# E3. Over-diagnosis and breast cancer screening

A. Robin M. Wilson and Andrew J. Evans

Nottingham Breast Institute, City Hospital, Hucknall Road, Nottingham NG5 1PB, UK

## Introduction

This abstract outlines some of the recent data published on over-diagnosis of breast cancer and introduces the debate on this subject.

The efficacy of breast cancer screening as a valid method of reducing mortality from breast cancer continues to be the subject of debate. In 2001 Olsen and Gotzsche published a Cochrane review on breast screening that questioned the validity of the various randomised trials and concluded that screening was not an effective intervention for the reduction in breast cancer mortality and, furthermore, that screening caused harm through over-diagnosis and over-treatment [1,2]. Subsequent commentaries and reviews of long-term follow-up of the screening data, most notably those carried out by the World Health Organization through its International Agency for Research on Cancer (IARC) and reviews of the Swedish trials, have dismissed the concerns raised by Olsen and Gotzsche and concluded that breast screening can result in a significant mortality benefit [3].

The IARC report also assessed the issue of over-diagnosis and over-treatment associated with breast screening. No firm conclusions are given in the IARC report on the significance of over-diagnosis, other than to conclude that there may be cause for concern and that detailed studies of the different biology and behaviour of breast malignancy are required so that better understanding can lead to better management of borderline malignant disease. Therefore there remains some concern that breast screening results in the detection of breast cancer that would not have come to light in the lifetime of the host in the absence of screening – so-called pseudo-disease – and that this inevitably leads to unnecessary treatment. Quantifying this side-effect of screening is the subject of much debate, with widely differing views on the level of any detrimental effect of over-diagnosis on populations being offered screening. Clearly, if screening results in the diagnosis of a significant amount of disease that would never threaten life this would be cause for concern, and women considering participating in screening would need to be fully informed of this potential risk.

The widely differing views and interpretations of over-diagnosis appear to have been expressed from analyses of apparently the same data [4–9]. Much of this conflicting opinion is based on data from trials and assumptions that are up to 40 years old. In the light of more recent information on screening outcomes and better understanding of breast cancer biology the European Group for Breast Cancer Screening (EGBCS) in partnership with the European Society of Mastology (EUSOMA) reviewed current evidence in a multidisciplinary consensus forum held in Nottingham in September 2003. This abstract outlines some of the data presented and the presents the conclusions to prompt further discussion.

## Concept of over-diagnosis of breast cancer

Breast cancer represents a spectrum of disease, including *in situ* and invasive disease with widely differing potential to threaten life. It is now generally agreed that the vast majority of high-grade ductal carcinoma *in situ* (DCIS) if left untreated will progress to disease that will threaten life. However, at the other end of the spectrum some low-grade *in situ* and invasive carcinoma (particularly tubular cancer) has such indolent behaviour that it is unlikely to progress to life-threatening disease in the lifetime of the host. It is difficult, if not impossible, with a diagnosis of low-grade malignant disease to predict which cancers may progress and which will not. Some disease is likely to be so indolent that it will never threaten life no matter how long it has been present – so-called pseudo-disease. Current practice in these circumstances is to recommend treatment assuming that the disease will progress. Inevitably this means that some women will receive treatment that is unnecessary.

We know that low-grade malignant breast disease is more likely to progress the longer it is present. Therefore, over-diagnosis is not only a function of the biology of the disease but also of the age of the individual at the time it is detected. Over-diagnosis on this basis is therefore more likely to occur in older women in whom there are more frequent competing causes of death. Invasive cancers in the elderly are more likely to be low grade, compounding the problem of over-diagnosis when screening older women. It has also been suggested that in younger women indolent malignant disease may be diagnosed and treated when, had it been left alone, it would not have progressed or even may have regressed as a result of host immunity. Over-diagnosis may affect younger women because screening mammography in younger women has

1359-6349/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.

a bias in the dense breast for detecting a higher proportion of *in situ* disease, represented by microcalcifications. Invasive cancer in younger women tends to be high grade and given the long life expectancy of this group, over-diagnosis of invasive disease is likely to be unusual. It is therefore reasonable to assume that over-diagnosis may be a significant issue at all ages. Quantifying the degree of over-diagnosis is difficult because knowledge of the natural history of DCIS and low-grade invasive lesions is incomplete. The natural history of low-grade DCIS is well-understood, that of high-grade DCIS less so, while the natural histories of intermediate-grade DCIS and tubular cancer are poorly understood [10,11]. It is likely that current techniques used for the investigation of symptomatic breast problems, particularly the widespread use of mammography and ultrasound combined with needle biopsy, will detect some incidental non-life-threatening malignant disease and that this disease will be over-treated. Taking this into account, a reasonable definition of over-diagnosis as a result of screening is the detection by screening of malignant disease that would not otherwise have been detected in the lifetime of the host (over-diagnosis over and above that which would have occurred because of the assessment of symptomatic breast problems).

## Evidence for over-diagnosis and over-treatment

Introducing screening will always appear to result in an increase in the incidence of breast cancer by detecting the disease earlier than would otherwise have been the case. In the age group most commonly screened (50–70 years) this lead-time is estimated to be about 3.7 years. Screening a population of women aged 50 years will result in the detection of breast cancer numbers expected in women aged 53–54 years in the absence of screening. This lead-time effect is not over-diagnosis and is simply anticipated early diagnosis. It is this early diagnosis and resultant down-staging of disease that results in the mortality reduction seen in randomised trials of screening.

However, some of the extra cancers diagnosed through screening will be over and above those anticipated through the lead-time effect, and it is these cancers that represent over-diagnosis. Estimates of the degree of over-diagnosis caused by screening mammography vary from 35% to none. The debate references the results of the randomised controlled trails, observations on the comparative breast cancer incidence rates in population screening programmes, theoretical extrapolation from autopsy studies carried out on women who have died from other causes, and knowledge regarding the natural history of lesions found by screening mammography.

Seven autopsy studies [3] provide evidence that both non-invasive and invasive breast cancer may be present in women not known to have the disease, with up to 1.8% (mean 1.3%) having invasive disease and up to 14.7% (mean 8.9%) *in situ* malignancy. Doubt about the validity of the classification of the malignancies in some of these studies, particularly the diagnosis of *in situ* cancer, means that accurate prediction of the likely levels of potential diagnosis at any given age cannot be made on the basis of these data. However, the data clearly establish the principle that there is potential for over-diagnosis through screening and that this side-effect is most likely to involve *in situ* disease.

Evidence from the randomised trials is conflicting. In four trials where the control group were not subsequently screened (Edinburgh, Canada 1 and 2 and Malmø) the observed increases in incidence of breast cancer in the study groups were 24–38%. However, the Canadian studies were not population-based and all women were physically examined, while the Edinburgh study used cluster not individual randomisation, leading to a decrease in the comparability of active and passive study populations. This means that these estimates of over-diagnosis may not be accurate. In the three other Swedish trials (Stockholm, Two-counties and Göteborg) the control groups were screened at the end of the study periods and the incidences of breast cancer in the study and control groups were then found to be the same. If over-diagnosis was a problem, this finding suggests that it is confined to the first (prevalent screen). However, the Two-counties study has reported that the incidence of breast cancer in the study and control groups at subsequent (incident) screening are the same, suggesting that over-diagnosis was not a significant problem.

Comparing rates of breast cancer in two neighbouring communities in the Netherlands, one offered screening (Nijmegen) and the other not (Arnhem), there was an observed increase in breast cancer incidence of 30% in the screened population at the prevalent screen, whereas the incidences in the two populations were similar in the two populations at subsequent follow-up. How much of this increased incidence was due to lead-time bias and how much to over-diagnosis is unknown.

Overall, the opinion from the published commentaries and discussion at the consensus meeting is that data from randomised trials and population-based comparisons indicate that over-diagnosis is occurring as a result of screening and that it occurs to a greater extent at the first (prevalent) screen, but that an accurate quantitative assessment of what proportion of screen-detected breast cancer represents over-diagnosis is not possible.

An alternative method of trying to quantifying over-diagnosis is to look at the incidences of breast cancer diagnoses in screened populations and compare these with the incidences found immediately prior to the

introduction of screening. In Finland and the United Kingdom (UK) population, screening has been available since the late 1980s and in both countries the incidence of breast cancer in the target age group has risen by approximately 50%. Part of this clearly represents the lead-time effect of early diagnosis, but the rise in incidence exceeds what can be explained simply by early diagnosis. Again, the data from the UK indicate that excess over expected incidence of breast cancer diagnosis occurs particularly in the cohort of the population screened for the first time. Assuming that an age-specific increase in underlying incidence of around 1.5% observed before screening was introduced is likely to have continued, it has been estimated that in the UK programme around 6% of cancers diagnosed represent over-diagnosis. However, there is good evidence that the underlying incidence has been increasing significantly more than this, particularly as a result of the extensive use of hormone replacement therapy by women in the screening age group. On the basis of the data from the Million Women Study [12], it has been estimated that the use of hormone replacement therapy accounts for half of the increase in the underlying incidence of breast cancer currently observed in the UK. This means that the amount of true over-diagnosis is likely to be less than the 6% calculated using the model of Boer and colleagues.

Paci and colleagues [13] have recently published a review of the Florence screening program, with particular reference to the issue of over-diagnosis. Using the model of calculating the predicted increase in incidence due to the lead-time effect and comparing this with the observed increase in incidence in the screened population, they report no evidence of over-diagnosis of invasive cancer, but a 5% over-diagnosis of *in situ* cancer. They concluded that screening should concentrate on the diagnosis of small invasive cancers rather than the diagnosis of *in situ* disease.

The importance of detecting *in situ* breast cancer through screening is also the subject of much debate with widely differing views. Some suggest that DCIS is extremely important as this represents minimal disease that will progress to invasive disease and threaten life, and that early detection and treatment offers the opportunity for cure of malignancy that would have otherwise inevitably have developed into life-threatening disease. Others suggest that the detection of DCIS at screening has little impact on screening mortality. Therefore assessment of the influence of the diagnosis of DCIS on the effectiveness of screening and its contribution to the unwanted side-effect of over-diagnosis are clearly important and pertinent to current screening practice. Duffy and colleagues [6] reviewed the results of the Swedish Two-counties trial to assess the contribution of DCIS to the down-staging of disease. They concluded that diagnosis of DCIS through screening makes only a relatively small contribution (5%) to the overall mortality reduction. However, in the Two-counties trial only 9% of cancers detected were DCIS. Duffy and colleagues recognise that higher detection rates of DCIS are likely to have more of an impact on mortality, but suggest that to achieve sufficient diagnosis rates of DCIS to produce a more significant mortality reduction would also result in considerable over-diagnosis of DCIS. They also point out that there appears to be no correlation in the screening trials between the detection rates of DCIS and the mortality reduction achieved where the proportion of all cancers detected that were DCIS varied from 5% to 16%. However, they also express the opinion that treatment of screen-detected DCIS is important as the excess of DCIS in the study group was balanced by a deficit of invasive cancers, suggesting that, at least in the Two-counties trial, over-diagnosis of DCIS was not significant. Current detection rates of DCIS observed in population-screening programmes are considerably higher than were observed in randomised controlled trials. In the UK, DCIS represents approximately 25% of the cancers detected at the prevalent screen and 20% of cancer at subsequent screens. This suggests that, as previously observed, around 5% of cancer detected at screening (20% of screen-detected DCIS) represents over-diagnosis. The fact that the rates of detection of DCIS at subsequent screens remain high, assuming that the signs were not missed previously, suggests that the majority of DCIS develops rapidly and therefore that DCIS detected by screening is less likely to be of low grade. Furthermore, concern that increased proportional rates of DCIS are more likely to represent over-diagnosis seem to be unfounded as the large majority of screen-detected DCIS, in keeping with the persisting high rates of detection at subsequent screens, is either high- or intermediate-grade and only a relatively small proportion is low-grade. Evans and colleagues have reported that only 13% of screen-detected DCIS is of low grade, with 18% representing intermediate-grade disease, but the vast majority (69%) being of the high-grade type. This reflects the detection bias of screening mammography, where the vast majority of DCIS is diagnosed through the detection of microcalcifications which is seen much more commonly with high-grade DCIS than low-grade disease because of the comedo necrosis characteristic of high-grade disease. The additional importance of the detection of microcalcifications that represent DCIS with necrosis has been emphasised by Evans and colleagues, who reported that detection of small clusters of microcalcifications that represent high-grade DCIS often leads to the detection of very small high-grade invasive carcinomas that are at a stage where treatment is likely to be beneficial.

Understanding of the biology of DCIS has improved considerably over recent years and this new information on the behaviour of the different types of DCIS contributes to the debate on the significance of detecting DCIS at screening. The genetic profile of DCIS shows that there is a direct link between low-grade types of DCIS and low-grade invasive cancers, both showing deletion of chromosome 16q, while high-grade DCIS shows similar characteristics to moderate- and high-grade invasive cancer with over-expression of 17q. These findings fit well with previous observations that low-grade DCIS is likely to be associated with low-grade invasive cancer and high-grade DCIS with high-grade cancer. These findings are important for screening, as it has been shown that screening has a bias towards the detection of high-grade DCIS that is highly likely otherwise to develop into life-threatening intermediate- or high-grade invasive cancer. If this were the case then it would be expected that screening services showing a high detection of DCIS would be reflected in later years by a lower incidence of invasive disease. In fact most of the observations from the screening trials and the data from screening programmes appear not to show this effect. However, data presented at the consensus meeting comparing two regions in the UK that have had traditionally low and high comparative DCIS detection rates show a trend towards lower invasive cancer detection incidence in the region with the higher DCIS rate compared with the region with lower DCIS rate.

There was also considerable discussion about the significance of the generally observed higher proportion of special-type and low-grade invasive cancers detected at screening in comparison with the profile of cancers detected in unscreened women. Tubular carcinoma is an example. This special type of low-grade breast cancer appears to have a very indolent behaviour and is very rarely associated with metastatic disease. Even when it is, death attributable to metastatic spread occurs some 10–15 years later than is observed with other tumour types and grades. Tubular cancer represents some 8% of screen-detected invasive breast cancer (6% of all screen-detected cancer). What is not known is how rarely or frequently tubular cancer de-differentiates into higher-grade tumours of mixed tubular and ductal non-special-type morphology with enhanced potential to threaten life (so-called phenotypic drift).

The consensus view of over-diagnosis and resultant over-treatment of breast cancer as a result of screening is that it does occur, mainly at the prevalent screen, and is largely confined to the diagnosis of low-grade *in situ* and low-grade invasive breast cancer. If it is assumed that all low-grade DCIS and all tubular carcinomas represent over-diagnosis, then screening results in an over-diagnosis rate of around 10%. The consensus view may be that at this level of over-diagnosis the mortality benefits of breast screening by mammography may still significantly outweigh any negative effects.

Another issue is whether screen-detected breast cancer leads to over-treatment compared with symptomatic breast cancer of the same stage and grade. The consensus view was that there is insufficient evidence on which to base an opinion on this particular point. However, views have been expressed that a false impression of over-treatment may be given where screen-detected cancer is treated in specialist centres while symptomatic cancer is treated in non-specialist centres – under-treatment of non-screen-detected breast cancer. The only recent evidence considered on over-treatment was that published by the Florence group – this evidence suggests that there is no evidence of differential over-treatment of screen-detected breast cancer [14].

## References


[1] Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001, **358**, 1340–2.

[2] Olsen O, Gotzsche PC. Systematic review of screening for breast cancer mammography. Available from http://image.thelancet.com/lancet/extra/fullreport.pdf

[3] World Health Organization. *WHO 7th Handbook on Cancer Prevention*. Lyons: IARC, 2002.

[4] Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammographic screening: update overview of the Swedish randomised trials. *Lancet* 2002, **359**, 909–19.

[5] Tabar L, Vitak B, Tony HH, *et al.* Beyond randomised controlled trials: organised mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001, **91**, 1724–31.

[6] Duffy S, Tabar L, Chen HH, *et al.* The impact of organised mammographic screening on breast carcinoma mortality in seven Swedish counties. *Cancer* 2002, **95**, 458–69.

[7] Zahl PH, Heine B, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *Br Med J* 2004, **328**, 921–4.

[8] Baum M. Commentary: False premises, false promises and false positives – the case against mammographic screening for breast cancer. *Int J Epidemiol* 2004, **33**(1), 66–7.

[9] Thornton H, Edwards A, Baum M. Women need better information about routine mammography. *Br Med J* 2003, **327**, 101–3.

[10] Evans AJ, Pinder SE, Ellis IO, Wilson ARM. Screen detected Ductal Carcinoma in Situ (DCIS): Overdiagnosis or an obligate precursor of invasive disease. *J Med Screen* 2001, **8**, 149–51.

[11] Evans AJ, Burrell HC, Pinder SE, Ellis IO, Wilson ARM. Detecting which invasive cancers at mammographic screening saves lives? *J Med Screen* 2001, **8**, 86–9.

[12] Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003, **362**, 419–27.

[13] Paci E, Warnick J, Falini P, Duffy SW. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 2004, **11**, 23–7.

[14] Paci E, Duffy S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005, **7**, 266–70.