



# Evaluation issues in the Swedish Two-County Trial of breast cancer screening: An historical review

*J Med Screen*  
2017, Vol. 24(1) 27–33  
© The Author(s) 2016   
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0969141316631375  
journals.sagepub.com/home/msc  


Laszlo Tabar<sup>1</sup>, Tony Hsiu-Hsi Chen<sup>2</sup>, Chen-Yang Hsu<sup>2</sup>,  
Wendy Yi-Ying Wu<sup>2</sup>, Amy Ming-Fang Yen<sup>3</sup>, Sam Li-Sheng Chen<sup>3</sup>,  
Sherry Yueh-Hsia Chiu<sup>4</sup>, Jean Ching-Yuan Fann<sup>5</sup>, Kerri Beckmann<sup>6</sup>,  
Robert A Smith<sup>7</sup> and Stephen W Duffy<sup>8</sup>

## Abstract

**Objectives:** To summarize debate and research in the Swedish Two-County Trial of mammographic screening on key issues of trial design, endpoint evaluation, and overdiagnosis, and from these to infer promising directions for the future.

**Methods:** A cluster-randomized controlled trial of the offer of breast cancer screening in Sweden, with a single screen of the control group at the end of the screening phase forms the setting for a historical review of investigations and debate on issues of design, analysis, and interpretation of results of the trial.

**Results:** There has been considerable commentary on the closure screen of the control group, ascertainment of cause of death, and cluster randomization. The issues raised were researched in detail and the main questions answered in publications between 1989 and 2003. Overdiagnosis issues still remain, but methods of estimation taking full account of lead time and of non-screening influences on incidence (taking place mainly before 2005) suggest that it is a minor phenomenon.

**Conclusion:** Despite resolution of issues relating to this trial in peer-reviewed publications dating from years, or even decades ago, issues that already have been addressed continue to be raised. We suggest that it would be more profitable to concentrate efforts on current research issues in breast cancer diagnosis, treatment, and prevention.

## Keywords

Breast cancer screening, mammography, randomised trial

Date received: 9 December 2015; accepted: 18 January 2016

## Introduction

The Swedish Two-County Trial of mammographic screening for breast cancer was initiated in 1977, with 77,080 women aged 40–74 invited to screening (Active Study Population, i.e. ASP) and 55,985 women not invited (Passive Study Population, i.e. PSP). After approximately three rounds of screening the PSP was invited to screening for the first time, and the trial closed thereafter.<sup>1</sup> The first mortality results were published in 1985, showing a significant 31% reduction in breast cancer mortality associated with an invitation to screening.<sup>2</sup> In the years that followed, a number of methodological and clinical issues related to the Two-County Trial arose, and findings were investigated and scrutinized. Many of these issues arose as a result of challenges raised by sceptics, including concern about overdiagnosis,<sup>3</sup> trial design aspects, including cluster randomization and the closure screen of the control group, and allegations of bias in classification of cause of death.<sup>4–6</sup> In this paper, we summarize the issues and

concerns raised, then review the research from the Two-County Trial and elsewhere that has addressed and responded to these critiques. We conclude by summarizing the current state of knowledge, and giving some implications for future research and its conduct.

<sup>1</sup>University of Uppsala, Uppsala, Sweden

<sup>2</sup>National Taiwan University, Taipei, Taiwan

<sup>3</sup>Taipei Medical University, Taipei, Taiwan

<sup>4</sup>Chang Gung University, Tao-Yuan, Taiwan

<sup>5</sup>Kainan University, Tao-Yuan, Taiwan

<sup>6</sup>University of South Australia, Adelaide, Australia

<sup>7</sup>American Cancer Society, Atlanta, GA, USA

<sup>8</sup>Queen Mary University of London, London, UK

## Corresponding author:

SW Duffy, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK.

Email: s.w.duffy@qmul.ac.uk

## Evaluation issues

### Trial design

Two major attributes of the Two-County Trial design have received critical attention: the cluster randomization, in which 45 geographic areas (rather than the 133,065 individual women) were randomized to invitation to screening or not; and the screen of the control group at the end of the screening phase of the trial.<sup>4-8</sup> The major criticism was that the cluster randomization could have led to imbalances between ASP and PSP, that could have biased the breast cancer mortality comparison.<sup>4,6</sup> In fact, the Two-County investigators had reported a small age imbalance between the ASP and PSP in 1989.<sup>9</sup> In addition, concern was expressed in 1998 that the screening of the PSP at closure of the trial had caused inclusion of deaths from cancers in the PSP whose counterparts in the ASP would not have been included,<sup>5</sup> thus inflating deaths in the control (PSP) group, with the net result that mammography screening would appear more beneficial.

### Cause of death ascertainment

The possibility that cause of death was differentially classified between the ASP and PSP has been raised.<sup>4,6-8,10</sup> The grounds for these allegations (different numbers of deaths reported in different publications) have been shown to be spurious, due to (i) the inevitable differences between numbers of breast cancer deaths reported from the original trial endpoint committee classification of cause of death and those determined by the Swedish overview committee in a separate analysis of all the Swedish trials,<sup>11,12</sup> and (ii) the also inevitable increase in numbers of deaths with time during the follow-up period. Naturally, successive publications with longer follow-up have larger numbers of deaths.<sup>12</sup> Despite these explanations, the allegations have persisted.<sup>8</sup> It also has been argued that the appropriate endpoint should be all-cause mortality, which is described as free from bias in cause of death ascertainment. The call for all-cause mortality ignores the fact that an intervention for a single cause of death cannot be expected to prevent death from all causes. It also either overlooks the sample size requirements to measure a statistically significant difference in all-cause deaths associated with a disease-specific intervention, or judges this not to be a problem. One commentator has argued that to evaluate breast screening using all-cause mortality, 'a trial of 1.2 million women would suffice', presumably (emphasis on "suffice") regarding this as a readily achievable figure.<sup>7</sup>

### Overdiagnosis

Overdiagnosis, of both invasive disease and ductal carcinoma in situ (DCIS) is increasingly cited as a serious risk associated with mammography screening programmes.<sup>13,14</sup> While these programmes aim to detect breast cancer early, in order to reduce mortality and morbidity from breast cancer, it is argued that they cause

significant harm by finding cancers that do not need to be found. Overdiagnosis is usually defined as the detection of cancers, through screening, that would never have developed symptoms, and therefore never would have been diagnosed or been a threat to life during the patient's life time, had she not participated in screening. It is argued that the diagnosis of non-progressive cancers results in unnecessary treatment and associated psychological effects that would not otherwise have been experienced in the absence of screening. Specifically in the Two-County Trial, Gøtzsche claimed an overdiagnosis rate of 33%.<sup>7</sup>

Theoretically, overdiagnosis could occur in one of several ways: when the preclinical, screen-detected cancer is progressive but the person dies prematurely of another cause before the time at which symptoms would have occurred; if the growth rate of a truly progressive cancer is not rapid enough to give rise to symptoms during the person's life time; if the cancer stops growing and becomes indolent for some reason; or due to regression of the cancer, as is claimed by some.<sup>15</sup> Of these four possibilities, it is challenging to measure the first three with confidence, and there is very little sound evidence of the fourth.<sup>16</sup>

Overdiagnosed cancers are pathologically confirmed cases, that is, they exhibit all the features by which pathologists classify cancer. They are distinct from "false positives" which are suspicious lesions that are found not to be malignant after further investigation. There are currently no markers that can distinguish overdiagnosed cases from those which would have become symptomatic in a woman's life time. Hence it is not possible to directly observe overdiagnosis. In essence, overdiagnosis is an epidemiological concept related to the apparent excess in breast cancer incidence that occurs following screening.

Estimating the level of overdiagnosis is complex. Methods used include:

- Comparison of cumulative incidence in long-term follow-up of populations exposed and unexposed to screening. This could be follow-up of the mammography screening trials in which the control group was not screened at closure of the trial or thereafter, or observational studies comparing breast cancer incidence with and without screening.<sup>17,18</sup>
- Estimation from disease progression models allowing heterogeneity in lead time. Overdiagnosed cancers can be thought of as an extreme of length bias cases or as an extreme of lead time, in which the latter applies to those whose lead time extends beyond the future expected lifetime of the patient.<sup>19</sup>

While it is impossible to know if an individual cancer is overdiagnosed, it is, by definition, confined to screen-detected cancer, and is expected that it is more prevalent in DCIS and in small, localized invasive cancers.<sup>20</sup> A major issue of debate is the necessity to distinguish the effect of lead time from that of overdiagnosis, that is, to avoid attributing excess incidence from cancers

diagnosed earlier than they would otherwise have been diagnosed from the excess incidence due to cancers that would not have been diagnosed at all without screening.<sup>13,21</sup> In estimation from observational studies it is also necessary to take account of trends in breast cancer incidence occurring independently of screening.<sup>18</sup> While it might be expected that these trends would be the same in a group exposed to screening as in one not exposed to screening, this is not guaranteed in a non-trial setting, and indeed, the changes in incidence over time are also confounded by lead-time.

## Research in response to evaluation issues

### Trial design

In any randomized trial, difficult decisions arise in relation to design. In particular, in the design of population-scale screening trials, practicability of different randomization strategies and dealing with issues of lead time are important issues. From early in the evolution of the Two-County Trial results, the issue of the potential influence of cluster randomization on outcomes has been taken seriously. It was observed in the first publication of the breast cancer mortality results that 'the excess variation resulting from randomization being at the community rather than the individual level was negligible',<sup>2</sup> however, it was the Two-County investigators themselves who first noted and published the age imbalance,<sup>9</sup> whereby the ASP was slightly older on average than the PSP. Concerns about cluster randomization always are more focused on the risk of biases that lead to type 1 errors, although in this instance, it was noted that if the age imbalance had any effect at all on the primary result, it would be expected to bias it *against* screening, as breast cancer incidence and mortality increase with age.<sup>1</sup>

From 1992, the Two-County Trial investigators repeatedly published conservative estimates to take account of the cluster randomization, and the results have remained virtually unchanged, i.e. a significant reduction in breast cancer mortality in the ASP, of the order of 30%.<sup>1,22,23</sup> Of particular interest is the analysis published in 2000 by Nixon et al.,<sup>22</sup> which used Bayesian hierarchical modelling to fit variation between and within clusters. Dr AB Miller, who has frequently expressed scepticism about mammography screening, wrote in 2004 that he found this analysis, 'particularly compelling in largely dealing with the cluster randomization issue'.<sup>24</sup>

Perhaps more pertinently, in 2003, we noted that the important issue was whether the cluster randomization had introduced an imbalance in underlying rates of breast cancer mortality between ASP and PSP.<sup>25</sup> We found no evidence of this, but nonetheless reanalyzed the results, adjusting for breast cancer mortality in the 10 years prior to the trial, and found the same result, a substantial and significant approximate 30% reduction in breast cancer mortality in the ASP. We also performed an analysis adjusting for prior breast cancer mortality in the

randomization clusters, that is, adjusting for a potential difference between ASP and PSP clusters in baseline breast cancer mortality, and still observed a significant 27% reduction in mortality in the ASP.<sup>25</sup> Two summary points are important. First, cluster randomization is a legitimate methodological strategy, although more challenging than individual randomization, due to the requirement that the clusters produce similar groups for comparison. The secondary analyses performed on these data do not support conjectures that the positive findings in the Two County Trial are in doubt due to alleged vagaries in the randomization.

In relation to the issue of closure screening in the PSP, the initial trial results published in 1985 pertained almost entirely to the period *before* invitation to screening of the PSP, and showed the same results as later analyses of longer follow-up, which include deaths from breast cancer arising during closure screening.<sup>2</sup> Thus, fears that the screening of the PSP was in some way anticonservative, have been shown for many years to be misplaced. Despite accumulation over the years of analyses, a recent paper<sup>26</sup> has again challenged the Two-County findings, based on the assumption that the cluster randomization and the closure screen of the PSP biased the results in favour of screening, with no mention of previous analyses excluding the closure screen, taking account of the cluster randomization, or methodological investigations that showed that the fears of bias were unwarranted.<sup>1,2,22,25</sup> The main thesis of this recent paper<sup>26</sup> is based on the incorrect assumption that, immediately after a period of screening, incidence and mortality should be equal to that in a population that has never been screened. In fact, screening in one period will reduce incidence and mortality in the years immediately following, largely due to the shift in the period of diagnosis (i.e. lead time), as has been observed in trials and in service screening.<sup>1,27,28</sup>

We recently published a paper reviewing methodological strategies in screening trial design, highlighting the degree to which all strategies with a fixed number of screening rounds and a post-screening follow-up period introduced different levels of bias against measuring the effect of screening. We found that a closure screen of the PSP gave the least biased result (albeit still conservative), and most closely approximated the approximately unbiased, but often impractical design, in which screening in the ASP and usual care for the PSP continued for a long period.<sup>29</sup>

### Cause of death ascertainment

In 1989, the first mortality update of the Swedish Two-County Study published the details of determination of cause of death.<sup>9</sup> Briefly, if there were clinical, histological, or autopsy evidence of distant metastases, and *no* clear evidence that the metastases originated from a non-breast primary tumour, breast cancer was recorded as the cause of death. If the patient died of another cause, but not from a proven non-breast malignancy, but had local or regional breast cancer, or an asymptomatic

distant metastasis present at death, then the death was recorded as from breast cancer. Otherwise, the death was recorded as from other causes. The classification as death from breast cancer of intercurrent disease death with breast cancer present is likely to have been conservative, as women in the ASP with asymptomatic breast cancer would be more likely to have the disease diagnosed due to the screening.

In the same paper, we investigated the issue of a potential difference in classification of cause of death between the ASP and PSP.<sup>9</sup> If such a difference existed, it would be manifested as a significant difference between ASP and PSP in rates of death from other causes among breast cancer cases, as only those with breast cancer diagnosed can be classified as having died from the disease. For example, if there were a subjective tendency to ascribe deaths in breast cancer cases in the ASP as from other causes, there would be an excess rate of death from other causes in breast cancers in the ASP. No such difference was observed. The analysis was repeated with further follow-up in 1992,<sup>1</sup> and again, no significant difference was observed.

The issue was revisited in 2002, when we noted again that only those with breast cancer can die from the disease, and used this fact to demonstrate a significant reduction in mortality among breast cancer cases in the ASP compared with the PSP without classifying cause of death.<sup>30</sup> Thus, the mortality benefit was established without reliance on cause of death classification. Indeed, we found a significant all-cause mortality reduction of 13–19%, depending on the method of lead time adjustment.<sup>30</sup> It is also worth noting that the Swedish overview published an excess mortality analysis that similarly avoided classifying cause of death, and found, essentially, the same result as when explicitly using breast cancer death as the endpoint.<sup>31</sup>

Finally, in collaboration with the Swedish Overview, the Two-County investigators compared the cause of death, as determined by the local endpoint committee, with cause of death as classified in the Swedish overview.<sup>32</sup> All disagreements were investigated to establish the cause of disagreement. The authors concluded: 'The vast majority of these pertained to a disagreement in inclusion/exclusion and not to disagreement in determination of cause of death.' For example, the original trial defined eligibility on the basis of year of birth, whereas the overview used exact age. The results vindicated the original trial endpoint committee's classification of cause of death. Despite this, in their paper on the closure screen of the control group, Autier et al. again raised this issue,<sup>26</sup> with no reference to publications providing evidence that the concern is misplaced.<sup>11,30,31</sup>

There is a minority view that all-cause mortality is an appropriate endpoint for cancer screening trials.<sup>4,7</sup> This view is mistaken, as can be seen from the following example. There is currently considerable interest in the possibility of ovarian cancer screening, and the ovarian cancer mortality results of the UK Trial have recently

been published.<sup>33</sup> Let us consider the implications of all-cause mortality as an endpoint in evaluation of ovarian cancer screening. Ovarian cancer is responsible for approximately 4% of all deaths in a typical middle-aged female population. Suppose that the effect of the offer of ovarian cancer screening were to reduce ovarian cancer mortality by 20%, without affecting deaths from other causes. In a very large trial with 100,000 all-cause deaths expected in the control group, the expected number of deaths in the study group would be  $99,200 (0.04 \times 0.2 \times 100,000 = 800 \text{ fewer ovarian cancer deaths})$ . Thus the all-cause mortality relative risk would be 0.992 with a 95% confidence interval (CI) = 0.9834–1.0008. That is, even with 100,000 expected all cause deaths in each arm, the variation in mortality from other causes in this example completely swamps the beneficial effect of screening. A study with 300,000 all cause deaths expected in each arm would, arguably, be powered for this effect, but it means that to evaluate ovarian cancer screening with an all-cause mortality endpoint, a trial with 12 million women is needed, 6 million in each arm, and follow-up such that 5% in each arm die from any cause. This demonstrates that all-cause mortality is an impractical, inefficient, and unaffordable endpoint in a trial of an intervention which can only be expected to affect a single (minority) cause. Thus, the effect of ovarian cancer screening on all-cause mortality is essentially unverifiable. The answer is surely to have death from ovarian cancer, or the sequelae of screening or treatment for ovarian cancer, as the endpoint, and to adopt very rigorous cause of death determination policies, with a high rate of autopsy if necessary.

The use of all-cause mortality is sometimes advocated on the grounds of objectivity. While human judgement is required in all areas of medicine and health, it might be observed that its use distrusts and discards so many things that we know. These include the fact that only those with a cancer can die of it, only those irradiated can have a radiation-induced disorder, and so on. The way to effective evaluation is to use our knowledge when designing and running trials, not to throw it away on the assumption that proper safeguards cannot be implemented and regularly reviewed.

It is worth noting that, if a statistically significant effect on all-cause mortality (in those with and without the relevant disease) were the universal criterion by which medical interventions were assessed, in all likelihood, there are very few preventive public health interventions that would satisfy this criterion.

### Overdiagnosis

In 2003, we published estimates from the Two-County Trial, and from a number of service screening sources, of the division of DCIS into progressive and non-progressive disease.<sup>34</sup> Results indicated that the majority of DCIS cases would have progressed to invasive disease if left untreated.

In 2005, in collaboration with the Gothenburg breast screening trial, the Two-County Trialists published estimates of overdiagnosis using the observed data in the trials to estimate rates of diagnosis of non-progressive (i.e. overdiagnosed) cancers, adjusting for progression rates and screening sensitivity in non-overdiagnosed cancers.<sup>19</sup> The latter estimates separate overdiagnosis from additional incidence due to lead time, while ensuring that overdiagnosed cases are not included in the estimation of lead time. This work and its results addressed the overdiagnosis questions previously raised,<sup>3,7</sup> but despite this, some authors have continued to assert these same critiques, i.e. that screening in the Two County Trial resulted in substantial overdiagnosis.<sup>13,35</sup> We estimated that 3–4% of cancers diagnosed at first screen, and less than 1% at subsequent screens, were non-progressive. In terms of DCIS, an excess incidence of around one per thousand recruits in the ASP was balanced by a deficit of similar size in invasive cancers, strongly suggesting that diagnosis of DCIS is preventing subsequent diagnosis of invasive breast cancer. Results from both trials were very similar.

We revisited the issue using 29-year incidence data from one county of the Trial in 2012.<sup>36</sup> At the end of the 29 years, no excess incidence was observed in the ASP over the PSP, despite screening having started some years earlier in the ASP, and the latter having had around 100,000 more screening episodes than the PSP.

All the above suggest that overdiagnosis, and in particular non-progressive breast cancer, is a rare phenomenon. The UK Independent Review arrived at larger estimates from follow-up of the Canadian and Malmö trials, which had reported no screening of the control groups, and for which excess long-term incidence in the screened groups might be interpreted as overdiagnosis.<sup>17</sup> However, the UK review did not include the HIP trial, which also did not screen the control group, and which showed only a very small excess incidence at long-term follow-up.<sup>37</sup> In the Malmö trial, screening continued in the study group beyond the nominal screening age, and in those for whom screening stopped at age 70, and the excess incidence was only 1%.<sup>38</sup>

It should be acknowledged that much of the focus of overdiagnosis research and discussion in recent years has been on its presence in service screening with mammography, rather than in the trials.<sup>13–15,18,20,21</sup> It should also be noted that, given the unobservable nature of overdiagnosis, there is still scope for further quantification with long-term follow-up of current screening programmes. The point that adjusting for lead time based on all cases may result in over-adjustment, and hence underestimation of overdiagnosis, is fair.<sup>13</sup> However, the fact remains that when this point is honoured in analysis, and when adjustments are made for lead time and underlying incidence trends independent of screening, the resulting estimates of overdiagnosis are modest, at worst.<sup>19,34,39,40</sup>

## Discussion and implications for future research

This paper, which formally inaugurates the Swedish Two-County Research Group after many years of productive collaboration, concentrates on three issues which have been a particular focus for published concerns about the validity of the Two County trial, and on the trialists' willingness to meet these concerns head on. We note in passing that, in addition to this work, the trialists have carried out a considerable programme of downstream research into tumour biology, progression, and natural history.<sup>41–49</sup> Our research will continue in this direction, to further inform prevention, diagnosis, and management of breast cancer.

Some general points can be made in relation to the subjects dealt with above. First, the concerns about design and analysis issues in the Swedish Two-County Trial have been answered repeatedly in the past, and there seems little point in further argument, or at the very least, argument that does not describe in detail why published explanations are insufficient to rule out these concerns.<sup>1,9,40,24,25</sup> Secondly, the arguments about cause of death determination seem similarly pointless, particularly in view of the observation above that all-cause mortality is essentially useless in evaluation of screening for most cancers, and indeed for most health interventions.

As a case study of reviving issues long after their resolution, perhaps the most egregious example is the recent paper by Autier et al.,<sup>26</sup> which raises again issues relating to the design feature of the closure screen of the PSP, potential imbalances between the ASP and PSP, and ascertainment of cause of death, with little acknowledgement that the questions posed have been repeatedly answered in previous publications, and makes a number of assumptions which a brief study of the literature would show are unwarranted. Specifically:

- The authors assert, 'During post-intervention periods, because screening (or absence of screening) activities are similar in the screening and in the control group, cancer detection rates in the two groups are also similar.' This argument forms the basis for the claim that the closure screen of the PSP has biased the results in favour of screening, but is incorrect because in the screening group a large number of cancers will have been screen detected in the intervention period, which otherwise would have been detected later, in the post-intervention period. Cancer detection rates in the screening group will thus be lower than the rates in the control group in the post-intervention period, at least in the early years. Breast cancer mortality will, therefore, also be lower. As noted above, lower incidence after screening has been observed in trials and service screening programmes,<sup>1,27,28</sup> and we have shown that the closure screen of the PSP is slightly conservative.<sup>29</sup>
- Even if the argument of Autier et al. against the closure screen of the PSP were accepted, their adjusted

estimates are wrong, as they subtract the deaths from cancers diagnosed at that screen of the control group, but not those from corresponding cancers diagnosed contemporaneously in the study group. Of greater concern, the reduction in mortality from breast cancers prior to the closure screen of the control group has been in the public domain since 1985,<sup>2</sup> and was 31%, almost identical to that observed including the deaths from the closure screen and their counterparts in the study group.<sup>1</sup> The argument that the smaller numbers of advanced cancers in the study group contemporaneously with the control group's closure screen somehow invalidates the design and analysis including such a closure screen again displays a fundamental misunderstanding of the influence of screening on incidence and mortality rates. The control group screen is a prevalent screen, whereas at the end of the screening phase, the study group is in incident screen mode. The appropriate comparison is with the prevalent screen of the study group, which shows very similar numbers to the control group closure screen.<sup>1</sup> This too has been in the public domain for decades.

- The paper asserts that there are significant imbalances in missing values between ASP and PSP. The authors claim "the histological grade of cancers found during the Two-County trial was unknown for 19% of patients in the control group vs. 10% in the screening group ( $p < 0.0001$ ). Lymph node status was missing for 5.0% of patients in the screening group and 7.3% of patients in the control group ( $p = 0.0396$ )." These figures are incorrect. We speculated that they may have been derived from an early publication, reporting results before all the pathology data had been collated, but this does not seem to be the case. Not only were we unable to verify these figures from the trial data, we cannot find them in the paper which Autier et al cite as the source.<sup>1</sup> The proportions with missing node status were 6% in the ASP and 7% in the PSP, and with missing grade were 7% and 9%, respectively (not significant in either case). No reference is made to two crucial papers exploring the cluster randomization and the possibility of bias.<sup>22,25</sup>
- The paper again raises the issue of potential bias in cause of death, without reference to work investigating this issue in detail and showing that the concerns were unwarranted.<sup>9,25,30,31</sup>

The above illustrates how easy it is to overlook the fact that speculative concerns about specific research projects have been raised and answered in the past. We do acknowledge that there is scope for further research on overdiagnosis, but we qualify this with two observations. First, methodologically sound estimates of overdiagnosis which take account of lead time and underlying incidence, and which avoid some of the mistakes of the past, yield estimates that are modest.<sup>18</sup> Second, it is worth noting that scepticism about the benefits of screening seems to be accompanied by credulity about extreme estimates of

the quantity of non-progressive or regressive disease,<sup>8,10</sup> despite there being very little evidence of such cases.<sup>11</sup>

In our view, the time has come to put aside arguing about the past, and concentrate on the important issues for breast cancer control today. These include development of early detection protocols which better serve the population with dense breast tissue, and a greater understanding of the risk factors and potential preventive actions for non-hormone dependent breast cancers. Also, now that breast cancer is typically detected at an early stage, there is a need to develop 'light-touch' patient communication, treatment, and management protocols for breast cancers of low risk to life. We should also work towards better biological and prognostic stratification of breast cancer, to improve prevention, early detection, and treatment of breast cancer in the future.

### Swedish Two-County research group

Laszlo Tabar (University of Uppsala, Sweden); Tony Hsiu-Hsi Chen, Chen-Yang Hsu, Wendy Yi-Ying Wu (National Taiwan University, Taipei, Taiwan); Amy Ming-Fang Yen, Sam Li-Sheng Chen (Taipei Medical University, Taipei, Taiwan); Sherry Yueh-Hsia Chiu (Chang Gung University, Tao-Yuan, Taiwan); Jean Ching-Yuan Fann (Kainan University, Tao-Yuan, Taiwan); Kerri Beckmann (University of South Australia, Adelaide, Australia); Robert A Smith (American Cancer Society, Atlanta, GA); Stephen W Duffy (Queen Mary University of London, UK).

### Acknowledgements

Thanks are due to the women who took part in the Two County Trial and the personnel of the Departments of Mammography in the two counties who carried out more than 200,000 mammography examinations with great skill and dedication.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the American Cancer Society through a gift from the Longaberger Company's Horizon of Hope Campaign®.

### References

1. Tabar L, Fagerberg G, Duffy SW, et al. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin N Am* 1992; 30: 187–210.
2. Tabar L, Fagerberg CJG, Gad A, 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; i: 829–832.
3. Zahl PH, Strand BH and Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004; 328: 921–924.
4. Gøtzsche PC and Olsen O. Is screening for breast cancer with mammography justified? *Lancet* 2000; 355: 129–134.
5. Berry DA. Benefits and risks of screening mammography for women in their forties: a statistical appraisal. *J Natl Cancer Inst* 1998; 90: 1431–1439.
6. Skrabanek P. The debate over mass mammography in Britain. The case against. *BMJ* 1988; 297: 971–972.

7. Gøtzsche PG. On the benefits and harms of screening for breast cancer. *Int J Epidemiol* 2004; 33: 56–64.
8. Gøtzsche PC and Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013; 6: CD001877.
9. Tabar L, Fagerberg G, Duffy SW, et al. The Swedish two-county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Commun Health* 1989; 43: 107–114.
10. Black WC, Haggstrom DA and Welch HG. All-cause mortality in randomized trials of cancer screening. *JNCI* 2002; 94: 167–173.
11. Nyström L, Larsson LG, Rutqvist LE, et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. *Acta Oncol* 1995; 34: 145–152.
12. Duffy SW, Tabar L and Smith RA. Screening for breast cancer with mammography. *Lancet* 2001; 358: 2166.
13. Zahl PH, Jørgensen KJ and Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. *Br J Cancer* 2013; 109: 2014–2019.
14. Bleyer A and Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *NEJM* 2012; 367: 1998–2005.
15. Zahl PH, Maehlen J and Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008; 168: 2311–2316.
16. Kopans DB, Smith RA and Duffy SW. Mammographic screening and “overdiagnosis”. *Radiology* 2011; 260: 616–620.
17. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; 380: 1778–1786.
18. Puliti D, Miccinesi G and Paci E. Overdiagnosis in breast cancer: design and methods of estimation in observational studies. *Preve Med* 2011; 53: 131–133.
19. Duffy SW, Agbaje O, Tabar L, et al. Estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005; 7: 258–265.
20. Paci E and Duffy SW. Overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005; 7: 266–270.
21. Duffy SW and Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res* 2013; 15: R41.
22. Nixon RM, Prevost TC, Duffy SW, et al. Some random effects models for the analysis of cluster randomized trials: application to the Swedish two-county trial of breast cancer screening. *J Epidemiol Biostat* 2000; 5: 349–358.
23. Tabar L, Vitak B, Chen THH, et al. Swedish Two-County Trial: impact of mammographic screening on breast cancer mortality during three decades. *Radiology* 2011; 260: 658–663.
24. Miller AB. Commentary: a defence of the Health Insurance Plan (HIP) study and the Canadian National Breast Screening Study (CNBSS). *Int J Epidemiol* 2004; 33: 64–65.
25. Duffy SW, Tabar L, Vitak B, et al. The Swedish Two-County Trial of mammographic screening: cluster randomisation and endpoint evaluation. *Ann Oncol* 2003; 39: 1746–1754.
26. Autier P, Boniol M, Smaus M, et al. Statistical analyses in Swedish randomised trials on mammography screening and in other randomised trials on cancer screening: a systematic review. *J Roy Soc Med* 2015; 108: 440–450.
27. Tabar L, Fagerberg G, Day NE, et al. What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer* 1987; 55: 547–551.
28. Bennett RL, Sellars SJ and Moss SM. Interval cancers in the NHS breast cancer screening programme. *Br J Cancer* 2011; 104: 571–577.
29. Duffy SW and Smith RA. A note on the design of cancer screening trials. *J Med Screen* 2015; 22: 65–68.
30. Tabar L, Duffy SW, Yen MF, et al. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an endpoint. *J Med Screen* 2002; 9: 159–162.
31. Larsson LG, Nyström L, Wall S, et al. The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. *J Med Screen* 1996; 3: 129–132.
32. Holmberg L, Duffy SW, Yen AMF, et al. Differences in endpoints between the Swedish W-E (two-county) trial of mammographic screening and the Swedish overview: methodological consequences. *J Med Screen* 2009; 16: 73–80.
33. Jacobs IJ, Menon U, Ryan A, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet*. Epub ahead of print 2015. DOI: 10.1016/S0140-6736(15)01224-6.
34. Yen MF, Tabar L, Vitak B, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003; 39: 1746–1754.
35. Zahl PH, Jørgensen KJ and Gøtzsche PC. Lead time models should not be used to estimate overdiagnosis in cancer screening. *J Gen Intern Med* 2014; 29: 1283–1286.
36. Yen AM, Duffy SW, Chen TH, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. *Cancer* 2012; 118: 5728–5732.
37. Shapiro S, Venet W, Strax P, et al. *Periodic screening for breast cancer: The Health Insurance Plan Project and its sequelae, 1963–1986*. Baltimore: Johns Hopkins University Press, 1988.
38. Njor SH, Garne JP and Lynge E. Over-diagnosis estimate from The Independent UK Panel on Breast Cancer Screening is based on unsuitable data. *J Med Screen* 2013; 20: 104–105.
39. Beckmann KR, Lynch JW, Hiller JE, et al. A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening. *Int J Cancer* 2015; 136: 1411–1421.
40. Beckmann K, Duffy SW, Lynch J, et al. Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects. *J Med Screen* 2015; 22: 127–135.
41. Tabar L, Chen HH, Duffy SW, et al. A novel method for prediction of long-term outcome of women with T1a, T1b and 10–14mm invasive breast cancer: a prospective study. *Lancet* 2000; 355: 429–433.
42. Tabar L, Chen HHT, Yen MFA, et al. Mammographic tumor features can reliably predict the long-term outcome of women with 1–14 mm invasive breast cancer: suggestions for revision of current therapeutic practice and the TNM classification system. *Cancer* 2004; 101: 1745–1759.
43. Chiu SY, Duffy SW, Yen AM, et al. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol Biomarkers Prev* 2010; 19: 1219–1228.
44. Tabar L, Dean PB, Chen SL, et al. Invasive lobular carcinoma of the breast: the use of radiological appearance to classify tumor subtypes for better prediction of long-term outcome. *J Clin Exp Pathol* 2014; 4. DOI:10.4172/2161-0681.1000179.
45. Chen THH, Kuo HS, Yen MF, et al. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* 2000; 56: 167–172.
46. Tabar L, Duffy SW, Vitak B, et al. The natural history of breast cancer: what have we learned from screening? *Cancer* 1999; 86: 449–462.
47. Chen HH, Duffy SW and Tabar L. A mover-stayer mixture of markov chain models for the assessment of dedifferentiation and tumour progression in breast cancer. *J Appl Stat* 1997; 24: 265–278.
48. Chen HH, Duffy SW, Tabar L, et al. Markov chain models for progression of breast cancer, part 1: tumour attributes and the preclinical screen-detectable phase. *J Epidemiol Biostat* 1997; 2: 9–23.
49. Chen HH, Duffy SW, Tabar L, et al. Markov chain models for progression of breast cancer, part 2: prediction of outcomes for different screening regimes. *J Epidemiol Biostat* 1997; 2: 25–35.