, Allen M, Wright WE (2003) Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol* **157**(6): 539–545

Phillips KA, Milne RL, Buys S, Friedlander ML, Ward JH, McCredie MR, Giles GG, Hopper JL (2005) Agreement between self-reported breast cancer treatment and medical records in a population-based Breast Cancer Family Registry. *J Clin Oncol* **23**(21): 4679–4686

Schootman MJD, West MM, Aft R (2005) Self-reported by elderly breast cancer patients was an acceptable altenative to surveillance, epidemiology, and end results (SEER) abstract data. *J Clin Epidemiol* **58**(12): 1316–1319

Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP (1994) Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol* **139**(4): 408–414

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

© 2012 Cancer Research UK