administered to large masses of the general population who do not exhibit any signs or symptoms of the disease. The purpose of screening is to detect in a preclinical phase the presence of a disease or precursor to a disease whose subsequent clinical course can then be ameliorated or even eliminated with treatment at that stage, in comparison with beginning treatment when the patient develops signs or symptoms of the disease (I am intentionally ignoring the issue of screening for genetic diseases, where a better understanding of prognosis in the absence of effective treatment can be considered an important outcome, as well as the prevention of a genetic disease in any offspring).

What needs to be considered then, are the outcomes of the disease for which screening is being contemplated. Although death is of course an important outcome, it is not the only outcome, and for some diseases may not even be the most important outcome. Many chronic diseases, such as heart failure, diabetes and chronic obstructive pulmonary disease (all three included in the set of diseases studied in this analysis) have numerous symptoms and outcomes other than mortality, such as dyspnoea, blindness, kidney failure and amputation. Even in the absence of any effect on mortality it is easy for me, as a primary care clinician, to imagine that patients would highly value any screening test and intervention that decreased the risk or severity of these outcomes. So I do not agree with the authors that, for these diseases, any screening test should be assessed with mortality as the main outcome. Whether these values are common among a broad community of patients deserves further study. Then there is the issue of patient preferences for different outcomes. Even if there is no effect on all-cause mortality, my clinical experience is that most patients would prefer some other cause of death to a death from cancer. Therefore for most diseases, I do not think that all-cause mortality should be considered the main outcome. Where the authors' data are most compelling is the evidence or lack thereof for a disease-specific effect on mortalities of cancers. Here, my clinical experience is that what patients are most concerned about is death from that cancer, be it lung, prostate, breast etc., and reducing the risk of that outcome is their paramount concern. It is hard for me to imagine having any enthusiasm for a screening test for cancer without convincing evidence that it would reduce disease-specific mortality.

The second issue I wish to comment on is what constitutes convincing evidence. The authors claim that this must come from randomized controlled trials with one group being offered screening and the other group not getting screened. For the most part, I agree with them. But there are exceptions. Cervical cancer screening has not been subjected to the kind of randomized controlled trial advocated by the authors, yet the observational evidence that mass screening programmes have had a beneficial effect is sufficiently strong to conclude that there is a cause-and-effect relationship. However, this is a historical issue, and I can agree with the authors that newly proposed tests should be subject to randomized trials assessing their benefits and harms.

In sum, the evidence synthesized by Drs Saquib, Saquib and Ioannides should be considered by anyone contemplating clinical practice guidelines about screening or proposing new screening tests. We have let too much get into routine practice without an adequate evaluation, and once widely disseminated, it can be very difficult to re-orient patient expectations and clinical behaviours to an understanding that a randomized trial comparing screening with no screening is ethically justified.

Conflict of interest: The author has received royalties for writing two chapters for *UpToDate*. In addition he receives an honorarium from ECRI for serving on the National Guidelines Clearing House and National Quality Clearing House (NGC/NQMC) Committee.

Commentary: Screening: a seductive paradigm that has generally failed us

International Journal of Epidemiology, 2015, 278–280
doi: 10.1093/ije/dyu267
Advance Access Publication Date: 15 January 2015

PC Gøtzsche

Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark. E-mail: pcg@cochrane.dk

Screening healthy people has face value and great public and political appeal. It looks so simple, and yet screening is fraught with difficulties. These start already with the terminology, and common slogans like, 'Catch the disease early, before it has produced any symptoms!' are misleading on two counts.

First, disease means lack of ease, which is not what we understand by being healthy; but people who work

with screening tend to forget that they deal with healthy people. For example, women being invited to mammography screening are often called patients in scientific articles.

The second error is the assumption that the disease is caught early. That is rarely the case, and breast cancer is again a good example. If we assume that the growth rate for a particular cancer is constant, then the women have harboured the cancer for 21 years on average before it is large enough to be detected by mammography screening.[1] Finding precursors to cancer is of course an entirely different matter. Screening with flexible sigmoidoscopy identifies polyps and vaginal smear finds carcinoma *in situ*.

A third problem with screening is that it always causes harm. Sometimes it also leads to benefits, and sometimes the benefits are sufficiently large to outweigh the harms. The main focus in screening trials should therefore be to quantify the harms, but this has rarely been the case, if ever. Screening trials focus on disease-specific mortality, which may seem natural, but it is the wrong outcome. Screening leads to overdiagnosis, and interventions that are beneficial for real patients can be lethal for healthy overdiagnosed people. Radiotherapy of overdiagnosed women may kill at least as many as those who are spared dying from breast cancer by attending breast screening.[2]

Total mortality should therefore be the primary outcome in screening trials of mortality, and Saquib *et al.* report a systematic review in this issue of the journal that aimed at clarifying whether screening lowers total mortality for diseases that carry a high disease-specific mortality.[3] They focused on cancer, cardiovascular diseases, type 2 diabetes and chronic obstructive pulmonary disease. They did not find any screening trials for hypertension or chronic obstructive pulmonary disease. Disease-specific mortality was reduced with ultrasound for abdominal aortic aneurysm in men, mammography for breast cancer and faecal occult blood test and flexible sigmoidoscopy for colorectal cancer, but the risk ratio point estimates for all-cause mortality were all very close to 1.00 (range 0.98–1.03).

Screening proponents often say that disease-specific mortality is the right outcome, arguing that in order to show an effect on total mortality, trials would become unrealistically large. I believe this argument is invalid, for both scientific and ethical reasons. We do randomized trials in order to avoid bias, and our primary outcome should therefore not be a biased one. Drug interventions are usually more common in a screened group, and they tend to increase mortality for a variety of non-disease related reasons.[4]

From an ethical perspective, it is problematic to screen the whole population in a certain age group without knowing whether this makes people live longer, while knowing almost certainly that it makes people less happy. It took 50 years after the first randomized trial of mammography started before we knew what the psychological consequences are of the many false-positive findings.[5] A specially designed questionnaire was developed using focus groups and women who had attended screening were followed up for 3 years. Even after so long a time, those who had experienced a false-positive diagnosis had an anxiety level (and other psychological problems) that fell between that for women with breast cancer and women who had always been told they did not have cancer. This study showed for the first time that the psychological harms of breast screening are substantial and long-lasting, and they affect a huge number of healthy women, as the cumulative risk of a false-positive result after 10 mammograms ranges from about 20% to 60%.[6] Added to this comes the psychological harm inflicted on all the overdiagnosed women who do not know that they are overdiagnosed but think that they suffer from a fatal disease. It is therefore pretty clear that any utility analysis that takes the psychological harms of breast screening into account will come out negative, as was recently reported by the Swiss Medical Board.[7]

Saquib *et al.* found no screening trials for hypertension and only one for diabetes, ADDITION-Cambridge, for which the risk ratio for all-cause mortality was 1.06. In our systematic review of general health checks,[8] 7 of the 16 trials screened for diabetes, and likely all of them screened for hypertension (in one, the screening tests were not specified). Although we had 11 940 deaths, we did not find an effect on total mortality (risk ratio 0.99, 95% confidence interval 0.95 to 1.03). We could not include the most recent trial, as it was published in 2014.[9] It investigated the effect of systematic screening for risk factors for ischaemic heart disease and lifestyle counselling. This trial was large but it also failed to find an effect on total mortality: 3163deaths occurred, and the hazard ratio was 1.00 (0.91 to 1.09).

It is worth noting that when screening does not work, it might be because beneficial effects are outweighed by harmful ones. Diabetes drugs, for example, are approved on the basis of their glucose-lowering effect without knowing what they do to patients. And the only large trial of tolbutamide ever performed was stopped prematurely because the drug increased cardiovascular mortality.[4] Rosiglitazone was once the most-sold diabetes drug in the world, but it was taken off the market in Europe in 2010 as it causes myocardial infarction and cardiovascular death; and pioglitazone has been linked to heart failure and bladder cancer.[4]

Screening is popular, but we need to be much more careful in the future when we contemplate approaching healthy people with our screening tests, and should demand much stronger evidence than when we treat patients.

Conflict of interest: None declared.

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

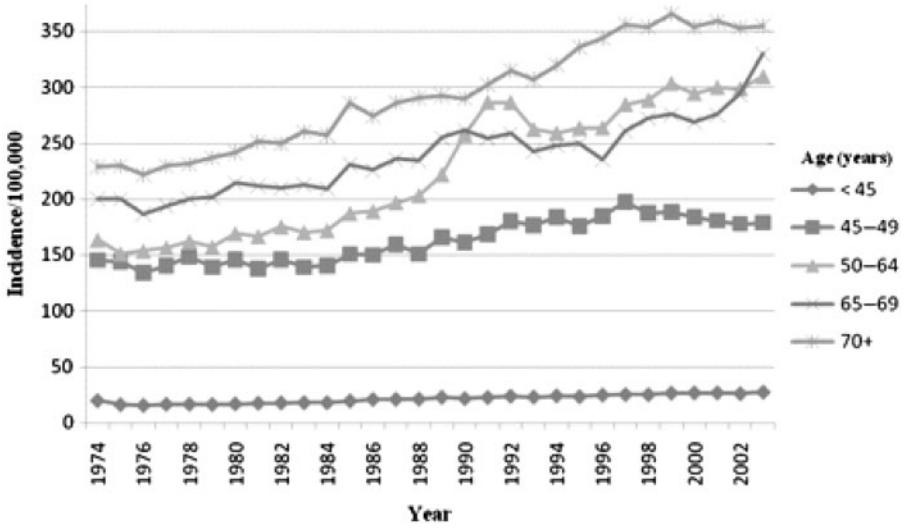

Figure 1. Incidence of breast cancer by age group in the UK from 1974 to 2004. (Reproduced from Duffy et al.[5]).