

Bond University
Research Repository



**Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence:
Improving information on the effects of attending screening mammography**

Jacklyn, Gemma; Glasziou, Paul; Macaskill, Petra; Barratt, Alexandra

Published in:
British Journal of Cancer

DOI:
[10.1038/bjc.2016.90](https://doi.org/10.1038/bjc.2016.90)

Licence:
CC BY-NC-SA

[Link to output in Bond University research repository.](#)

Recommended citation(APA):

Jacklyn, G., Glasziou, P., Macaskill, P., & Barratt, A. (2016). Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence: Improving information on the effects of attending screening mammography. *British Journal of Cancer*, 114(11), 1269-1276. <https://doi.org/10.1038/bjc.2016.90>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

Keywords: mammography; patient compliance; mass screening; meta-analysis; breast neoplasm; overdiagnosis

Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence: improving information on the effects of attending screening mammography

Gemma Jacklyn^{*1}, Paul Glasziou^{2,3}, Petra Macaskill³ and Alexandra Barratt⁴

¹Sydney School of Public Health, The University of Sydney, Edward Ford Building (A27), Sydney, NSW 2006, Australia; ²Centre for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, QLD 4229, Australia; ³Screening & Test Evaluation Program (STEP), Sydney School of Public Health, The University of Sydney, Edward Ford Building (A27), Sydney, NSW 2006, Australia and ⁴Centre for Medical Psychology & Evidence-based Decision-making (CeMPED), Sydney School of Public Health, The University of Sydney, Edward Ford Building (A27), Sydney, NSW 2006, Australia

Background: Women require information about the impact of regularly attending screening mammography on breast cancer mortality and overdiagnosis to make informed decisions. To provide this information we aimed to meta-analyse randomised controlled trials adjusted for adherence to the trial protocol.

Methods: Nine screening mammography trials used in the Independent UK Breast Screening Report were selected. Extending an existing approach to adjust intention-to-treat (ITT) estimates for less than 100% adherence rates, we conducted a random-effects meta-analysis. This produced a combined deattenuated prevented fraction and a combined deattenuated percentage risk of overdiagnosis.

Results: In women aged 39–75 years invited to screen, the prevented fraction of breast cancer mortality at 13-year follow-up was 0.22 (95% CI 0.15–0.28) and it increased to 0.30 (95% CI 0.18–0.42) with deattenuation. In women aged 40–69 years invited to screen, the ITT percentage risk of overdiagnosis during the screening period was 19.0% (95% CI 15.2–22.7%), deattenuation increased this to 29.7% (95% CI 17.8–41.5%).

Conclusions: Adjustment for nonadherence increased the size of the mortality benefit and risk of overdiagnosis by up to 50%. These estimates are more appropriate when developing quantitative information to support individual decisions about attending screening mammography.

Screening mammography can help and harm. The Independent UK Panel on Breast Cancer Screening stated that mammography has the potential to reduce a woman's risk of dying from breast cancer, but that this must be balanced against the possibility of physical, psychological and financial harm because of unnecessary diagnosis and treatment (Marmot *et al*, 2013). There is international recognition of the need to develop objective information on the benefits, harms and uncertainties of screening mammography

to accompany invitations to screening (Canadian Task Force on Preventive Health Care, 2011; Marmot *et al*, 2013; Biller-Andorno and Jüni, 2014; Depypere *et al*, 2014). To aid women with informed decision making, material should present information about, and numerical outcomes of, breast cancer mortality reduction and overdiagnosis – also termed overdiagnosis – of screening mammography (Elwyn *et al*, 2006). Clear and balanced information that is relevant and appropriate for individuals

*Correspondence: G Jacklyn; E-mail: gemma.jacklyn@sydney.edu.au

Received 14 December 2015; revised 24 February 2016; accepted 12 March 2016; published online 28 April 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16

can help women assess the trade-offs using their own values and preferences (Jørgensen *et al*, 2009; Pace and Keating, 2014; Hersch *et al*, 2015).

Randomised controlled trials (RCTs) are the most rigorous method of establishing numerical estimates and can determine whether a causal relationship exists between an intervention and outcome. Participants are generally analysed in the groups to which they were allocated, regardless of missing outcome data or nonadherence to the trial protocol (Hollis and Campbell, 1999; Moher *et al*, 2010). This is termed intention-to-treat (ITT) analysis and is the traditional statistical method when analysing data and presenting comparative results (Hollis and Campbell, 1999). It helps control selection bias and therefore confounding (Peto *et al*, 1976), and provides population estimates of the benefits and harms of an intervention in practice (Sommer and Zeger, 1991). But using an ITT analysis dilutes significant measures of effect. If we analyse groups exactly as randomised when participants do not adhere to the trial protocol, estimates attenuate towards the null and statistical power is reduced (Newcombe, 1988; Sommer and Zeger, 1991).

Although ITT estimates derived from RCTs are helpful when planning policy, they are not suitable for individual decision making about the potential effects of an intervention as received (Baker *et al*, 2002). With screening mammography, estimates for women invited to screening are attenuated compared with that for women who actually attend screening. Control group estimates contribute to attenuation when women seek out and attend screening outside of the trial. This is because women who are screened outside of the trial can be assumed to receive some benefit and harm. This reduces the total number of breast cancer deaths in the control group and inflates breast cancer incidence that would underestimate mortality benefit and overdiagnosis. Furthermore, individual participants are likely to have a preference for a particular group to which they are randomised that can influence adherence to the protocol.

This is often observed in RCTs of cancer screening (Hewitson *et al*, 2007; Ilic *et al*, 2013). Nonadherence in the intervention group – when participants do not attend screening – and in the control group – when participants attend screening outside of the trial – can be problematic. For example, in the most recent screening mammography RCT, the UK Age Trial, adherence in the intervention group was 68% (Moss *et al*, 2006). Although such adherence rates may reflect real-world participation in screening programmes, they are not helpful when estimating benefits and harms for women who regularly attend screening and receive the ‘package’ as recommended by screening organisations. Thus, alternative methods are required to provide unbiased estimates of the effect of screening mammography on individuals.

Several meta-analyses of screening mammography of women in the target age range of 50–75 years have been conducted using the ITT principal, yet estimates of the effects remain contested (Wald *et al*, 1993; Kerlikowske *et al*, 1995; Demissie *et al*, 1998; Blamey *et al*, 2000; Humphrey *et al*, 2002; IARC, 2002; Nyström *et al*, 2002; Canadian Task Force on Preventive Health Care, 2011; Duffy *et al*, 2012; Gøtzsche and Jørgensen, 2013). Because of this ongoing debate, in 2012 a group of independent experts from the United Kingdom reviewed the benefits and harms of screening mammography (Marmot *et al*, 2013). Their meta-analysis of the mortality benefit of screening from nine trials was primarily based on 13 years of follow-up data reported in the Cochrane Review (Gøtzsche and Nielsen, 2009). They reported a pooled relative risk reduction (RRR) in breast cancer mortality of 20% among women invited to screening. When determining harm due to overdiagnosis of breast cancer, only the Malmö I and Canada I and II trials were used (Miller *et al*, 2000, 2002; Zackrisson *et al*, 2006). After 6–15 years of follow-up, meta-analysis of these trials found an overdiagnosis proportion of 19% during the screening period. These ITT

estimates of benefit and harm are useful for assessing the effect of screening mammography on society. However, they do not account for nonadherent participants and thus provide an attenuated estimate of the effect of receiving screening and do not reflect the effect on an individual.

Methods to obtain summary estimates from meta-analysis that adjust for nonadherence yet respect the randomisation have been developed for binary outcomes (Glasziou, 1992; Baker and Lindeman, 1994; Cuzick *et al*, 1997), and pooled deattenuated mortality estimates have previously been calculated for screening mammography (Glasziou, 1992; Glasziou and Houssami, 2011). To our knowledge, pooled deattenuated estimates of overdiagnosis with meta-analysis have not been reported.

We attempt to answer the question from an individual woman’s perspective, how large is the benefit of screening mammography in terms of reduced risk of dying from breast cancer and how substantial is the risk of harm in terms of overdiagnosis if she decides to participate in a regular programme of screening mammography? We aim to do this by deattenuating, that is, adjusting for attenuation of the screening effect caused by women who do not adhere to the trial protocol. We present pooled estimates of the prevented fraction of mortality benefit and percentage risk of overdiagnosis for RCTs of screening mammography adjusted for nonadherence.

MATERIALS AND METHODS

Outline of approach. Our study built on the UK Panel 2012 systematic review of the benefit and harm of screening mammography (Marmot *et al*, 2013). The Independent UK Panel on Breast Cancer Screening reviewed the published work, assessed the risk of bias in included studies and pooled results to provide estimates of effect for the benefit and harm. Our intention was to put into practice the recommendations made by The Independent UK Panel report on Breast Cancer Screening to provide ‘*clear communication of these harms and benefits to women*’ (Marmot *et al*, 2013). We do this by deattenuating the estimates of effect to reflect the efficacy of screening mammography for women who actually attend. We combined data from relevant mammography trials and extended the approach previously described by Glasziou (1992) to conduct a random-effects meta-analysis for mortality benefit and overdiagnosis risk. This method enabled us to obtain a combined, deattenuated prevented fraction: the proportion of deaths attributable to breast cancer prevented by screening mammography. For the combined, deattenuated estimate of overdiagnosis risk we estimated the variance (with 95% confidence intervals (CIs)) for study-specific deattenuated estimates of overdiagnosis. Using the combined results we were able to determine the effectiveness of screening mammography on reducing breast cancer mortality and the risk of overdiagnosis from the perspective of an average-risk woman between the ages of 39 and 75 years who chooses to participate ‘regularly’ in screening. ‘Regularly’ refers to the screening interval across the trials that ranged from 12 to 33 months. Adherence refers to a person’s health-related behaviour in accordance with the group assignment in a trial. We use the term adherence rather than the tradition term compliance as this reflects the active decision making of participants in a trial. Compliance denotes passive obedience and has a paternalistic connotation.

Data sources: mammography trials. Relevant studies were identified from the Independent UK Breast Screening Review (Marmot *et al*, 2013) that in turn sourced publications and data from the Cochrane Review ‘Screening for breast cancer with mammography’ (Gøtzsche and Nielsen, 2009). Only RCTs were

used. When analysing the mortality benefit of screening mammography, similar to the Independent UK Breast Screening Review and Cochrane Review, we included 9 of the 11 available RCTs in our analysis (NY HIP trial; Malmö I; Swedish Two County; Koppberg and Östergötland; Canada I and II; Stockholm; Göteborg; and UK Age trial) (Shapiro *et al*, 1982; Tabár *et al*, 1985; Andersson *et al*, 1988; Frisell *et al*, 1997; Miller *et al*, 2000, 2002; Bjurstam *et al*, 2003; Moss *et al*, 2006). The Edinburgh trial was excluded because of imbalances between groups in baseline socioeconomic level: 26% of the women in the control and 53% in the screening group belonged to the highest socioeconomic level (Alexander *et al*, 1999). Malmö II was excluded because of insufficient follow-up time (Andersson and Janzon, 1997). Of the nine trials used to estimate mortality benefit, three clearly did not offer screening to the control group at the end of the active study period. Thus, when determining harm due to overdiagnosis of breast cancer, only the Canada I and II trials (Miller *et al*, 2000, 2002) and a subgroup of women aged 55–69 years from the Malmö I trial were used (Zackrisson *et al*, 2006).

Data extraction. One author (GJ) independently extracted outcome data from primary papers where possible. When the required information was not obtainable, or there were discrepancies in reporting of outcomes for the same trial, the author relied on the Cochrane review (Gøtzsche and Nielsen, 2009). Extracted data included: number of women randomised, adherence rates (attendance in the intervention group and opportunistic screening in the control group), number of breast cancers detected and breast cancer mortality.

Adherence rates. Adherence rates were extracted from the Cochrane review and primary studies (Shapiro *et al*, 1985; Tabár *et al*, 1985; Andersson *et al*, 1988; Frisell *et al*, 1997; Miller *et al*, 2000, 2002; Moss *et al*, 2006; Gøtzsche and Jørgensen, 2013). We estimated adherence for one study: the HIP trial. Adherence in the screened group was reported but adherence in the control group was not. We assumed 100% adherence in the control group as mammography was a relatively new technology and screening was unlikely to be available outside of the trial during the active screening period from December 1963 to June 1966 (National Cancer Institute, 1978).

Deaths attributable to breast cancer (mortality rates). Breast cancer mortality was taken from the Cochrane Review and primary papers that reported the same length of follow-up of 13 years (Shapiro *et al*, 1982; Andersson *et al*, 1988; Tabár *et al*, 1995; Frisell *et al*, 1997; Miller *et al*, 2000, 2002; Bjurstam *et al*, 2003; Moss *et al*, 2006; Gøtzsche and Jørgensen, 2013).

Breast cancer cases (incidence rates). Information on breast cancer diagnosis was obtained from the two Canadian trials (Miller *et al*, 2000, 2002) and the Malmö I trial for women aged 55–69 years (Zackrisson *et al*, 2006). Since the release of the Independent UK Breast Screening Review, the 25-year follow-up data have been published for the Canada trials (Miller *et al*, 2014). The two parts of this trial, however, were not reported separately and thus the data are not suitable for our analysis.

Trial numbers. Data on the number of women in each trial were obtained from the Cochrane Review and primary studies (Shapiro *et al*, 1982; Tabár *et al*, 1985; Tabár *et al*, 1995; Frisell *et al*, 1997; Miller *et al*, 2000, 2002; Bjurstam *et al*, 2003; Moss *et al*, 2006; Gøtzsche and Jørgensen, 2013). Reported numbers vary considerably for some trials. To ensure consistency and accuracy when the exact sample size was in doubt we used the numbers reported in the Cochrane Review.

Deattenuated estimates of effect

Mortality benefit. Glasziou (1992) proposed a method to adjust the ITT estimate of the prevented fraction (PF) from RCTs for nonadherence. The rationale is based on the linear model of Newcombe (1988) for calculating a deattenuated estimate in a single study that involves adjusting the observed difference between the mean outcome for the intervention (\bar{Y}_1) and control (\bar{Y}_2) group by a deattenuation factor (Δ) based on the adherence proportions in the two groups (P_1 and P_2 respectively). The equation for the deattenuation factor is given by:

$$\Delta = P_1 + P_2 - 1$$

Hence, for a specific study, the deattenuated estimate of the prevented fraction (DPF_i) of an intervention effect is:

$$DPF_i = \frac{\bar{Y}_{1i} - \bar{Y}_{2i}}{\Delta_i}$$

For ease of interpretation, readers can think of the deattenuated prevented fraction as a RRR, that is, the proportion of deaths prevented by the intervention.

Glasziou (1992) also described a fixed-effect procedure to combine the study-specific deattenuated estimates and compute a summary estimate (DPF^*) of the deattenuated Intervention effect. Details of this along with formulas for the 95% CIs and variance can be found in the Supplementary Material.

Overdiagnosis. For overdiagnosis, the Independent UK Panel report percentage of risk for screened women. The panel's preferred method for measuring percentage overdiagnosis from the perspective of an individual woman is method C, excess cancers (5–10 years after screening ends) as a proportion of all cancers diagnosed during the screening period in women invited for screening (Marmot *et al*, 2013). The calculation for this is:

$$\frac{(\text{Cumulative incidence in the screened group} - \text{cumulative incidence in the control group})}{\text{Total number of all breast cancers detected in the screened group during the active screening period}}$$

As this is an outcome that can only occur in the screened group, we use an attributable fraction. An explanation of the rationale for this method of calculating overdiagnosis with examples can be found in Welch and Black (2010). When conducting a meta-analysis of percentage risk of overdiagnosis it is important that the denominator includes screen-detected, interval and clinically detected breast cancers in the screened group. This is because the proportion of interval cancers relative to screen-detected cancers increases as the screening interval increases. Excluding interval breast cancers provides an estimate of overdiagnosis that is dependent on screening frequency and applies only to one particular trial. As we use trials with different screening intervals (Canada I and II, 12 months; Malmö I, 18–24 months) we must account for this by including in the denominator all cancers detected in the screened group.

The formula for the variance of the deattenuated estimate of percentage risk of overdiagnosis (ODP) in the screened group was computed assuming (1) an underlying binomial distribution for percentage overdiagnosis and (2) Δ to be a constant. Using standard variance formulae, the approximate variance of ODP for study I is given by:

$$\text{var}\left(\frac{ODP_i}{\Delta}\right) = \frac{1}{\Delta^2} \text{var}(ODP_i)$$

where the study-specific percentage risk of overdiagnosis (OD_i) is

expressed as a proportion, Δ is a constant and the variance formula is:

$$var(ODP_i) = \frac{ODP_i(100-ODP_i)}{n_i}$$

For both the deattenuated estimates of prevented fraction and overdiagnosis a random-effects summary estimate was calculated using the method of (DerSimonian and Laird, 1986). We used the I^2 measure of heterogeneity to assess the extent to which the results of individual studies were consistent (Higgins and Thompson, 2002). A spreadsheet was developed and statistical analyses performed using Microsoft Excel software for the analyses. In order to obtain and compare bias-corrected parameter estimates, we bootstrapped the DerSimonian and Laird random-effects model using the bdl option in metaan (Stata version 13, StataCorp LP, College Station, TX, USA) to estimate the between-study variance and heterogeneity parameters. Study-specific estimates and their corresponding variance from the spreadsheet were input to metaan to obtain summary estimates, 95% CIs and I^2 based on 1000 replications.

Finally, we performed sensitivity analyses to account for the uncertainty in the deattenuation factor, Δ . When data on the proportion of women screened across all screening rounds were available, we averaged this for individual trials (NY HIP trial; Malmö I; Kopparberg, Östergötland; Canada I and II; Göteborg; and UK Age Trial).

RESULTS

Adherence. Table 1 presents rates of adherence to trial protocol. Figures ranged from 100% to 65% for the screened group and 100% to 76% for the control group.

Breast cancer-specific mortality benefit. Table 2 compares death from breast cancer in screened vs control women. Figure 1 shows the results from random-effects meta-analyses for both the ITT and deattenuated estimates of the prevented fraction of breast cancer mortality due to screening mammography. Estimates of benefit increase with deattenuation, with a decrease in precision. The overall ITT prevented fraction, comparing invited vs control women, is 0.22 (95% CI 0.15–0.28). After adjusting for non-adherence, the prevented fraction is 0.30 (95% CI 0.18–0.42). There was moderate heterogeneity in the ITT prevented fractions across different trials ($I^2 = 36%$, 95% CI 0–70.7%) that decreased slightly with deattenuation ($I^2 = 34%$, 95% CI 0–69.8%).

Overdiagnosis. Table 3 compares breast cancer detection during the entire follow-up period in screened vs control women. Figure 2 shows the results from random-effects meta-analyses for the two estimates of percentage risk of overdiagnosis due to screening mammography. An increase in overdiagnosis is observed with deattenuation. The overall ITT percentage risk of overdiagnosis is 19.0% (95% CI 15.2–22.7%). After adjusting for nonadherence, the percentage risk of overdiagnosis is 29.7% (95% CI 17.8–41.5%). There was substantial heterogeneity in the ITT percentage risks of overdiagnosis from the three trials ($I^2 = 64.8%$, 95% CI 0–89.9%) that increased with deattenuation ($I^2 = 92.0%$, 95% CI 79.8–96.8%). The corresponding bootstrap results are very similar.

Sensitivity analysis. When using an average of attendance in the screened group across all screening rounds, the pooled deattenuated prevented fraction increased to 0.34 (95% CI 0.21–0.47) and the pooled deattenuated percentage risk of overdiagnosis also increased to 32.1% (95% CI 20.3–44.0%).

DISCUSSION

Summary of key findings. We present the first combined, deattenuated estimate of overdiagnosis from randomised controlled trials of breast cancer screening for use when developing quantitative information for individual women who participate regularly in screening. The deattenuated percentage risk of overdiagnosis of 30% for women who regularly attend screening mammography during the screening period is considerably higher than 19% estimate for invited women. The pooled deattenuated prevented fraction estimate for breast cancer mortality from trials is 0.30 for those who attended screening, again, considerably greater than 0.20 for those invited.

Strengths

Random-effects model. We used a random-effects model to pool the deattenuated estimates. When combining ITT estimates in a fixed-effect meta-analysis, more weight may be given to studies with poor adherence if they have a large sample size (Glasziou, 1992). A meta-analysis of deattenuated estimates avoids this problem and previous studies have used this approach to combine screening mammography trial estimates of mortality reduction (Glasziou, 1992; Glasziou and Houssami, 2011). We used a random-effects model as it is better suited to the heterogeneity between screening mammography trials, for example differences in the breast screening method, screening interval and number of screening rounds (DerSimonian and Laird, 1986). Whilst the assigned weights are more balanced, it produces wider confidence intervals around the pooled estimate, reflecting a decrease in precision (Higgins and Thompson, 2002). Although our analysis should better allow for heterogeneity between trials that is reflected in the reduced I^2 for the DPF, the I^2 for our deattenuated

Table 2. Number of breast cancer deaths and participants in randomised trials of screening mammography

	Number of breast cancer deaths		Number of participants	
	Screened	Control	Screened	Control
New York HIP	218	262	31 000	31 000
Malmö I	87	108	20 695	20 783
Kopparberg	126	104	38 589	18 582
Östergötland	135	173	38 491	37 403
Canada I	105	108	25 214	25 216
Canada II	107	105	19 711	19 694
Stockholm	66	45	40 318	19 943
Göteborg	88	162	21 650	29 961
UK Age trial	105	251	53 884	106 956

Table 1. Adherence to the study protocol in the randomised trials of screening mammography

	New York HIP	Malmö I	Kopparberg	Östergötland	Canada I	Canada II	Stockholm	Göteborg	UK Age Trial
Adherence									
Screened	65%	74%	89%	90%	100%	100%	81%	84%	68%
Control	100%	76%	87%	87%	74%	83%	92%	82%	92%

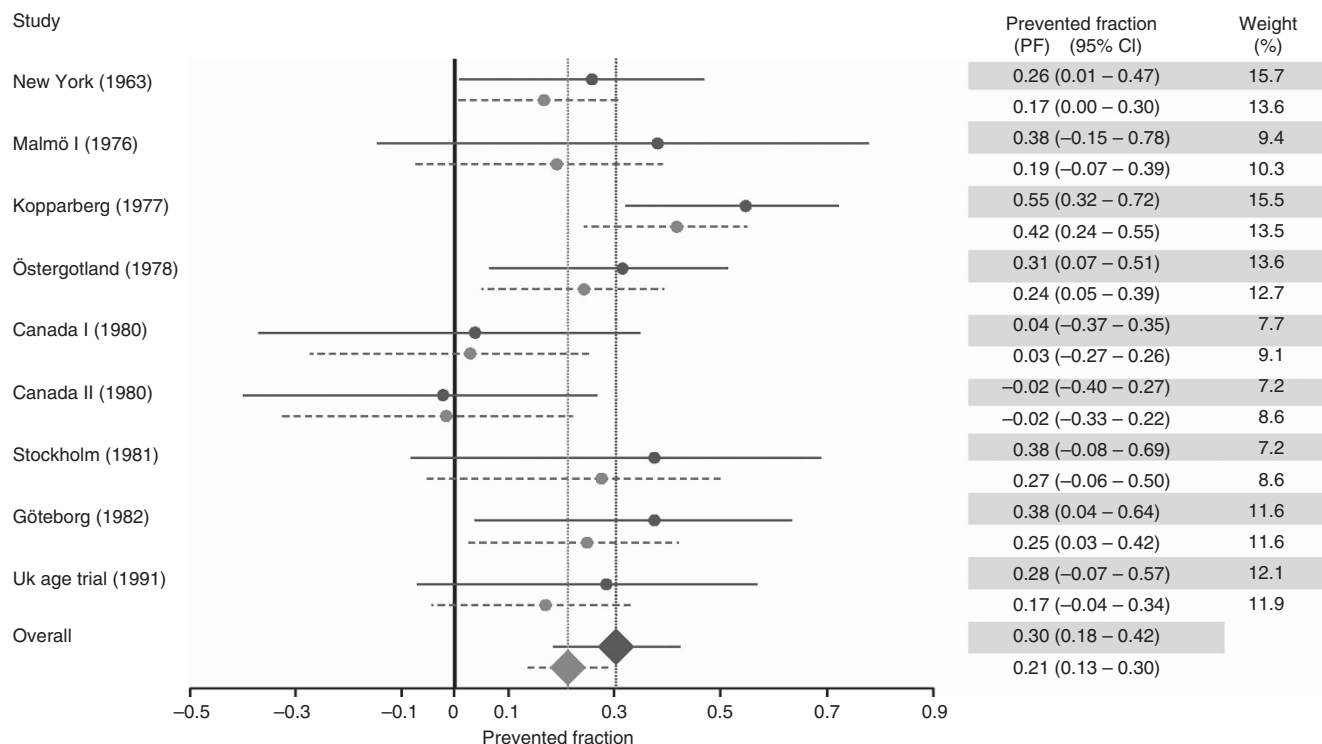


Figure 1. Meta-analysis of estimates of prevented fraction of breast cancer mortality after 13 years of follow-up with (blue) and without (grey) adjustment for adherence. A full colour version of this figure is available at the *British Journal of Cancer* journal online.

Table 3. Number of breast cancers diagnosed in the three randomised trials of screening mammography suitable for estimating overdiagnosis

	Number of breast cancers diagnosed during the entire follow-up period		Excess breast cancers diagnosed in the screened group at the end of follow-up	Breast cancers diagnosed during the screening period in women invited to screening
	Screened	Control		
Malmö I (55–69)	780	698	82	438
Canada I (40–49)	663	581	82	361
Canada II (50–59)	693	626	67	420

estimate of overdiagnosis still suggests considerable remaining heterogeneity.

Percentage risk of overdiagnosis during screening. When calculating percentage risk of overdiagnosis, different methods address different questions. For individual women who are considering screening, the most appropriate estimate of overdiagnosis is one that presents the risk for regularly screened women, during the active screening period (Marmot *et al*, 2013). This is because overdiagnosis can only occur in women who are screened, and including in the denominator cancers diagnosed after screening ends dilutes the estimate of effect. Furthermore, it is important to include both screen-detected and interval cancers as excluding interval breast cancers provides an estimate of overdiagnosis that is dependent on the screening interval. Our choice of estimate is in line with the Independent UK Panel preferred measure of percentage risk of overdiagnosis from the perspective of an individual woman, Method C (Marmot *et al*, 2013).

Our deattenuated analysis avoids introducing selection bias by preserving the ITT analysis. To calculate percentage risk of overdiagnosis we compare outcome event rates among all

participants and then adjust that estimate to a degree determined by the proportion of adherent participants in both arms of each trial. As demonstrated by Newcombe (1988), this deattenuates the estimate to what it would have been had there only been fully adherent participants. Thus, our results are more relevant to women who participate in screening and assumes that people who attend screening in the trials are similar to those who attend screening programmes. Deattenuation will not change the estimate if there is no intervention effect or 100% adherence in the study groups.

Limitations

Uncertainty. The deattenuated study-specific estimates diverge from the unadjusted results. In part, this reflects the heterogeneity between trials and uncertainty in study-specific estimates. Although it is important to note the statistical uncertainty, as indicated by the wide CIs that increase with deattenuation, there are multiple sources of uncertainty that cannot be quantified, including the methodological limitations of component studies. Although our method avoids selection bias, this is only to the extent that outcomes data are complete for all study participants.

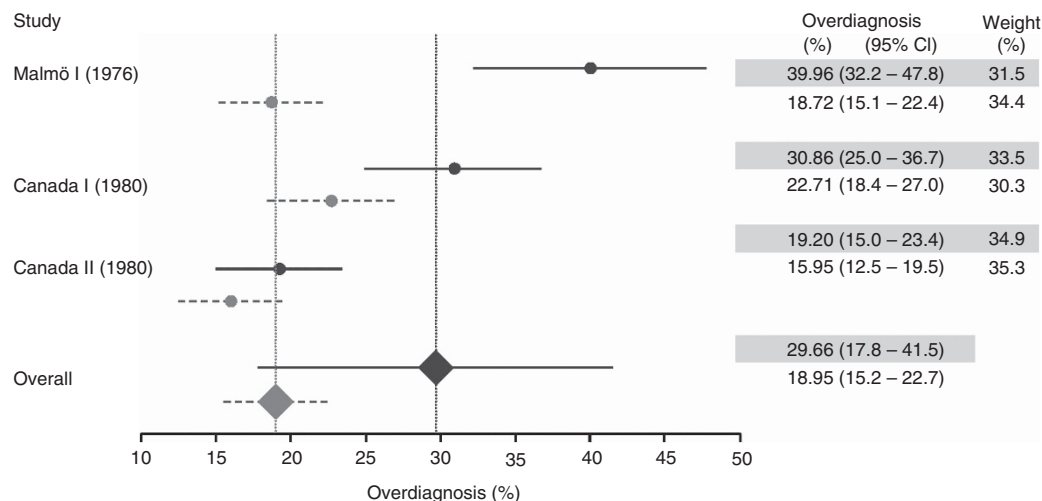


Figure 2. Meta-analysis of estimates of percentage risk of overdiagnosis of breast cancer with (blue) and without (grey) adjustment for adherence. Overdiagnosis is measured from the perspective of an individual woman: excess cancers over the entire follow-up period as a proportion of all cancers diagnosed during screening period in women invited for screening (as recommended by Independent UK Panel on Breast Cancer Screening (Marmot *et al*, 2013), Method C). A full colour version of this figure is available at the *British Journal of Cancer* journal online.

Fortunately with screening mammography trials, outcomes data are near complete as cancer incidence and mortality events were ascertained by reference to population-based cancer and death registries. We acknowledge, however, that although follow-up rates for outcomes were high (>90%), they were not 100%. Furthermore, our estimates reflect a screening duration that ranges from 3 to 12 years, and may underestimate the prevented fraction and overdiagnosis attributable to a 25-year screening period, the recommended 'package' by many organisations including those in Australia, the United States, Canada and, by 2016, the United Kingdom (Canadian Task Force on Preventive Health Care, 2011; Moser *et al*, 2011; Department of Health Australia, 2016; Siu, 2016), considerably longer than the screening periods evaluated in the randomised trials. Finally, there is uncertainty about the extent to which these estimates may be applicable to contemporary screening mammography because of advances in mammography technology, breast cancer treatment and an overall reduction in breast cancer mortality.

When calculating the variance of the percentage risk of overdiagnosis as a proportion of all cancers detected in the screened group (Marmot method C), we have used a binomial distribution in line with the Independent UK Panel (Marmot *et al*, 2013) and Baker *et al* (2014). When using a different denominator, alternative methods may be more appropriate. Dealing with such statistical issues is beyond the scope of this paper but warrants further exploration.

Adherence. We used the adherence rate for the first round of screening only, when attendance is highest. In reality, adherence is a continuous variable. To reflect the full attenuation of the estimates we would need to obtain adherence rates for all screening rounds as the estimates of mortality benefit and overdiagnosis from the trials reflect the impact of attendance at 2–10 mammograms. Published data however do not allow for the multi-stage, interval nature of screening. Even if it did, it is not clear how to partition the benefit and harm for each screen as the contribution of individual screening rounds to the overall observed mortality reduction and overdiagnosis is unknown. Thus, it is likely that we have underestimated the deattenuated effects.

Our analysis does not allow for participation in screening programmes by either the intervention or control group after the active screening period of the trial. Systematic screening of the

control group occurred in seven of the nine trials, but information on attendance is not available for most studies. Continued screening by the intervention group or uptake by the control group after the end of the active screening period may have been widespread because of the gradual introduction of mass screening mammography during the 1980s in countries where the trials were undertaken. Both of these issues would dilute the mortality reduction and overdiagnosis estimates.

Methods to adjust for nonadherence in RCTs. The method used in this paper to adjust for non-adherence was originally applied to screening mammography trials by Glasziou (1992). There are many reasons why our deattenuated breast cancer mortality estimate differs to that original analyses conducted over 20 years ago, where the pooled deattenuated prevented fraction for breast cancer mortality was reported as 0.37. Of the five trials included in the analysis of Glasziou (1992) (HIP, Swedish Two County, Malmö, Edinburgh and Stockholm trials), we have excluded the Edinburgh trial but used additional data from the Goteborg, Canada I and II and UK Age trials. The original analysis of Glasziou (1992) used women-years at risk as the denominator; these data are not reported for all trials. We used number of women as the denominator for mortality rate in line with the Cochrane Review and the Independent UK Breast Screening Review. Finally, using a random-effects model means the weights assigned to each study are more balanced compared with the fixed-effect analysis used by Glasziou.

Our deattenuated results also differ to those presented by The Independent UK Breast Screening Review (Marmot *et al*, 2013). They reported a RRR for breast cancer mortality, adjusted for nonadherence in the intervention group, of 25%. The methods however are inexact. The authors estimated trial adherence in the intervention group at 80% and divided the RRR of 20% by that average attendance when adherence differed for individual trials, ranging from 65% to 100%. Furthermore, they did not account for nonadherence in the control groups across the trials that would further attenuate the estimate of RRR and thus increase the deattenuated RRR.

Alternative methods for adjustment have been described by Cuzick *et al* (1997), Baker and Lindeman (1994) and Baker *et al* (2016). The method of Cuzick *et al* (1997) involves stratifying the probability of breast cancer mortality into four groups according to women who are screened and not screened in both the control and

screening arms. The assumption is that the underlying rate in attenders is equivalent to that in the control group adjusted for the rate in the nonattenders. Such detailed data are not reported for most mammography trials. Baker and Lindeman (1994) and Baker *et al* (2016) formulated the latent class IV method for binary outcomes. They use four principal strata: always receivers who would receive screening regardless of randomisation group, compliers who would receive screening only if randomised to screening, never takers and defiers. In each stratum, the response to treatment differs. Estimates are based only on compliers, but require assumptions about the different strata. A similar framework was used by McIntosh (1999) to deattenuate mortality benefit for the HIP Trial.

Implications. We present estimates that can be applied and used to develop information on the benefit and harms of screening mammography for individual women. To do so, we extended existing methods to derive, for the first time, a deattenuated, pooled estimate of percentage risk of overdiagnosis due to screening mammography from three trials. The results demonstrate the usefulness of this approach to other forms of cancer screening where overdiagnosis may be a potential downside, including prostate and lung cancer screening. The information can be applied to local data to develop evidence-based numerical outcomes for decision aids to help individuals weigh up the benefits and harms of screening and make an informed choice, as was done in a recent decision aid trial by Hersch *et al* (2015). Likewise, it will assist clinicians in shared decision making as well as policymakers when developing information materials about cancer screening for individuals.

Methodological flaws and inconsistent reporting of results within individual mammography trials highlight the need for better standards of data collection and reporting by investigators. Most lacking is information on the harms. More detailed information on adherence to the trial protocol per screening round could help to provide a more accurate reflection of attendance and its impact on screening outcomes. This is especially important for trials of emerging technology, including digital mammography and tomosynthesis.

Finally, the uncertainty in the estimates and divergence of individual study results after deattenuation, particularly with overdiagnosis, further highlight the fact that there are few sources of reliable data. Given that each year over 1.4 million women in Australia and 2 million women in the United Kingdom attend breast cancer screening, and they may rely on estimates such as ours to help them weigh up the benefits and harms of screening to make an informed choice, obtaining more precise numbers that reflect changes since the original trials were conducted should be a research priority.

CONCLUSION

Adjustment for nonadherence with the trial protocol increased the size of both the mortality benefit and risk of overdiagnosis by up to 50%. Deattenuated estimates better represent the effects of screening mammography at an individual level, and are useful for developing numerical outcomes for use by clinicians and programmes to communicate risk and help women make informed decisions – a guiding principle of modern health care. This approach to the calculation of estimates of mortality benefit and overdiagnosis of screening mammography that have been adjusted for nonadherence is applicable to other cancer screening trials.

ACKNOWLEDGEMENTS

This work was supported by a Grant (No. 1074626) awarded to GJ from the National Health and Medical Research Council of Australia and a Grant (No. 633003) awarded to the Screening and Test Evaluation Programme from the National Health and Medical Research Council of Australia. GJ acknowledges the support she has received from The Centre for Values, Ethics and the Law in Medicine, The University of Sydney, Australia.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Alexander F, Anderson T, Brown H, Forrest A, Hepburn W, Kirkpatrick A, Muir B, Prescott R, Smith A (1999) 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* **353**(9168): 1903–1908.
- Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, Ljungberg O, Ranstam J, Sigfusson B (1988) Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* **297**(6654): 943–948.
- Andersson I, Janzon L (1997) Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. *J Natl Cancer Inst Monogr* **1997**(22): 63–67.
- Baker SG, Kramer BS, Lindeman KS (2016) Latent class instrumental variables: a clinical and biostatistical perspective. *Stat Med* **35**(1): 147–160.
- Baker SG, Kramer BS, Prorok PC (2002) Statistical issues in randomized trials of cancer screening. *BMC Med Res Methodol* **2**(1): 11.
- Baker SG, Lindeman KS (1994) The paired availability design: a proposal for evaluating epidural analgesia during labor. *Stat Med* **13**(21): 2269–2278.
- Baker SG, Prorok PC, Kramer BS (2014) Lead time and overdiagnosis. *J Natl Cancer Inst* **106**(12): dju346.
- Biller-Andorno N, Jüni P (2014) Abolishing mammography screening programs? A view from the Swiss Medical Board. *NEJM* **370**(21): 1965–1967.
- Bjurstam N, Björnelid L, Warwick J, Sala E, Duffy SW, Nyström L, Walker N, Cahlin E, Eriksson O, Hafström LO (2003) The Gothenburg breast screening trial. *Cancer* **97**(10): 2387–2396.
- Blamey R, Wilson A, Patnick J (2000) ABC of breast diseases: screening for breast cancer. *BMJ* **321**(7262): 689.
- Canadian Task Force on Preventive Health Care (2011) Recommendations on screening for breast cancer in average-risk women aged 40–74 years. *Can Med Assoc J* **183**(17): 1991–2001.
- Cuzick J, Edwards R, Segnan N (1997) Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med* **16**(9): 1017–1029.
- Demissie K, Mills OF, Rhoads GG (1998) Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J Clin Epidemiol* **51**(2): 81–91.
- Department of Health Australia (2016) BreastScreen Australia Policy. Vol. 2016.
- Depypere H, Desreux J, Pérez-López FR, Ceausu I, Erel CT, Lambrinoukaki I, Schenck-Gustafsson K, van der Schouw YT, Simoncini T, Tremollieres F (2014) EMAS position statement: individualized breast cancer screening versus population-based mammography screening programmes. *Maturitas* **79**(4): 481–486.
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* **7**(3): 177–188.
- Duffy SW, Yen AM-F, Chen TH-H, Chen SL-S, Chiu SY-H, Fan JJ-Y, Smith RA, Vitak B, Tabar L (2012) Long-term benefits of breast screening. *Breast Cancer Manage* **1**(1): 31–38.
- Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, Thomson R, Barratt A, Barry M, Bernstein S (2006) Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* **333**(7565): 417.

- Frisell J, Lidbrink E, Hellström L, Rutqvist L-E (1997) Followup after 11 years—update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat* **45**(3): 263–270.
- Glasziou P (1992) Meta-analysis adjusting for compliance: the example of screening for breast cancer. *J Clin Epidemiol* **45**(11): 1251–1256.
- Glasziou P, Houssami N (2011) The evidence base for breast cancer screening. *Prev Med* **53**(3): 100–102.
- Gøtzsche PC, Jørgensen KJ (2013) Screening for breast cancer with mammography. *Cochrane Database Syst Rev* (6): CD001877.
- Gøtzsche PC, Nielsen M (2009) Screening for breast cancer with mammography. *Cochrane Database Syst Rev* (4): CD001877.
- Hersch J, Barratt A, Jansen J, Irwig L, McGeechan K, Jacklyn G, Thornton H, Dhillon H, Houssami N, McCaffery K (2015) Use of a decision aid including information on overdiagnosis to support informed choice about breast cancer screening: a randomised controlled trial. *Lancet* **385**(9978): 1642–1652.
- Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E (2007) Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev* (1): CD001216.
- Higgins J, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**(11): 1539–1558.
- Hollis S, Campbell F (1999) What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* **319**(7211): 670–674.
- Humphrey LL, Helfand M, Chan BK, Woolf SH (2002) Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* **137**(5_Part_1): 347–360.
- IARC (2002) Breast cancer screening. In *IARC Handbook of Cancer Prevention*, Vainio H, Bianchini F (eds). Vol. 7. IARC Press: Lyon.
- Ilic D, Neuberger MM, Djulbegovic M, Dahm P (2013) Screening for prostate cancer. *Cochrane Database Syst Rev*. (1) CD004720.
- Jørgensen K, Brodersen J, Hartling O, Nielsen M, Gøtzsche P (2009) Informed choice requires information about both benefits and harms. *J Med Ethics* **35**(4): 268–269.
- Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL (1995) Efficacy of screening mammography: a meta-analysis. *JAMA* **273**(2): 149–154.
- Marmot M, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M (2013) The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* **108**(11): 2205–2240.
- McIntosh MW (1999) Instrumental variables when evaluating screening trials: estimating the benefit of detecting cancer by screening. *Stat Med* **18**(20): 2775–2794.
- Miller AB, To T, Baines CJ, Wall C (2000) Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst* **92**(18): 1490–1499.
- Miller AB, To T, Baines CJ, Wall C (2002) The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* **137**(5_Part_1): 305–312.
- Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA (2014) Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* **348**: g366.
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devreux PJ, Elbourne D, Egger M, Altman DG (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**: c869.
- Moser K, Sellars S, Wheaton M, Cooke J, Duncan A, Maxwell A, Michell M, Wilson M, Beral V, Peto R (2011) Extending the age range for breast screening in England: pilot study to assess the feasibility and acceptability of randomization. *J Med Screen* **18**(2): 96–102.
- Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L (2006) Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* **368**(9552): 2053–2060.
- National Cancer Institute (1978) NIH/NCI Consensus Development Meeting on breast cancer screening. Issues and recommendations. *Yale J Biol Med* **51**(1): 3–7.
- Newcombe R (1988) Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Stat Med* **7**(11): 1179–1186.
- Nyström L, Andersson I, Bjurstram N, Frisell J, Nordenskjöld B, Rutqvist LE (2002) Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* **359**(9310): 909–919.
- Pace LE, Keating NL (2014) A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA* **311**(13): 1327–1335.
- Peto R, Pike M, Armitage P, Breslow N, Cox D, Howard S, Mantel N, McPherson K, Peto J, Smith P (1976) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *Br J Cancer* **34**: 585–612.
- Shapiro S, Venet W, Strax P, Venet L, Roeser R (1982) Ten-to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* **69**(2): 349–355.
- Shapiro S, Venet W, Strax P, Venet L, Roeser R (1985) Selection, follow-up, and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. *Natl Cancer Inst Monogr* **67**: 65–74.
- Siu AL (2016) Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* **164**(4): 279–296.
- Sommer A, Zeger SL (1991) On estimating efficacy from clinical trials. *Stat Med* **10**: 45–52.
- Tabár L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Gröntoft O, Ljungquist U, Lundström B, Månson JC, Eklund G (1985) 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* **1**(8433): 829–832.
- Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, Smith RA (1995) Efficacy of breast cancer screening by age. New results Swedish two-county trial. *Cancer* **75**(10): 2507–2517.
- Wald N, Chamberlain J, Hackshaw A (1993) Report of the European Society for Mastology breast cancer screening evaluation committee (1993). *Breast* **2**(4): 209–216.
- Welch HG, Black WC (2010) Overdiagnosis in cancer. *J Natl Cancer Inst* **102**(9): 605–613.
- Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP (2006) Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* **332**(7543): 689–692.

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 4.0 Unported License.

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