References

1. Gøtzsche PC, Jørgensen KJ, Zahl PH, Maehlen J. Why mammography screening has not lived up to expectations from the randomised trials. *Cancer Causes Control* 2012;23:15–21.
2. Baum M. Harms from breast cancer screening outweigh benefits if death caused by treatment is included. *BMJ* 2013;346:f385.
3. Saquib N, Saquib J, Ioannidis J. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol* 2015;44:264–77.
4. Gøtzsche PC. *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Health Care*. London: Radcliffe Publishing, 2013.
5. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med* 2013;11:106–15.
6. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013;6:CD001877.
7. Swiss Medical Board. *Systematisches Mammographie-Screening*. 2013. www.medicalboard.ch/fileadmin/docs/public/mb/Fachberichte/2013-12-15_Bericht_Mammographie_Final_Kurzfassung_e.pdf (date last accessed, 1 October 2014).
8. Krogsbøll LT, Jørgensen KJ, Grønhøj Larsen C, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease. *Cochrane Database Syst Rev* 2012;10:CD009009.
9. Jørgensen T, Jacobsen RK, Toft U, Aadahl M, Glümer C, Pisinger C. Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial. *BMJ* 2014;348:g3617.

International Journal of Epidemiology, 2015, 280–282
doi: 10.1093/ije/dyu269
Advance Access Publication Date: 15 January 2015

Commentary: Tempering expectations of screening: what is the most authoritative advice we can give, given the data that we have?

Paul Taylor

Institute of Health Informatics, University College London, 222 Euston Road, London NW1 2DA, UK.
E-mail: p.taylor@ucl.ac.uk

The most authoritative basis for supporting a medical intervention is a meta-analysis of all sufficiently rigorous relevant randomized controlled trials. In this issue Saquib, Saquib and Ioannidis present an unprecedentedly thorough survey of 9 meta-analyses and 48 trials representing the best available evidence for the effectiveness of a range of screening interventions.[1] Some of the evidence reviewed has been argued over before. In the case of breast cancer, probably

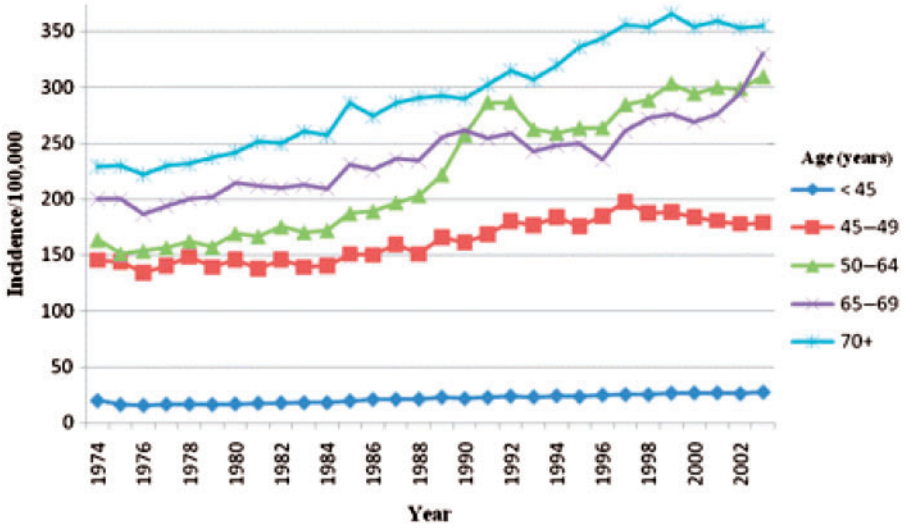

Figure 1. Incidence of breast cancer by age group in the UK from 1974 to 2004. (Reproduced from Duffy et al.[5]).

the most debated screening intervention, a series of large trials in the 1970s and 1980s provided what seemed to be clear evidence that screening saved lives, and countries across the developed world introduced programmes. Yet, in 2000, a meta-analysis concluded that there was no reliable evidence for breast screening.[2] The authors, Olsen and Gøtzsche, had identified eight trials but argued that the results of the six more or less favourable trials could not be trusted and that only the two more equivocal trials were sound.

The inclusion of one of the larger trials, the Swedish two-counties trial, should have been enough to reverse Olsen and Gøtzsche's conclusion.[3] They had argued that this trial had to be excluded in part because the average ages of the two groups were slightly different. The point is not that this difference—which was only 5 months—affected the results, but rather that any difference between two such large samples casts suspicion on the claim that they were randomly allocated. A further concern was that the investigators reviewing deaths among participants knew which arm of the trial a woman was in when they decided whether to count her death as caused by breast cancer death or not.

Whether or not these concerns are sufficient to warrant the exclusion of the trial is a matter of judgement and judgements, in this case, differed. Olsen and Gøtzsche were not the first or the last to attempt a meta-analysis of breast screening: there have been more reviews than there are trials to review. The arguments have been bitter, but have led towards consensus. Gøtzsche updated his analysis in 2011,[3] including more trials and finding overall support for the conclusion that screening reduces breast cancer deaths. An independent panel of UK experts, commissioned to look at the evidence, published a report in 2012 that drew on Gøtzsche's revised review to conclude that screening does reduce breast cancer deaths.[4] The United States Preventative Services Task Force has made a similar assessment.[6] Saquib, Saquib and Ioannidis, following Gøtzsche's updated analysis, give breast cancer as a case where screening reduces disease-specific mortality.[1]

But there's the rub. If breast cancer deaths are reduced, but all-cause mortality is unaffected, is this because detecting the latter requires that more statistical power be deployed? Or is it, as Gøtzsche has suggested, because the harms of screening increase deaths from other causes? The most serious cause of harm is overdiagnosis. The independent UK panel took the view that the best estimate of overdiagnosis could be provided by comparing the rates of cancer detection in the screened and the unscreened groups of randomized controlled trials. The problem is that when most trials ended, screening was offered to the women in the control groups, creating overdiagnosis in the follow-up period. The panel therefore restricted their attention to three trials in which no screening was offered to the control group during follow-up. This is a very limited set of data. Saquib, Saquib and Ioannidis ignore the question of harms presumably because there simply are not enough RCT data to review.

It is striking that almost all the patients screened in the reviewed trials that show a benefit due to screening, had their ultrasound, mammogram, sigmoidoscopy or faecal occult blood test in the past century, many of them in the 1970s and 1980s. For many cancers the benefits of early detection have been attenuated since then as a consequence of improvements in the treatment of late-stage disease. Trials of screening are expensive. Tens, sometimes hundreds, of thousands of participants are required and follow-up periods of 10 and 20 years are needed. Saquib, Saquib and Ioannidis's review lists only 48 trials. Restricting ourselves to this subset of the available data may be the best defence against methodological error, but in a changing world it clearly limits our capacity to base policy on relevant evidence.

Data other than those from trials could be used to provide evidence about the benefits and harms of screening.

The above graph, for example, shows a spike in the incidence of cancer in women of 50 to 64 years of age following the start of screening programme. We can use this to calculate how much overdiagnosis there is if we can estimate (i) the gradual increase in incidence observed before 1988—which presumably would have continued along the same trajectory had screening not been introduced—and (ii) the compensatory drop in incidence in older women who have been through screening. Unfortunately the aggregation of uncertainties in the calculation of these two figures means that wildly different estimates of overdiagnosis rates can be derived and indeed are derived.[6] We need a process similar to that which has allowed a degree of consensus to emerge on the validity of evidence from moderately flawed clinical trials, before we can use the data collected in the course of routine screening.

The abstract of this review[1] concludes: 'Among currently available screening tests for diseases where death is a common outcome, reductions in disease-specific mortality are uncommon and reductions in all-cause mortality are very rare or non-existent'. As I read it, 'uncommon' equates to 30% and 'very rare or non-existent' to 11%. The 30% figure is presented as disappointing. Perhaps it is, but remember that even an advocate of screening would expect a good proportion of trials to fail. One issue that is not discussed is the impact of our increasing capacity to stratify populations on the basis of risk. This should allow us to optimize screening programmes and improve outcomes. The cautious tempering of expectations advised by Saquib, Saquib and Ionnidis is prudent but should not be overdone.

Conflict of interest: None declared.

References

1. Saquib N, Saquib J, Ioannidis J. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol* 2015;44:264–77.
2. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129–34.
3. Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989;43:107–14.
4. Gøtzsche, Peter C, and Margrethe Nielsen. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2011;CD001877:4.
5. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380:1778–86.
6. Nelson H, Tyne K, Naik A, Bougatsos C, Chan B, Humphrey L. Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2009;151:727–37.
7. Duffy SW, McCann J, Godward S *et al*. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* 2010;17:25–30.
8. Duffy SW, Chen TH-H, Smith RA, Amy Ming-Fang Yen AM-F, Tabar L. Real and artificial controversies in breast cancer screening. *Breast Cancer Manag* 2013;2:519–28.