Journal of Public Health | Vol. 32, No. 4, pp. 565–571 | doi:10.1093/pubmed/fdq001 | Advance Access Publication 12 February 2010

Benign breast disease and subsequent breast cancer: English record linkage studies

M.J. Goldacre, J.D. Abisgold, D.G.R. Yeates, M.P. Vessey

Unit of Health-Care Epidemiology, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF, UK
Address correspondence to M.J. Goldacre, E-mail: michael.goldacre@dphpc.ox.ac.uk

ABSTRACT

Background Benign breast disease (BBD) increases the risk of breast cancer, but details of the relationship would benefit from further study in the UK.

Methods Analysis of linked statistical abstracts of hospital data, including a cohort of 20 976 women with BBD in an Oxford data set and 89 268 such women in an English national data set.

Results Rate ratios (RRs) for breast cancer, comparing BBD and comparison cohorts in these two data sets, were 2.3 (95% CI: 2.2–2.5) and 3.2 (3.0–3.3), respectively. RRs rose with increasing age at BBD diagnosis and remained elevated for at least 20 years after diagnosis. RRs were particularly high for a relatively small number of cancers occurring in the first few months after BBD diagnosis.

Conclusions Our findings accord well with those in other large studies, mostly done in the USA, in showing a sustained long-term cancer risk after BBD. They also demonstrate that known long-term risks of disease can be reliably identified from linked routine administrative hospital statistics. Most other studies omit cancers in the first few months after BBD. Such cases—presumably either misdiagnosed or miscoded—merit further study to determine whether in fact they include diagnoses of cancer that were initially missed.

Keywords benign breast disease, breast cancer, epidemiology, public health

Introduction

Benign breast disease (BBD) is common and well recognized to be a risk factor for subsequent breast cancer.[1–3] Important questions remain, however, about a number of aspects of the relationship in the UK.

The incidence of breast cancer varies substantially geographically.[4] For example, the age-standardized incidence per 100 000 women has varied from 25 in Beijing to 46 in Grenada in Spain, 85 in the Oxford region of England and 104 in non-Hispanic whites in Los Angeles.[4] Behavioural and environmental risk factors for breast cancer vary geographically; and it is possible that the contribution of BBD to risk varies between populations too.[5] The great majority of published studies on BBD and breast cancer risk have been undertaken in the USA.[2,6–8] The risk in English populations is less well-documented.

The possibility that the risk within a population may have changed over calendar time is also not well-documented. The duration of increased risk of cancer after a finding of BBD is uncertain,[2,6,7] as is whether the risk of breast cancer varies much according to women's age at diagnosis of BBD.[6,7,9,10] The increase in the identification of BBD in the past two decades in England as a result of the increased use of mammography (especially in screening) is a further reason why it is important to have breast cancer risk estimates in women diagnosed with BBD in England.

For these reasons, we have undertaken two record linkage studies to determine the risk of breast cancer after a diagnosis of BBD in English populations, over both short and long periods of follow-up, and comparing different time periods in

M.J. Goldacre, Professor of Public Health
J.D. Abisgold, Research Officer
D.G.R. Yeates, Computer Scientist
M.P. Vessey, Emeritus Professor of Public Health

which BBD was diagnosed. We used the long-standing Oxford record linkage study (ORLS)[11] and a more recent English national linked Hospital Episode Statistics (HES) data set.

Methods

Population and data

The ORLS includes brief statistical abstracts of records of all hospital admissions and day cases in National Health Service (NHS) hospitals, and all deaths regardless of where they occurred, in defined populations within the former Oxford NHS Region from 1963 to 1999. The ORLS continued from 1999, but the data before and after 1999 cannot be linked across the time periods. The original ORLS data from 1963 were collected routinely in the NHS as the region's hospital statistics system and were similar to English national HES data. Information about deaths derives from death certificates. The data for successive admissions and the death data for each person were linked together as they accrued. The English national linked HES file, similar in structure and content to the ORLS, spanned 1 April 1998–31 March 2005.

Using the ORLS, a cohort of patients with a diagnosis of BBD was constructed by identifying the first admission, or episode of day case care, in the study period. The study was confined to women aged 18 and over. BBD was defined as a hospital admission with the following International Classification of Disease (ICD) codes: ICD-7, 213 and 620; ICD-8, 217.0 and 610; ICD-9, 217 and 610; and ICD-10, D24 and N60. These codes encompass benign mammary dysplasia, fibroadenoma and a wide range of other less common benign conditions. Unfortunately, the ICD codes used in the data sets provide only site and not histology. A cohort for comparison, which we termed the 'reference cohort', was constructed by identifying the first admission for each individual with various medical and surgical conditions (see table footnotes). We followed the standard epidemiological practice, when hospital controls are used, of selecting a diverse range of conditions, rather than relying on a narrow range (in case, the latter are themselves atypical in their risk of subsequent breast cancer). This 'reference' group of conditions has been used in other studies of associations between non-malignant diseases and subsequent cancer,[11–13] and we have checked that individual conditions within the group do not give atypically high or low rates of cancer. We then searched the data set for any subsequent record of breast cancer in the Oxford data set. We considered that rates of breast cancer in the reference cohort would approximate those of healthy people in the general population of the region while allowing for migration into and out of it (data on migration of individuals were not available). We also analysed the ORLS data grouped into the time periods 1963–78, 1979–88 and from 1989. We did so to study any trends over time. The pre- and post-1989 periods define the periods before and after the introduction of the breast cancer screening programme. We split the pre-1989 era into two periods—1979–88 as a 10-year period immediately before screening (i.e. of similar length to the period after screening) and 1963–78 as the remainder.

The same methods and selection criteria were used to construct cohorts using the English national data. However, although the fact and date of death were available in this data set, the cause of death was not. Accordingly, the 'follow-up' for breast cancer was confined to the identification of hospital admissions for breast cancer.

Statistical methods

We calculated rates of breast cancer based on person-years at risk. We took 'date of entry' into each cohort as the date of first admission or day case care for BBD, or reference condition, and 'date of exit' as the date of first record of breast cancer, death or the end of the data file (31 March 1999 for the ORLS, 31 March 2005 for England), whichever was the earliest. In comparing the BBD cohort with the reference cohort, we first calculated rates for breast cancer, standardized using strata as described below, taking the combined BBD and reference cohorts as the standard population. We applied the stratum-specific rates in the combined cohort to the number of people in each stratum in the BBD cohort, separately, and then to those in the reference cohort. We calculated the ratio of the standardized rate of occurrence in the breast cancer cohort relative to that in the reference cohort. The confidence interval for the rate ratio (RR) and χ^2 statistics for its significance were calculated as described elsewhere.[14] We standardized the calculation of the rates by strata of age (in 5-year age groups), calendar year of first recorded admission (in single years), either district of residence (ORLS) or region of residence (England), and patients' deprivation score (England) grouped into quintiles based on their ward-of-residence score for the index of multiple deprivation (IMD). We used the IMD because it is provided in the English linked data set (but individual social class data are not). It is not available in the ORLS.

In comparing the BBD and reference cohorts, the precision of the RR depends on the number of people with each subsequent disease within each cohort. The size of the BBD cohort is fixed by the number in the data set with BBD.

Table 1 Age distribution of women admitted to hospital with BBD in Oxford and England: number and percentage of women in each age group at the time of admission

	Oxford			*England*		
	Number	*Percentage*	*MR*[a]	*Number*	*Percentage*	*MR*[a]
Age group (years)						
Under 25	2709	12.9	15	8627	9.7	17
25–29	2378	11.3	12	7002	7.8	18
30–34	2728	13	10	9370	10.5	14
35–39	3151	15	8	11 852	13.3	11
40–44	3550	16.9	7	12 396	13.9	10
45–49	3012	14.3	8	11 221	12.6	11
50–54	1539	7.3	14	10 947	12.3	12
55–59	667	3.2	23	6229	7	22
60–64	473	2.3	29	4369	4.8	31
65–69	348	1.7	42	2937	3.3	52
70–74	192	0.9	81	1984	2.2	87
75 +	229	1.2	168	2334	2.6	207
Total	20 976	100	14	89 268	100	22

[a]Matching ratio (MR) = the number of people in the reference cohort per person with BBD in each age stratum, see the 'Methods' section.

In the reference cohort, we included all the people in the data set in each stratum with the comparison conditions. We did this to maximize the numbers in each stratum in the reference cohort in order to maximize the statistical power of the study.

Results

Numbers and ages of subjects

There were 20 976 women with BBD in the ORLS cohort and 89 268 women with BBD in the English national cohort (Table 1). The mean ages at entry into the ORLS and English national BBD cohorts were 40.1 and 43.9 years, respectively. Of each total, 16.4% of the women in the ORLS cohort and 32.3% in the English national cohort were aged 50 years and over at the time of BBD diagnosis. The difference may reflect the influence of the national breast cancer screening programme, targeted at women aged 50 and over, in the period covering the English cohort. The mean period of follow-up was 10.7 years for the ORLS patients and 3.3 years nationally.

All ages and calendar time periods

Oxford record linkage study

Overall, there was a significantly high risk of breast cancer after BBD compared with the reference cohort, with a RR of 2.3 (95% CI: 2.2–2.5, Table 2). The RR for breast cancer within the first month of an admission for BBD was particularly high (18.4, 13.9–24.5). The RR was 4.4 (3.4–5.7) between 1 month and the end of the first year. It remained significantly high thereafter, throughout the study period, although it declined to 1.5 (1.2–2.0) at 20 years and more after BBD diagnosis.

Table 3 shows that there was an overall increase in the risk of breast cancer among women with BBD, relative to that in the reference cohort, during the period 1963 to 1978 (RR 2.2, 2.0–2.4). Corresponding RRs for the periods 1979–88 and 1989–99 were 2.1 (1.9–2.4) and 3.9 (3.2–4.7), respectively. The risk of breast cancer was significantly higher for women diagnosed with BBD after the introduction of routine screening, during 1988, than before it. The difference was only found for breast cancer diagnosed within the first year after BBD; excluding the first year, the risks before and after screening were very similar (Table 3).

Table 2 Occurrence of breast cancer in women aged 18 and over after benign breast disease in the ORLS area: time between admission for BBD and admission for breast cancer, number of women with breast cancer in the reference cohort, observed (Obs) number of women with cancer in the BBD cohort, expected (Exp) number of women with cancer in the BBD cohort, ratio of rate in BBD cohort to that in the reference cohort[a] and 95% confidence intervals for the RR

Time intervals	*Obs*	*Exp*	*Adjusted RR*[b] *(95% confidence interval)*
All (4442)	850	400.3	2.3 (2.2–2.5)
1st month (96)	113	12.5	18.4 (13.9–24.5)
1–11 months (308)	70	18.5	4.4 (3.4–5.7)
1–4 years (1172)	173	78.0	2.4 (2.0–2.8)
5–9 years (1102)	180	96.2	2.0 (1.7–2.4)
10–19 years (1303)	241	144.9	1.8 (1.6–2.0)
20+ years (461)	73	50.2	1.5 (1.2–2.0)

[a]Conditions used in the reference cohort with ICD9 code for diagnosis (with equivalent codes used for other coding editions): varicose veins (454), haemorrhoids (455), upper respiratory tract infections (460), deflected septum, nasal polyp (470–471), impacted tooth and disorders of teeth (520–521), inguinal hernia (550), in-growing toenail and other diseases of nail (703), sebaceous cyst (706.2), internal derangement of knee (717), bunion (727.1), squint (378), otitis externa and otitis media (380–382), selected fractures (810–816, 823–826), dislocations sprains and strains (830–839, 840–848) and superficial injury and contusion (910–919, 920–924).
[b]Adjusted for age in 5-year age bands, time period in single calendar years and district of residence; comparing Obs/Exp in the BBD cohort (data in table) with Obs/Exp in the reference cohort (data not shown separately).

Table 3 Breast cancer in Oxford after benign breast disease in year periods[a]: time between admission for BBD and admission for breast cancer, number of women with breast cancer in the reference cohort, observed (Obs) number of women with cancer in the BBD cohort, expected (Exp) number of women with cancer in the BBD cohort, ratio of rate in BBD cohort to that in the reference cohort and 95% confidence intervals for the RR

Year periods	*Time intervals*[b]	*Obs*	*Exp*	*Adjusted RR*[c]	*95% confidence interval*
1963–1978	All (2059)	422	211.4	2.2	2.0–2.4
1979–1988	All (1604)	296	151.5	2.1	1.9–2.4
1989–1999	All (779)	132	37.8	3.9	3.2–4.7
1963–1978	1st month	32	4.6	12.2	7.4–20.2
1979–1988	1st month	41	4.2	20.3	12.5–32.9
1989–1999	1st month	40	3.8	26.0	15.4–44.7
1963–1978	1–11 months	21	5.8	4.4	2.6–7.2
1979–1988	1–11 months	17	7.5	2.5	1.4–4.1
1989–1999	1–11 months	32	5.2	7.5	4.9–11.2
1963–1978	1–4 years	75	31.4	2.7	2.1–3.5
1979–1988	1–4 years	64	30.5	2.3	1.7–3.0
1989–1999	1–4 years	34	16.1	2.2	1.5–3.1
1963–1978	5–9 years	72	36.5	2.2	1.7–2.8
1979–1988	5–9 years	83	47.8	1.9	1.5–2.3
1989–1999	5–9 years	25	11.9	2.2	1.4–3.4
1963–1978	10–19 years	149	83.3	1.9	1.6–2.3
1979–1988	10–19 years	91	60.9	1.6	1.3–2.0
1963–1978	20+ years	73	49.9	1.5	1.2–2.0

[a]The number of women admitted with BBD from 1963 to 1978 was 8350, from 1979 to 1988 was 9988 and from 1989 to 1999 was 5134.

[b]Number of women with breast cancer in the reference cohort in parentheses.

[c]Adjusted for age in 5-year age bands, time period in single calendar years and district of residence; comparing Obs/Exp in the BBD cohort (data in table) with Obs/Exp in the reference cohort (data not shown separately).

Table 4 Occurrence of breast cancer in women aged 18 and over after benign breast disease in the English data set: time between admission for BBD and admission for breast cancer, number of women with breast cancer in the reference cohort, observed (Obs) and expected (Exp) number of women with cancer in the BBD cohort, ratio of rate of BBD cohort to that in the reference cohort and 95%confidence intervals for the RR

Time intervals[a]	*Obs*	*Exp*	*Adjusted RR*[b] *(95% confidence interval)*
All (17 338)	1812	612.5	3.2 (3.0–3.3)
1st mo. (1355)	299	44.1	7.7 (6.8–8.7)
1–11 mo. (4117)	612	151.6	4.5 (4.1–4.9)
1–4 y (10 615)	811	369.1	2.3 (2.1–2.5)
5–6 y (1251)	90	47.8	2.0 (1.6–2.4)

[a]Number of women with breast cancer in the reference cohort in parentheses.

[b]Adjusted for age in 5-year age bands, time period in single calendar years and district of residence; comparing Obs/Exp in the BBD cohort (data in table) with Obs/Exp in the reference cohort (data not shown separately).

English record linkage study

There was a significantly high RR for breast cancer in women with BBD relative to rates in the reference cohort (RR 3.2, 3.0–3.3, Table 4). The RR was particularly high in the first month and first year after BBD diagnosis (7.7, 6.8–8.7 and 4.5, 4.1–4.9, respectively) but an elevated RR was maintained over the full period of follow-up.

Individual age groups

In the Oxford data, the significant elevation of risk of cancer following BBD diagnosis compared with the reference cohort was found in each age group. Its magnitude generally increased with increasing age (Table S1). The RR was 2.2 (2.0–2.4) in women aged 18–49 years and it rose to 4.2 (2.9–5.9) in women aged 70 and over. The RR in the first month after discharge with BBD was particularly high in women under 50 years of age (54.2, 7.8–118.0, based on 71 observed cases). Within each time interval from BBD to breast cancer, the risk of breast cancer reduced with increasing calendar time from BBD, but it nonetheless generally remained significantly high (Table S1). A broadly similar pattern of risk to that seen in the ORLS data was found in the English data in each age group up to 6 years of follow-up (Table S2), although the increase in RRs with increasing age was less apparent than that in the ORLS.

An approximate calculation of absolute risk can be made using the English data in Table S2. Assuming no sizeable migration out of England, the percentage of women who, following an admission for BBD, had a subsequent admission for breast cancer within 7 years was 1.14% (691 of 60 468) for women aged 18–49, 3.28% (564 of 17 176) aged 50–59, 3.82% (279 of 7306) aged 60–69, 4.93% (213 of 4318) aged 70 and over, and 1.96% overall (1747 of 89 268). This increased gradient with increasing age no doubt reflects the increase in the absolute risk of breast cancer with age in the population as a whole.

Discussion

Main findings of this study

There was a 2–3-fold elevation of the risk of breast cancer following hospital care for BBD relative to that in the reference cohort. Elevated risk persisted for at least 20 years after the initial admission for BBD, although the magnitude of the increase in risk steadily declined with the passage of time. A diagnosis of BBD must be regarded as marking a long-term and persistent increase in breast cancer risk.

We found, in both data sets, that the RRs for breast cancer increased with increasing age at diagnosis of BBD, although the effect in the English national data was small.

The RR for breast cancer following BBD, compared with the reference cohort, was higher after admission for BBD in the period after the introduction of breast cancer screening than in the period before. The difference was only found in the first year after BBD. After the first year, the RRs were very similar in the periods before and after screening.

Our RRs for the first year after admission for BBD, especially the figures for the first month, are strikingly high, although the numbers of cases on which they are based are small (an average of 3 cases per year in the 36 years of the Oxford study and 43 per year in the whole of England).

What is already known on this topic

Our findings of an elevated RR persisting for at least 20 years after BBD diagnosis are in line with those of most other publications.[6,9,15–17] However, most large studies of the relationship between BBD and breast cancer have been done in the USA.

What this study adds

We show that findings in the UK, based on two very large data sets, do indeed lead to similar conclusions to the American studies about the level and long-term persistence of risk. Other large studies have generally excluded cancers occurring soon after a diagnosis of BBD (usually within the first 6 months) presumably on the assumption that many of them must represent misdiagnosis or misrecording of the original lesion as benign.[2,6,7,18]

There are a number of possible explanations for the early cluster of breast cancer cases. First, the coded BBD diagnosis might have been a miscode but the patient knew of the cancer diagnosis when she left hospital after the BBD episode. Second, the patient might have been discharged too soon for a diagnosis to be available; the hospital record or coding staff assumed the lesion was benign; but the woman was then correctly told of the cancer diagnosis. Third, misdiagnosis might have occurred, the patient being reassured initially but wrongly so. Finally, there might have been a rapid change from a benign to a malignant lesion; but this seems highly unlikely, given the very short intervals.

We cannot go back to original medical records to seek further information about the circumstances of the cases at short intervals because of privacy regulations. However, it is important to know the reasons for the high RRs at short-time intervals if only to provide reassurance that they did not represent misdiagnosis and false reassurance, if that is the case, in the context of any medico-legal challenge. Short-interval cases deserve further study in a research environment where privacy regulations do not preclude access to case notes, clinicians and patients.

Our data on risks in different calendar time periods of BBD diagnosis (Table 3) are, as far as we know, novel. The principal finding is that there has been no major change over time in risk. However, within this overall finding, we note the increase in risk, confined to the first year after BBD, in the period following the introduction of the breast cancer screening programme. This finding is likely to reflect, in some way, both the increasing frequency of breast biopsy consequent on mammographic screening and changing approaches to methods of undertaking biopsy in recent years. Most biopsies are now undertaken in the outpatient setting rather than as a day case or inpatient. Patients in the latter two categories are thus likely to be those in whom outpatient biopsy has been unsuccessful or uninformative and may thus be atypical of the group as a whole. However, an implication for public health is that, by identifying women with BBD through screening, some women are identified who, though reassured that they do not have cancer at the time of screening, nonetheless do have a long-term higher-than-average risk of breast cancer.

Our finding that the RRs for breast cancer increased with increasing age at diagnosis contrasts with the results reported in an important paper from the USA[6] that showed somewhat higher RRs for breast cancer following a diagnosis of BBD in younger women. In the US study, women aged 30–39 years had a RR of 1.85 (1.45–2.34), whereas in those aged 70 or more the corresponding figure was 1.40 (1.08–1.78). Other studies have produced variable results in this respect: some found an increase in RRs with age[10,16,18] and others a decrease.[7,9,17,19,20] Of necessity, we included all types of BBD in our cohorts, whereas some other studies have focussed on benign mammary dysplasia. Fibroadenoma, for example, an important type of BBD, makes up a larger proportion of such disease in younger women than in older women.[7,16,21] Since most studies have found fibroadenoma to be associated with a lesser increase in risk than benign mammary dysplasia,[3,18,21] our RRs in

younger women may have been diluted to some extent by the inclusion of this and other lower risk conditions. It is worth noting that the first large multiethnic cohort study of BBD and breast cancer in the USA[20] has shown a similar prevalence of the major histological subtypes of BBD in African-American to that in other American women and that the breast cancer risks associated with the different subtypes were also similar comparing ethnic groups.

Limitations of this study

An important limitation is the absence of clinical and histological information about the different types of BBD. This is partly a consequence of the inadequacy of the ICD coding system for BBD, used in HES, which needs to be improved. Most other studies of BBD and breast cancer reported in recent years have had histological information available and important differences have been found in the subsequent risk of breast cancer according to the nature of the benign pathology; in particular, the presence of proliferative disease with atypia has been associated with high cancer risk.[2,6,8,19,21] Unfortunately, as noted above, we could not seek further data such as histology from original case notes because of privacy regulations. There are also complexities of practice in pathology and screening that probably could not be resolved even with access to case notes. These include variation over time in case definition for pathologically defined breast disease, variation between pathologists in their use of diagnostic criteria, the fact that some cancers deemed to be malignant may not progress to invasive disease and ascertainment biases associated with the introduction of screening. We cannot address these issues. However, the study provides a perspective on health care received by the women themselves: for example, the English data show that ~2% of women given a diagnosis of BBD are readmitted, with a diagnosis of breast cancer, within 7 years; and it shows that admission for breast cancer is two to three times more common for such women than that for women without a previous hospital admission for BBD.

Other limitations of record linkage studies using routinely collected administrative data, such as ours are well known, and include the facts that the data are limited to hospitalized patients (day cases are nonetheless included) and that information about some variables of potential interest, such as family history,[2,6,7,15] individual social class and ethnicity are generally unavailable. As noted above, the risk of breast cancer after BBD was not related to ethnicity in a large multi-ethnic study in the USA.[20] The similarities between our study and others, in respect of the size and duration of risk found, suggest that unmeasured confounding and bias—e.g. by ethnicity or social class, variation in the use of diagnostic criteria or the interpretation of histopathology—are unlikely to have had major effects on our findings.

Migration of subjects into and out of the area covered by Oxford record linkage prevents the calculation of absolute event rates or events per woman-years at risk. However, migration is likely to be small, over a fairly short period of time covered by English national linked data, and we show estimates of absolute risk.

A strength of using the Oxford and England linked data sets is that they cover two very large populations of women with BBD. The ORLS data provide long duration of follow-up; the English data provide a much larger and more recent population but with much shorter follow-up. The two data sets are independent of each other and it is reassuring that they give very similar results to one another and to other published studies.

An important methodological point is that we demonstrate clearly that the linkage of routinely collected administrative hospital data can identify known associations between two clinical conditions—in this case, BBD and breast cancer—at a level of risk that accords with the literature from resource-intensive, patient-based, long-term clinical follow-up. It would be very daunting to undertake a prospective cohort study, using personal tracking of individual patients, on the scale of our study over a 20+-year period of follow-up.

In conclusion, our data confirm that BBD is an important risk factor for breast cancer in England which persists for at least 20 years after BBD diagnosis. Further work is required to provide a firm explanation for the apparently very high risk of breast cancer found during the first few months after BBD diagnosis.

Ethical approval

The English NHS Central Office for Research Ethics Committees approved the current work programme of analysis using the linked data sets (reference number 04/Q2006/176).

Supplementary data

Supplementary data are available at the *Journal of Public Health* online.

Acknowledgements

The English NIHR Co-ordinating Centre for Research Capacity Development funds the Unit of Health-Care

Epidemiology to undertake research using the Oxford record linkage study and the English national linked data set. The views in this paper do not necessarily reflect those of the funding body.

Funding

English NIHR Co-ordinating Centre for Research Capacity Development.

References

1 Connolly JL, Schnitt SJ. Benign breast disease: resolved and unresolved issues. *Cancer* 1993;**71**:1187–9.

2 Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med* 1985;**312**:146–51.

3 Fitzgibbons PL, Henson DE, Hutter RV. Benign breast changes and the risk for subsequent breast cancer: an update of the 1985 consensus statement. Cancer Committee of the College of American Pathologists. *Arch Pathol Lab Med* 1998;**122**:1053–5.

4 Tomatis L, Aitio A, Day NE *et al. Cancer: Causes, Occurrence and Control. IARC Scientific Publications No. 100*. Lyon: International Agency for Research in Cancer, 1990,69–72.

5 Hartmann LC, Ghosh K. Benign breast disease: emerging findings in a diverse population. *Breast J* 2007;**13**:113–4.

6 Hartmann LC, Sellers TA, Frost MH *et al*. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;**353**:229–37.

7 Carter CL, Corle DK, Micozzi MS *et al*. A prospective study of the development of breast cancer in 16,692 women with benign breast disease. *Am J Epidemiol* 1988;**128**:467–77.

8 Schnitt SJ. Benign breast disease and breast cancer risk: potential role for antiestrogens. *Clin Cancer Res* 2001;**7(Suppl.)**:4419s–22s.

9 London SJ, Connolly JL, Schnitt SJ *et al*. A prospective study of benign breast disease and the risk of breast cancer. *JAMA* 1992;**267**:941–4.

10 Wang J, Costantino JP, Tan-Chiu E *et al*. Lower-category benign breast disease and the risk of invasive breast cancer. *J Natl Cancer Inst* 2004;**96**:616–20.

11 Goldacre M, Kurina L, Yeates D *et al*. Use of large medical databases to study associations between diseases. *QJM* 2000;**93(10)**:669–75.

12 Goldacre MJ, Wotton CJ, Seagroatt V *et al*. Cancer following hip and knee arthroplasy: record linkage study. *Br J Cancer* 2005;**92(7)**:1298–301.

13 Goldacre MJ, Abisgold JD, Seagroatt V *et al*. Cancer after cholecystectomy: record-linkage cohort study. *Br J Cancer* 2005;**92(7)**:1307–9.

14 Breslow NE, Day NE. *Statistical Methods in Cancer Research. Volume II—The Design and Analysis of Cohort Studies. IARC Scientific Publications*. Lyon: International Agency for Research in Cancer, 1987;1–406.

15 Dupont WD, Page Dl, Parl FF *et al*. Long-term risk of breast cancer in women with fibroadenoma. *N Engl J Med* 1994;**331**:10–5.

16 Krieger N, Hiatt RA. Risk of breast cancer after benign breast diseases. Variation by histologic type, degree of atypia, age at biopsy and length of follow-up. *Am J Epidemiol* 1992;**135**:619–31.

17 Degnim AC, Visscher DW, Berman HK *et al*. Stratification of breast cancer risk in women with atypia: a Mayo Clinic study. *J Clin Oncol* 2007;**25**:2671–7.

18 Ashbeck EL, Rosenberg RD, Stauber PM *et al*. Benign breast biopsy diagnosis and subsequent risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:467–72.

19 Marshall LM, Hunter DJ, Connolly JL *et al*. Risk of breast cancer associated with atypical hyperplasia of lobular and ductal types. *Cancer Epidemiol Biomarkers Prev* 1997;**6**:297–301.

20 Worsham MJ, Abrams, Raju U *et al*. Breast cancer incidence in a cohort of women with benign breast disease from a multiethnic, primary health care population. *Breast J* 2007;**13**:115–21.

21 Santen RJ, Mansel R. Benign breast disorders. *N Engl J Med* 2005;**353**:275–85.