Why the results from the EUROSCREEN Working Group are false

Comment by

Peter C Gøtzsche, DrMedSci, MSc, MD
Professor and Director
Nordic Cochrane Centre

11 October 2012

On 14 September 2012, the EUROSCREEN Working Group published a series of papers in a supplement of Journal of Medical Screening, which included a summary paper, 'Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet' (J Med Screeen 2012;19 Suppl 1:5-13, DOI: 10.1258/jms.2012.012077). The summary paper received a lot of media attention and many people have asked me to post my critique of it on our centre's website. I have done that and shall briefly explain the major issues.

Choice of journal

The papers were published in a journal that is known for having published many seriously flawed articles, using wrong methods, incorrect assumptions, misleading graphs, and having internal contradictions in numbers. I have given examples of some of the worst violations of good scientific practice in my book about mammography screening from January 2012 (1) and elsewhere, e.g. in (2,3). In particular, we have explained in detail why statistician Stephen Duffy et al. got it wrong by a factor of 20-25 when they published their estimate of the balance between the benefit of screening and the level of overdiagnosis (1-3). A factor of 20-25 is an extreme exaggeration, almost unheard of in science, as it is an inflation of 1900-2400%.

Journal of Screening is not a prestigious journal; its impact factor is only 1.7, compared to 14.1 for the BMJ. In addition, supplements to journals often lack editorial control and the papers are usually not peer reviewed.

In contrast, the Nordic Cochrane Centre has published most of its research in high-impact journals, particularly in the BMJ. Its editor-in-chief, Fiona Godlee, said in an interview in the Independent in 2010 (1):

'The screening lobby thinks the BMJ has got a bee in its bonnet about screening. That is not the case - we follow the evidence. The Danish team [from The Nordic Cochrane Centre] do good work which is very thorough and of good quality. It is fair to say that we have not had the same quality of submissions from the other side. We would be delighted if someone came forward with a robust defence of the screening programme - I don't think they have done that.'

Conflicts of interest

The papers in the supplement issue were written by self-appointed experts who have published a number of the papers they evaluate. Most of them are either directly responsible for breast screening programmes, or their professional esteem are tied to the success of the programmes in other ways. Self-appointed expert groups tend to preferentially include members who agree on central issues. While this increases the chance of reaching consensus, it is more doubtful whether it increases the chance of reaching correct conclusions.

Unsurprisingly, this expert group found that their own methods and results were the most reliable ones. They repeated their criticism of papers from independent research groups that have been fully adressed and rejected (1), and failed to mention important criticisms of their own papers, which have not (1).

What is most interesting in the papers is not what they mention but what they don't mention. No mention is made of the convincing evidence from many countries that screening has not reduced the occurrence of large tumours (bigger than 20 mm) or of tumours in stages III and IV (4,5), although this means that screening cannot have an effect on breast cancer mortality. This omission says a lot about

conflicts of interest, as the rationale for breast screening when it was introduced 25 years ago was to reduce the occurrence of advanced cancers.

The authors also fail to mention the pretty strong evidence there is that many of the tumours detected at screening sessions are harmless overdiagnosed tumours (5-8), and that many would have regressed spontaneously if left alone without treatment (9,10).

Further, no mention is made of the indisputable evidence that breast screening increases the use of mastectomies because of overdiagnosis (6,11,12).

Problems with the methods

The authors' summary paper is not science. It is wishful thinking and involves totally inappropriate methods, including extrapolations far beyond the data, instead of telling us what has been observed in terms of a reduction in breast cancer mortality.

In their attempt at estimating the effect of screening, the authors accept results of case-control studies, despite previous consensus among breast cancer screening experts that such studies cannot say anything reliably about the effect of screening (13). The bias introduced by such studies is huge. This has been beautifully demonstrated by data from the Malmö randomised screening trial. When properly analysed as a randomised trial, the reduction in breast cancer mortality was 4%, but when analysed as a case-control study (comparing breast cancer mortality in attenders with non-attenders within the screening arm), the reduction in breast cancer mortality was 58% (14). The flaw is referred to as the 'healthy screenee effect,' and it is inexcusable, but telling, that the authors of the supplement papers include such meaningless data.

The most detailed systematic reviews that have been carried out of the randomised trials by independent researchers are the Cochrane reviews and the reviews prepared by the US Preventives Services Task Force. The two teams came up with a similar estimate, a 15-16% reduction in breast cancer mortality (6,15). In contrast, using flawed methods, the EUROSCREEN Working Group arrive at two estimates, a 38% and a 48% reduction. What is remarkable about these estimates is that they correspond fairly well to the declines in breast cancer mortality we have seen in many countries, which have occurred at the same time as highly effective adjuvant therapy was introduced. Thus, the EUROSCREEN Working Group seems to say that there has been no effect of adjuvant therapy.

Using the implausible estimates of the effect of screening, the EUROSCREEN Working Group performs a stunning act of hocus-pocus. Their starting point is the number of breast cancer deaths occurring in a 30-year period when the women are between 50 years and 79 years of age. The authors then calculate, using their flawed estimates of the effect of screening, that over 30 years, so many lives will be saved that it corresponds to saving 7-9 out of every 1000 women screened. The average of this, 8 lives, is an overestimate of a factor of 16, or 1500%, compared to one woman saved out of 2000 that follows directly from the estimate derived from the randomised trials (6,15; the calculation is very simple and is explained in (6)).

Now, what do we know about the effect when we screen for longer than the 10 years in the randomised trials? We know a good deal, actually. Denmark is unique in the world for observational studies of mammography screening because it has a concomitant non-screened control group throughout 17 years, where screening was only offered in about 20% of the country, measured as population size (16). We found that the decline in breast cancer mortality was 1% per year in women who could benefit from screening (ages 55–74 years) in the screening areas during the 10-year period when screening could have had an effect (1997–2006), whereas it was 2% per year in the non-screening areas. The decline was larger in women who were too young to benefit from screening (ages 35–55 years), namely 5% per year in the screened areas, and 6% per year in the non-screened areas in the same time period.

These observations fit very well with recent papers in the BMJ by Philippe Autier et al. These researchers have described the breast cancer mortality trends in 30 European countries and found that the declines in young, non-screened age groups had been almost twice those in screened age groups (17). Further, they found that there was no discernible difference in the declines when comparing neighbouring countries that had introduced screening 10-15 years apart, and there was no relation whatsoever between start of screening and the reduction in breast cancer mortality (18). The fall in breast cancer mortality was not only about the same in all countries; it was also about the same as that seen in the United States (19).

So, whatever the effect of screening was when the randomised trials were performed many years ago, it seems to have disappeared. This is due to much better treatments and because women attend a doctor much earlier today when they have noticed anything unusual in their breast. The data are clear, but the EUROSCREEN Working Group nevertheless wants us to believe the opposite has occurred:

'The above suggests that service screening effectiveness may be greater today[34] than at the time when the studies on which the present European balance sheet is based were conducted.'

Inappropriate quotations

When I had read the EUROSCREEN Working Group's summary article and browsed the reference list, I became highly surprised to see that some of our own papers were mentioned there, as I hadn't noticed the slightest hint at any quotations of our work in the text, apart from our first Cochrane review from 2001, which 'questioned most of the RCTs.' So I looked at the text again to find out how the most important papers in the reference list had been discussed in the text.

The authors' references 8 and 9 are both to our papers in the BMJ and they are both very important. They showed that there has been no effect of screening in Denmark (16), and that screening increases mastectomies in Norway (12). The authors did not say a word about our results but said:

'While population-based service screening programmes have continued to be implemented without substantial changes in screening policy, there is still discussion over its effectiveness.[7–10],'

Autier et al.'s two pivotal papers on breast cancer mortality in Europe (17,18) were also listed, as references 36 and 37. Both papers strongly suggest that screening has no effect, but again, the EUROSCREEN Working Group's summary article says nothing about this:

'There is increasing interest in the literature in evaluation of outcomes of service screening, but methodology across the studies is not uniform.[29,34,36–41],'

Finally, reference 40 in the EUROSCREEN Working Group's summary article is our devasting criticism of the UK leaflet, 'Breast Screening, the facts - or maybe not?' (20), which we published in the BMJ in 2009. But yet again, what the EUROSCREEN Working Group writes in the text is totally meaningless:

'Recently published balance sheets differ considerably with regard to the sources of information used for the estimates, the modality of presentation of the outcomes and the communication implications for decision-making.[29,38–40],'

Try to read this pompous statement again. What is the take-home message in this?

Overdiagnosis

The EUROSCREEN Working Group's summary article says about overdiagnosis:

'In the absence of over-diagnosis, the initial increase in breast cancer occurrence in the screened group would be fully compensated by a similar decrease in cancers among older age groups no longer offered screening - the so-called "compensatory drop".[20] The compensatory drop method requires that the screening programme has been running long enough to achieve a full adjustment for lead time.'

Ironically, our study of overdiagnosis in Denmark lives fully up to these requirements, as we had a control group for 17 years without screening (8). It was even us who used the term 'compensatory drop' and studied it in our papers on overdiagnosis (8,21) but the EUROSCREEN Working Group does not quote our studies. We found 33% overdiagnosis in Denmark after adjustment for a small compensatory drop. This was somewhat less than the 52% we found in our systematic review of countries with organised screening programmes (21), likely because of lower uptake, lower recall rates and lower detection rates of carcinoma in situ.

Although our study is the most ideal there is, according to their own criteria, the EUROSCREEN Working Group's summary article does not mention our paper in the text or in their reference list.

Conclusions

More than anything, what the EUROSCREEN Working Group has documented so clearly is the crucial necessity of an independent and honest assessment of the evidence for health care interventions, free from conflicts of interest.

From now on, whenever I meet a person who believes mammography screening is effective, I shall ask how it is possible for screening to work when it doesn't reduce the occurrence of advanced cancers, which even the strongest screening advocates acknowledge is the point with screening (22):

'The key feature of a successful mammographic screening program is a reduction in the incidence rate of advanced tumors.'

Unfortunately, what the screening advocates have done time and again is to provide misleading information about this (1). They say that there are now relatively fewer large tumours than before screening. Of course there is, as all the small overdiagnosed tumours have been added to the denominator when such a percentage is calculated. I have illustrated in my book about breast screening why this manipulation is highly misleading (1):

Imagine a town with a certain level of crime. You divide the crimes into serious ones and less serious ones. Over a period of time, the rate of serious crime increases by 20% and the rate of less serious crime increases by 40%. This is clearly a development for the worse. But although more people are exposed to serious crime and more people are exposed to less serious crime as well, a trickster would say that, as there are now relatively fewer cases of serious crime, the situation has improved.

Finally, allow me to say that our leaflet about mammography screening is available at our homepage, www.cochrane.dk. The Center for Medical Consumers in the United States has described it as 'the first honest mammography information for women written by health professionals.' We think this is the reason that volunteers have translated it so that it now exists in 14 languages, and soon in 16, including Arabic, Chinese and Urdu. We conclude in our leaflet that it 'no longer seems reasonable to attend for breast cancer screening. In fact, by avoiding going to screening, a woman will lower her risk of getting a breast cancer diagnosis.'

References

1 Gøtzsche PC. Mammography screening: truth, lies and controversy. London: Radcliffe Publishing; 2012.

2 Gøtzsche PC, Jørgensen KJ. The Breast Screening Programme and misinforming the public. J R Soc Med 2011;104(9):361-9.

3 Gøtzsche PC, Jørgensen KJ, Zahl PH. Breast screening: why estimates differ by a factor of 20-25. J Med Screen 2010;17(3):158-9.

4 Autier P, Boniol M, Middleton R, et al. Advanced breast cancer incidence following population-based mammographic screening. Ann Oncol 2011;22(8):1726-35.

5 Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med 2012;156(7):491-9.

6 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2009;4:CD001877.

7 Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 2009;339:b2587.

8 Jørgensen KJ, Zahl P-H, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark: a comparative study. BMC Women's Health 2009;9:36.

9 Zahl P- H, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168:2311–16.

10 Zahl PH, Gotzsche PC, Maehlen J. Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study. Lancet Oncol 2011;12(12):1118-24.

11 Jørgensen KJ, Keen JD, Gøtzsche PC. Is mammographic screening justifiable considering its substantial overdiagnosis rate and minor effect on mortality? Radiology 2011;260(3):621-7.

12 Suhrke P, Maehlen J, Schlichting E, et al. Effect of mammography screening on surgical treatment for breast cancer in Norway: comparative analysis of cancer registry data. BMJ 2011;343:d4692.

13 Vainio H, Bianchini F. IARC Handbooks of Cancer Prevention. Vol 7: Breast Cancer Screening. Lyon: IARC Press, 2002.

14 Jørgensen KJ. Flawed methods explain the effect of mammography screening in Nijmegen. Br J Cancer 2011;105:592-3.

15 Humphrey LL, Helfand M, Chan BK, et al. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137(5 Part 1):347–60.

16 Jørgensen KJ, Zahl P-H, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: comparative study. BMJ 2010;340:c1241.

17 Autier P, Boniol M, Gavin A, et al. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ2011;343:d4411.

18 Autier P, Boniol M, La Vecchia C, et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. BMJ 2010;341:c3620.

19 Bleyer A. US breast cancer mortality is consistent with European data. BMJ 2011;343:d5630.

20 Gøtzsche P, Hartling OJ, Nielsen M, Brodersen J, Jørgensen KJ. Breast screening: the facts - or maybe not. BMJ 2009;338:446-8.

21 Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 2009; 339: b2587.

22 Smith RA, Duffy SW, Gabe R, et al. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am 2004;42(5):793–806.