

## Review

### Overdiagnosis and overtreatment of breast cancer

# Estimates of overdiagnosis from two trials of mammographic screening for breast cancer

Stephen W Duffy<sup>1</sup>, Olorunsola Agbaje<sup>1</sup>, Laszlo Tabar<sup>2</sup>, Bedrich Vitak<sup>3</sup>, Nils Bjurstam<sup>4</sup>, Lena Björneld<sup>5</sup>, Jonathan P Myles<sup>1</sup> and Jane Warwick<sup>1</sup>

<sup>1</sup>Cancer Research UK Department of Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, Queen Mary College, University of London, London, UK

<sup>2</sup>Department of Mammography, Falun Central Hospital, Sweden

<sup>3</sup>Department of Radiology, University Hospital, Linköping, Sweden

<sup>4</sup>Department of Radiology, Center for Breast Imaging, University Hospital of North Norway, Tromsø, Norway

<sup>5</sup>Department of Radiology, Sahlgrenska University Hospital, Göteborg, Sweden

Corresponding author: Stephen W Duffy, [stephen.duffy@cancer.org.uk](mailto:stephen.duffy@cancer.org.uk)

Published: 10 November 2005

This article is online at <http://breast-cancer-research.com/content/7/6/258>

© 2005 BioMed Central Ltd

*Breast Cancer Research* 2005, **7**:258-265 (DOI 10.1186/bcr1354)

## Abstract

Randomised controlled trials have shown that the policy of mammographic screening confers a substantial and significant reduction in breast cancer mortality. This has often been accompanied, however, by an increase in breast cancer incidence, particularly during the early years of a screening programme, which has led to concerns about overdiagnosis, that is to say, the diagnosis of disease that, if left undetected and therefore untreated, would not become symptomatic. We used incidence data from two randomised controlled trials of mammographic screening, the Swedish Two-county Trial and the Gothenburg Trial, to establish the timing and magnitude of any excess incidence of invasive disease and ductal carcinoma *in situ* (DCIS) in the study groups, to ascertain whether the excess incidence of DCIS reported early in a screening trial is balanced by a later deficit in invasive disease and provide explicit estimates of the rate of 'real' and non-progressive 'overdiagnosed' tumours from the study groups of the trials. We used a multistate model for overdiagnosis and used Markov Chain Monte Carlo methods to estimate the parameters. After taking into account the effect of lead time, we estimated that less than 5% of cases diagnosed at prevalence screen and less than 1% of cases diagnosed at incidence screens are being overdiagnosed. Overall, we estimate overdiagnosis to be around 1% of all cases diagnosed in screened populations. These estimates are, however, subject to considerable uncertainty. Our results suggest that overdiagnosis in mammography screening is a minor phenomenon, but further studies with very large numbers are required for more precise estimation.

## Introduction

Randomised controlled trials have shown that the policy of mammographic screening confers a substantial and significant reduction in breast cancer mortality [1-3]. There is

continuing interest in the human costs associated with the mortality benefit, in particular, whether overdiagnosis occurs in breast cancer screening and, if so, its magnitude [4,5]. In this context, overdiagnosis means the diagnosis of cancer as a result of screening, usually histologically confirmed, that would not have arisen clinically during the lifetime of the host had screening not taken place.

When a mammographic screening programme is initiated, usually a large increase in breast cancer incidence is observed in the early years of the programme, and a relatively small increase later [4,6]. This in itself is not sufficient to imply overdiagnosis, for the following reasons:

1. In most parts of the world, breast cancer incidence was increasing prior to the epoch of mammography. Thus at least part of any excess incidence observed in the screening epoch is probably due to an existing increasing trend in incidence.
2. In addition, the early diagnosis of cancers due to lead time may exacerbate the underlying temporal increase by bringing forward in time future higher rates of disease.
3. In relation to this, screening also causes an artificial increase in age-specific incidence. With two years lead time on average, we would observe age 52 incidence at age 50, and so on.
4. There will be a substantial excess in incidence in the first few years of the programme due to the prevalence screen: large numbers of asymptomatic tumours in the prevalence pool will have their diagnosis date brought forward to the time of the prevalence screen.

5. There will be a continuing excess thereafter at the lower end of the age range for screening, due to prevalence screens of subjects reaching the age for screening eligibility.

That said, the increase could still be partly due to overdiagnosis.

One would expect the excess incidence due to lead time to be followed by a deficit in incidence in screened cohorts at ages higher than the upper age limit for screening, as was observed in the UK [6]. Estimation from the deficit, however, is not straightforward, because usually one can identify screened cohorts only at aggregate rather than individual level, and it takes some years after screening before the subsequent deficit becomes observable.

An issue of particular interest is overdiagnosis of ductal carcinoma *in situ* (DCIS) [7]. Here, the question of most interest is: how much of the DCIS diagnosed at screening would be expected to progress to invasive cancer if left untreated? The DCIS that would have progressed represents invasive cancers prevented, a major benefit of screening. Those that would not have progressed represent overdiagnosis and unnecessary treatment.

Essential to the concept and existence of overdiagnosis is the duration of the preclinical screen-detectable period, the sojourn time. Overdiagnosis can be thought of as a combination of two disease entities. The first is the diagnosis of a potentially progressive cancer in a subject who is going to die of other causes in the near future in any case, possibly from an accident, another occult disease or an unexpected cerebrovascular or cardiovascular event, before the tumour would have given rise to clinical symptoms. The second is an extreme form of length bias whereby there are, in theory, subclinical tumours with little or no potential to progress to symptomatic disease, that is, whose sojourn time has a radically different distribution from that of the general tumour population.

The first of these must undoubtedly happen, but given the low all-cause mortality rates of women in the age groups invited for screening, and the likely mean and distribution of sojourn time, this type of overdiagnosis is liable to be very rare [4]. It would, therefore, seem more potentially productive in terms of estimation to focus on the latter form of overdiagnosis, a subpopulation of non-progressive or low-progression tumours.

In this paper, we use two randomised controlled trials of mammographic screening, the Swedish Two-county Trial and the Gothenburg Trial, to address the following issues: the timing and magnitude of excess incidence of invasive disease and DCIS in the study groups compared to the control groups; whether there is evidence that the excess incidence

of DCIS is balanced by a later deficit in invasive disease; and explicit estimation of rates of 'real' tumours and non-progressive 'overdiagnosed' tumours from the study groups of the trials.

## Methods

The design features of the two trials have been described in detail elsewhere [1,8]. Briefly, in the Swedish Two-county Trial, 77,080 women aged 40 to 74 years were randomised to regular invitation to screening, and 55,985 to no invitation. Screening was by single-view mammography, with an interscreening interval of 2 years in women aged 40 to 49 years and 33 months in women aged 50 to 74 years at randomisation. The trial began in late 1977. Around 7 years later, after approximately 3 rounds of screening in the older group and 4 rounds of screening in the younger, a mortality reduction of 30% was observed and published [9], the control group invited to screening and the screening phase of the trial closed. Follow-up was continued for mortality from the tumours diagnosed during the screening phase [1].

In the Gothenburg Trial, 21,650 women aged 39 to 59 years were randomised to invitation to screening and 29,961 to no invitation [8]. The screening was by two-view mammography at first screen, with number of views thereafter dependent on breast density. Screening took place at 18 month intervals. The trial began in 1982. After five rounds of screening in the 1933 to 1944 birth cohorts (approximately the 39 to 49 year age group at randomisation), the corresponding control group members were offered screening and the screening phase of the trial closed. In the 1923 to 1932 birth cohorts (the 50 to 59 year age group), the control group was invited to screening after four rounds. As in the Swedish Two-county Trial, follow-up has continued for mortality from the tumours diagnosed during the screening phase of the trial.

In both trials, the control group was offered screening at the close of the screening phase, so we cannot estimate overdiagnosis by a simple comparison of long term incidence rates in the study and control groups. We can, however, study the size and timing of excess incidence during the screening phase to obtain clues to when overdiagnosis may occur. Accordingly, our first analysis was to estimate cumulative incidence rates of invasive, *in situ* and total cancers in the study and control groups of each trial. It has already been noted that in both trials incidence equalised between study and control groups with the first screen of the control group, suggesting that if there is overdiagnosis, it occurs mainly at the first screen [2,8].

In the Gothenburg Trial, each individual year of birth cohort (from 1923 to 1944) was randomised in succession, with a study to control ratio chosen on the basis of the capacity of the mammography facilities to screen the study group [8]. The variation of the randomisation ratio by year of birth induced an age imbalance (albeit a very small imbalance)

between study and control groups. To take account of this, the Gothenburg study group incidence is compared not with the raw control group incidence but with the standardised incidence that would have been observed in the control group if it had had exactly the same year of birth distribution as the study group [8].

Our second analysis involved explicit estimation of the incidence of 'real' and 'overdiagnosed' cases from the numbers of cases detected at screening and between screens in the two trials. We assumed a uniform annual incidence  $I$  of preclinical but screen detectable, truly progressive cancers, an exponential distribution of time from inception of these to clinical symptoms with rate  $\lambda$ , and a screening test sensitivity  $S$ . In addition, we assume exponential incidence of overdiagnosed (non-progressive) preclinical screen-detectable cancers, with rate  $\mu$ . Because a tumour is only overdiagnosed if it is actually detected at screening, we define the screening test sensitivity to be 100% for overdiagnosed cancers. In this model, there are four states: no detectable disease, non-progressive (overdiagnosed) preclinical disease, progressive preclinical disease, and clinical symptomatic disease. The expected rates of cancers diagnosed at first, second and third screens, and in the intervals following those screens with an average interval time of  $t$  are as follows.

First screen:

$$\frac{SI}{\lambda} + (1 - e^{-\mu a})$$

where  $a$  is average age (50 years in the Gothenburg Trial and 58 years in the Swedish Two-county Trial). The second component in the expected rate represents the overdiagnosed cancers.

This allows a constant incidence rate of non-progressive disease from birth to age at first screen. This is arbitrary, biologically unverifiable and it may be wrong. However, the expected rates predicted for any multiplier of  $\mu$  from 15 or 20 years upwards are very similar, and it seemed to us less arbitrary to allow the subjects' age to dictate our time limit than to choose one ourselves, given the current low level of knowledge of non-progressive disease.

Between first and second screen:

$$\frac{I}{\lambda} \{-S(1 - e^{-\lambda t}) + \lambda t\}$$

As these are symptomatic tumours there is no term for overdiagnosis.

Second screen:

$$\frac{SI}{\lambda} (1 - Se^{-\lambda t}) + (1 - e^{-\mu t})$$

The second component in the expected rate represents the overdiagnosed cancers.

Between second and third screen:

$$\frac{I}{\lambda} \{(1 - S) (1 - Se^{-\lambda t}) (1 - e^{-\lambda t}) + \lambda t - (1 - e^{-\lambda t})\}$$

As these are symptomatic tumours there is no term for overdiagnosis.

Third screen:

$$\frac{SI}{\lambda} \{(1 - S) (1 - Se^{-\lambda t}) e^{-\lambda t} + (1 - e^{-\lambda t})\} + (1 - e^{-\mu t})$$

The second component in the expected rate represents the overdiagnosed cancers.

Interval after third screen:

$$\frac{(1 - S)I}{\lambda} \{(1 - S) (1 - Se^{-\lambda t}) e^{-\lambda t} + (1 - e^{-\lambda t})\} + It - \frac{I(1 - e^{-\lambda t})}{\lambda}$$

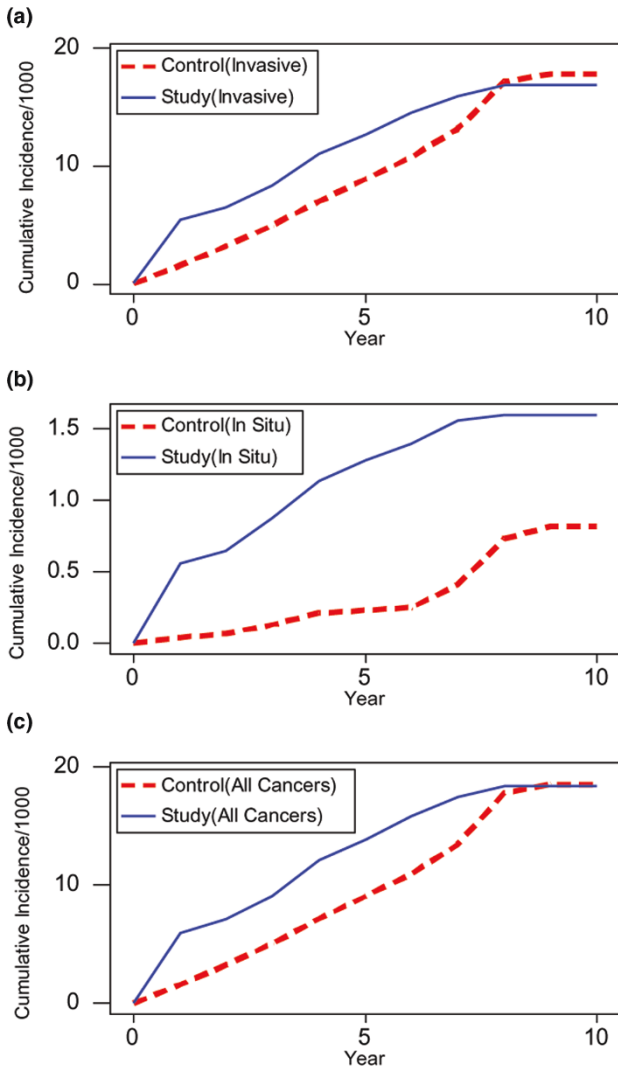
Since these are symptomatic tumours there is no term for overdiagnosis.

From the data on screen-detected and interval cancers, we estimated  $I$ ,  $\lambda$ ,  $S$  and  $\mu$  by fitting Poisson distributions to the numbers of cases at the three screens and in the three intervals with expectations as above. For the Swedish Two-county Trial,  $t = 2.56$  years (the average interval for the 19,844 women younger than 50 years and the 57,236 women aged 50 to 74 years). For the Gothenburg Trial,  $t = 1.5$  years. The estimation algorithm used was Markov Chain Monte Carlo (MCMC), implemented in the computer programme WinBUGS [10]. The diagnostic criteria of Geweke, Raftery and Lewis, and Heldelberger and Elch in Convergence Diagnostics and Output Analysis Software (CODA) were used to assess convergence of the MCMC parameters [11]. The results for the chain provided no evidence against convergence for all the parameters. We intentionally chose uninformative prior distributions to approximate a maximum likelihood solution. Results are presented as mean posterior distribution values and 95% credible intervals. The WinBUGS program updated a single chain with 15,000 samples (with thinning of 1), from which the first 5,000 samples were discarded (burn-ins) and the remaining 10,000 samples were used in estimation. Prior distributions used for the parameters  $I$ ,  $\lambda$ ,  $S$  and  $\mu$  were as follows:  $I$ , lognormal(0.0, 0.0001);  $\lambda$ , gamma(0.01, 0.01);  $S$ , logit( $S$ ) =  $\alpha$ ,  $\alpha \sim$  normal(0.0, 0.0001);  $\mu$ , lognormal(0.0, 0.01). Note that the second parameter in the normal and lognormal distributions is the precision, and not the variance or the standard deviation [10].

## Results

Figure 1a-c shows the cumulative incidence of invasive breast cancer, DCIS, and all breast cancers in the study and

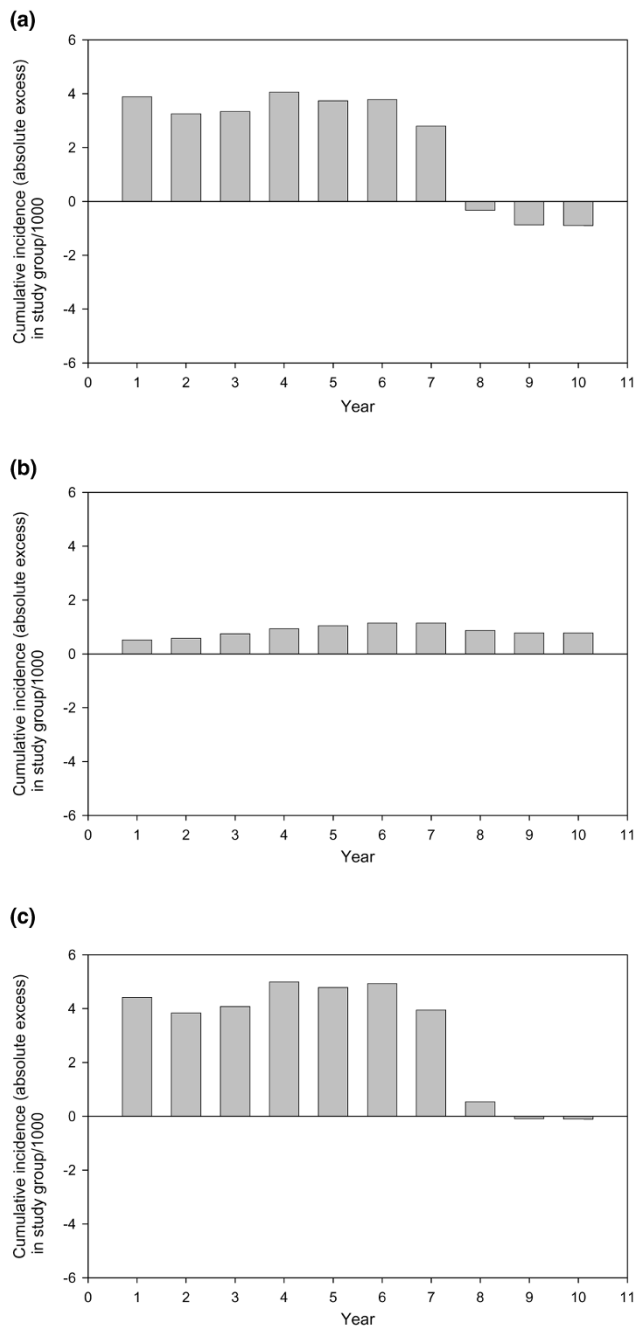
**Figure 1**



Cumulative incidence of breast cancers in study and control groups of the Swedish Two-county Trial. **(a)** Invasive cancers. **(b)** *In situ* cancers. **(c)** All cancers.

control groups of the Swedish Two-county Trial. Figure 2a-c shows the corresponding absolute excesses/deficits in the study group over time, per thousand women randomised. As noted above, the overall rates equalised at years 8 to 9, once the first screen of the control group was complete. The study group excess in DCIS rates peaked at 6 to 7 years and was balanced by a deficit in invasive tumours at 8 to 9 years, with the screening of the control group. The absolute excess of DCIS cases in the study group was 60 tumours, and the deficit of invasive tumours was 68, suggesting no overdiagnosis at all. If, conservatively, we exclude DCIS cases diagnosed at the first screen of the control group, there was an excess of 86 DCIS cases in the study group, suggesting a total overdiagnosis of 18 DCIS cases. This amounts to 15%

**Figure 2**

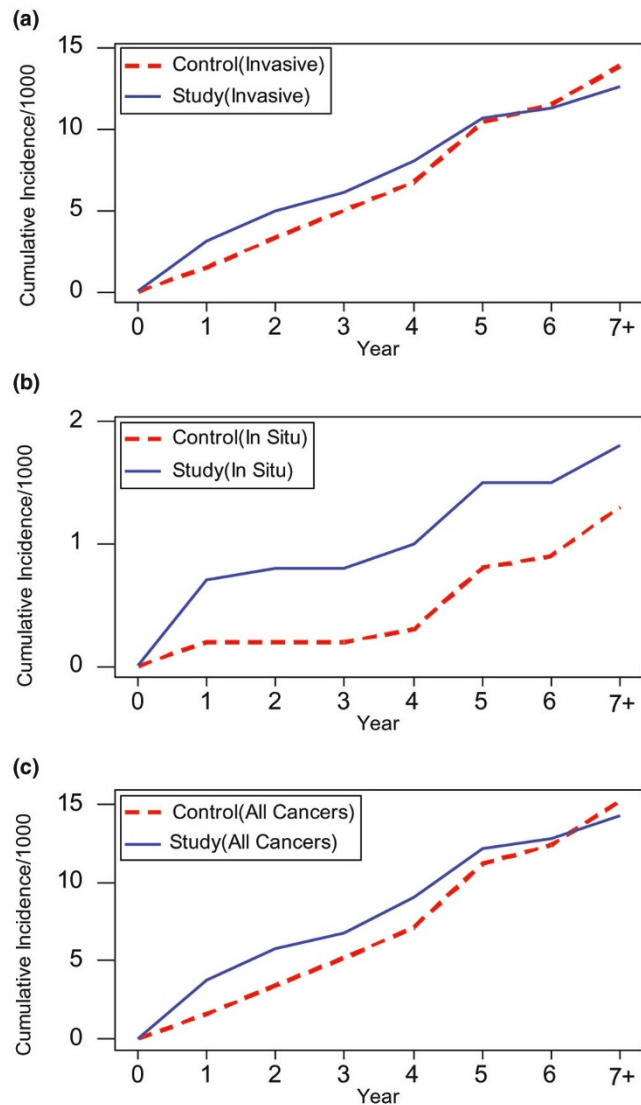


Cumulative excess incidence (study versus control) of breast cancers in the Swedish Two-county Trial. **(a)** Invasive cancers. **(b)** *In situ* cancers. **(c)** All cancers.

of all DCIS cases and 1% of all tumours. This can be regarded as an upper limit on the amount of overdiagnosis of DCIS in the trial.

Figure 3a-c shows the corresponding cumulative incidences in the Gothenburg Trial.

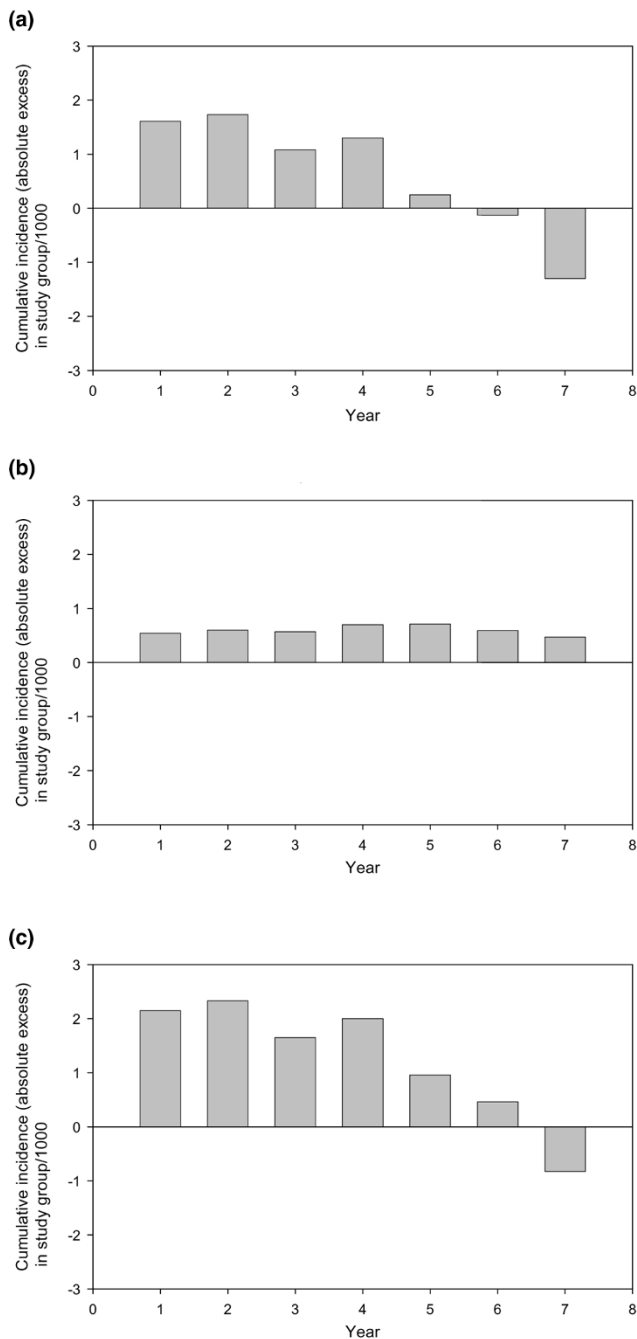
Figure 3



Cumulative incidence of breast cancers in study and control groups of the Gothenburg Trial. (a) Invasive cancers. (b) *In situ* cancers. (c) All cancers.

There was a substantial proportional excess, but very small absolute excess of *in situ* cancers, which was again balanced by a deficit in invasive cancers (Fig. 4). The excess of *in situ* cancers peaked at 4 to 5 years. Overall rates equalised at 6 to 7 years, around the time of screening the control group. The absolute excess of DCIS cases was 10, and the deficit of invasive cases was 28, again suggesting no overdiagnosis of DCIS. After exclusion of DCIS cases diagnosed at the first screen of the control group, the excess in the study group was 35, and the overall balance of all tumour types therefore suggested 7 overdiagnosed cases, 18% of DCIS and 2% of all study group cancers, a likely upper limit on overdiagnosis of DCIS in this study.

Figure 4



Cumulative excess incidence (study versus control) of breast cancers in the Gothenburg Trial. (a) Invasive cancers. (b) *In situ* cancers. (c) All cancers.

Table 1 shows the numbers screened and cancers detected at the first three screens and in the interval after each of the first three screens in the study group of the Swedish Two-county Trial. Applying the overdiagnosis model to these data gives the results in Table 2. These results pertain to all

**Table 1****Cancers diagnosed at and after the first three screens, Swedish Two-County Trial**

Detection occasion	Number screened	Invasive cancers only	All cancers
First screen	68,770	384	426
First interval	68,770	123	134
Second screen	58,601	214	244
Second interval	58,601	78	82
Third screen	43,320	173	193
Third interval	43,320	89	91

**Table 2****Estimates from formal overdiagnosis modelling, Swedish Two-county Trial data, ages 40 to 74 years**

Quantity	Estimate	95% CI
Incidence (true cases)/1000 ( <i>I</i> )	2.2	(2.0-2.4)
Screening test sensitivity ( <i>S</i> ) as %	99.9	(99.6-100)
Progression to clinical cancer ( $\lambda$ )	0.34	(0.31-0.38)
Incidence of overdiagnosed cases/1000 ( $\mu$ )	0.0038	(0.0001-0.0011)
Percent overdiagnosis (first screen)	3.1	(0.1-10.9)
Percent overdiagnosis (second screen)	0.3	(0.1-1.0)
Percent overdiagnosis (third screen)	0.3	(0.1-1.0)

CI, confidence interval.

cancers, invasive and *in situ*, but it should be noted that very similar results were obtained using invasive cancers only. Results indicate percentages of tumours overdiagnosed of 3.1%, 0.3% and 0.3% at the first, second and third screens, respectively. This implies a total of 14 tumours overdiagnosed, 1% of all tumours, screen-detected and clinical, arising during the period of observation. We also re-estimated the parameters restricting the data to the 40 to 69 year age group, as the 70 to 74 year age group was only invited to the first two screens. Results were very similar, giving overdiagnosis rates of 3%, 0.2% and 0.2% at the first three screens, and an overall percentage overdiagnosed of 1% of all tumours diagnosed in the programme.

Table 3 shows the corresponding data for the Gothenburg Trial, and Table 4 the results of overdiagnosis modelling from the Gothenburg data. Results show 4.2% overdiagnosis at first screen and 0.3% at subsequent screens. This corresponds to three cancers diagnosed, two percent of all tumours

**Table 3****Cancers diagnosed at and after the first three screens, Gothenburg Trial**

Detection occasion	Number screened	Invasive cancers only	All cancers
First screen	18,197	55	70
First interval	18,197	7	9
Second screen	17,005	27	28
Second interval	17,005	14	14
Third screen	17,093	36	41
Third interval	17,093	21	23

**Table 4****Estimates from formal overdiagnosis modelling, Gothenburg Trial data, ages 39 to 59 years**

Quantity	Estimate	95% CI
Incidence (true cases)/1000 ( <i>I</i> )	1.9	(1.6-2.3)
Screening test sensitivity ( <i>S</i> )	99.9	(99.6-100)
Progression to clinical cancer ( $\lambda$ )	0.52	(0.41-0.67)
Incidence of overdiagnosed cases/1000 ( $\mu$ )	0.0034	(0.0000-0.0238)
Percent overdiagnosis (first screen)	4.2	(0.0-28.8)
Percent overdiagnosis (second screen)	0.3	(0.0-2.0)
Percent overdiagnosis (third screen)	0.3	(0.0-2.0)

CI, confidence interval.

diagnosed in the first three screening rounds. Restriction of the analysis to invasive tumours only reduces the overdiagnosis estimates by around one-third.

**Discussion**

We have derived formal estimates of overdiagnosis from empirical breast screening data. The estimates take into account the effect of lead time and use direct estimation of the underlying incidence of both 'true' and 'overdiagnosed' cases from the screened populations. We found overdiagnosis to be a minor phenomenon, with less than 5% of cases diagnosed at prevalence screen and less than 1% of cases at incidence screens being overdiagnosed. Overdiagnosis was estimated at around 1% of all cases diagnosed in the screened populations.

Examination of absolute incidence rates of DCIS and invasive disease suggest further that overdiagnosis of DCIS is not the major problem it is claimed to be [12]. While large relative

increases in DCIS rates have been cited as evidence for such overdiagnosis [12], absolute rates of detection of DCIS remain low, at around one per thousand screened [13]. Previous detailed estimation of DCIS progression is in agreement with our results [14].

Other estimates of overdiagnosis in the literature range from 5% or less [4] up to 30% [15]. The latter, however, does not formally take into account the lead time effect, and does not fully identify screened and unscreened cohorts. We would suggest that simple estimation of rates at an aggregate level, while useful, is not sufficient in itself to derive conclusive estimates of overdiagnosis rates.

Our estimates of incidence of preclinical disease in the two trials are similar to the clinical incidence rates in the respective control groups before their exit screen (2.1 per 1,000 and 1.8 per 1,000 for the Swedish Two-county and Gothenburg Trials, respectively). It should be noted that we have wide confidence intervals on our overdiagnosis estimates, and the estimate of screening test sensitivity tends to drift to its boundary at 100%. Also, there is some sensitivity to the prior distribution for  $\mu$ , the incidence rate of overdiagnosed cancers, uniform priors tending to give higher estimates of  $\mu$ . For more stable estimation, perhaps overview estimates from several screening programmes, as in Yen *et al.* [14], are indicated.

In both trials, our estimate of sensitivity drifted towards its upper bound of 100%. Two points should be noted here. Firstly, the part of the likelihood related to the prevalence screen is monotonic increasing in  $S$ , as are the parts related to incidence screens under most circumstances. The likelihood component related to the interval cancers is not, but if there are very few interval cancers, this can be outweighed by the likelihood pertaining to screen-detected cancers. This reflects the fact that a very high sensitivity is implied if there are very low interval cancer rates. Secondly, our sensitivity estimate is of test sensitivity, not program sensitivity, which includes all interval cancers as false negatives. Our estimate differs from that of others [16], largely because it takes account of the sojourn time in estimation of the proportion of interval cancers that are really newly arising since the screen, as opposed to those missed at the screen. As noted above, if the observed number of interval cancers is small, the estimate of  $S$  must be close to 100%. It should be noted that the maximum likelihood estimate of  $S$  would also be 100%.

The models we have fitted here are rather simple. Only a single overdiagnosis parameter is estimated. There is room for improvement, in terms of estimation of age-specific overdiagnosis rates, for example. Multiple overdiagnosis parameters, and the small numbers resulting when analysis is restricted to age subgroups, both give rise to instability of estimation. Solving this problem is a target of ongoing research.

This article is part of a review series on *Overdiagnosis and overtreatment of breast cancer*, edited by Nick E Day, Stephen Duffy and Eugenio Paci.

Other articles in the series can be found online at [http://breast-cancer-research.com/articles/review-series.asp?series=BCR\\_Overdiagnosis](http://breast-cancer-research.com/articles/review-series.asp?series=BCR_Overdiagnosis)

It would be of some interest to see estimates from formal models from other screening trials and service screening programmes. In the meantime, the results here suggest that overdiagnosis in mammography screening is a minor phenomenon. We need more data to reduce the uncertainty around these estimates.

### Competing interests

The author(s) declare that they have no competing interests.

### Authors' contributions

SWD conceived and designed the study, supervised the statistical analyses and drafted the manuscript. OA carried out statistical analyses. LT led the Swedish Two-county Trial and contributed to the manuscript. BV conducted the Ostergotland part of the Swedish Two-county trial and contributed to the manuscript. NB led the Gothenburg Trial and contributed to the manuscript. LB managed the Gothenburg Trial administration and contributed to the manuscript. JPM contributed to the statistical modelling. JW carried out statistical analyses and contributed to the manuscript.

### Acknowledgements

We thank the women who participated in the trials and the clinical and support staff who contributed time and expertise to them. We also thank Jayne Mead for secretarial support.

### References

1. Tabar L, Vitak B, Chen HH, Duffy SW, Smith RA: **The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long term follow-up.** *Radiol Clin Nth Am* 2000, **38**:625-651.
2. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AMF, Chen HHT: **The randomized trials of breast cancer screening: what have we learned?** *Radiol Clin Nth Am* 2004, **42**:793-806.
3. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE: **Long-term effects of mammography screening: updated overview of the Swedish randomised trials.** *Lancet* 2002, **360**:724.
4. Paci E, Warwick J, Falini P, Duffy SW: **Overdiagnosis in service screening: should the increase in breast cancer incidence rates necessarily be a cause for concern?** *J Med Screen* 2004, **11**:23-27.
5. Anttila A, Koskela J, Hakama M: **Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland.** *J Med Screen* 2002, **9**:153-158.
6. McCann J, Treasure P, Duffy SW: **Modelling the impact of detecting and treating carcinoma in situ in a breast screening programme.** *J Med Screen* 2004, **11**:117-125.
7. Evans AJ, Pinder SE, Ellis IO, Wilson AR: **Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease?** *J Med Screen* 2001, **8**:149-151.
8. Bjurstam N, Björnelid L, Warwick J: **The Gothenburg Breast Screening Trial.** *Cancer* 2003, **97**:2387-2396.

9. Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, Ljungquist U, Lundstrom B, Manson JC, Eklund G, *et al.*: **Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare.** *Lancet* 1985, **1**:829-832.
10. **WinBUGS 1.4.1** [<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>]
11. **CODA: Convergence Diagnostics and Output Analysis Software 0.4** [<http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml>]
12. Ernster VL, Barclay J: **Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma.** *J Natl Cancer Inst Monogr* 1997, **22**:151-156.
13. Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, Yankaskas BC, Rosenberg R, Carney PA, Kerlikowske K, *et al.*: **Detection of ductal carcinoma in situ in women undergoing screening mammography.** *J Natl Cancer Inst* 2002, **94**:1546-1554.
14. Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW: **Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening.** *Eur J Cancer* 2003, **39**:1746-1754.
15. Zahl PH, Strand BH, Maehlen J: **Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study.** *Br Med J* 2004, **328**:921-924.
16. Vainio H, Bianchim F: *Breast Cancer Screening*. Lyon: IARC Press; 2002.