

Comparative effectiveness research in cancer screening programmes

Michael Bretthauer,^{1,2} Geir Hoff^{2,3,4}

¹University of Oslo, Oslo, Norway

²Oslo University Hospital, Oslo

³Cancer Registry of Norway, Oslo

⁴Department of Research, Telemark Hospital, Skien, Norway

Correspondence to: M Bretthauer, Institute of Health and Society, Department of Health Management and Health Economy, University of Oslo, 0317 Oslo, Norway michael.bretthauer@medisin.uio.no

Accepted: 26 March 2012

Cite this as: *BMJ* 2012;344:e2864
doi: 10.1136/bmj.e2864

Large scale cancer screening programmes are not amenable to generating or responding to new evidence about their effectiveness. The authors outline a new approach in Norway intended to overcome these drawbacks by means of comparative effectiveness research

In recent decades, cancer screening programmes (screening that is publicly organised and includes invitation procedures for eligible people of the average risk population in the screening area) have been established in many countries. While cancer screening in the context of clinical trials is innovative and investigative, cancer screening programmes themselves are largely static and not designed to generate new, evidence based knowledge. However, screening programmes themselves affect health and healthcare, which may in turn substantially affect the effectiveness of the programmes. Many screening programmes today can be regarded as supertankers; once under way, they are difficult to halt or alter in direction or content. In mammography, increasing concern about the benefits and harms of screening programmes has led to the announcement of an independent review of mammography screening in the United Kingdom.¹ Here, we outline new approaches that aim to overcome this obstacle, using the principles of comparative effectiveness research.

Screening programmes change medicine

Breast cancer mortality in Norway has declined since the introduction of the Norwegian breast cancer screening programme in the 1990s. According to new evidence, however, most of the observed decline in mortality is not due to the screening itself but is because of the improved patient care that resulted from the introduction of the screening programme (which was accompanied by reorganisation

of breast cancer care, improving quality and awareness).² This surprising finding indicates that screening tests themselves may be of minor importance for the reduced morbidity and mortality achieved by implementing a screening programme.

In the case of the Norwegian breast cancer screening programme, it was possible to tease out the effect of the reorganised care from that of the screening itself only because the programme was introduced in stages, and control groups that were parallel in time with the intervention groups could be established. Such control groups (so called concurrent groups) are important for the evaluation of screening programmes because of changes in risk factors and improvements in diagnostics and treatment over time and socioeconomic imbalance between screening areas and non-screening areas. Most screening programmes, however, do not include concurrent control groups, which makes it difficult to obtain valid comparisons between individuals who are invited to screening and those who are not. This precludes a scientific evaluation of the effects of the programmes.

Comparative effectiveness research

Comparative effectiveness research is “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition, or to improve the delivery of care ... The purpose of [comparative effectiveness research] is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.”³ This approach has been used most often in the context of clinical trials comparing different therapies such as drugs or devices. However, it may also be applicable for other types of interventions, such as disease prevention and screening. Comparative effectiveness research can also be used to improve the quality and effectiveness of applied concepts of healthcare services, such as screening programmes for cancer.

Lack of evidence

Cancer screening trials are costly and difficult to perform compared with most clinical trials. There is a long lag time—often exceeding 10 years—between a screening intervention and the time when the most important effect of screening, disease specific mortality, can be observed. Because relatively few people die from cancer, screening trials have to include a large number of participants; and high compliance rates are difficult to obtain because screening trials approach presumptively healthy people who will have no immediate gain from participation (unlike trials among symptomatic patients). These challenges may account for the lack of scientific evidence for

SUMMARY POINTS

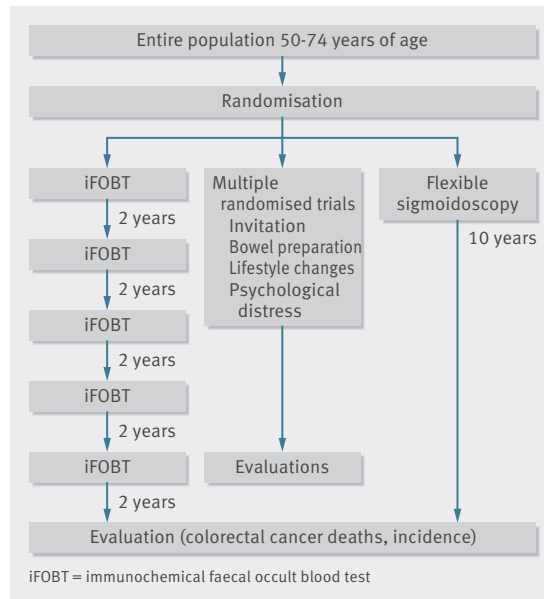
Cancer screening programmes are generally static and are not designed to include research studies

Cancer screening programmes change the environment they are operating in, and this may change the effectiveness of the programme

Comparative effectiveness research may be used to continuously optimise effectiveness of screening services within running programmes

In Norway, the entire population of two areas is being randomised within a new bowel cancer screening programme from 2012, to generate new evidence for effectiveness and harms of screening

Design of the comparative effectiveness research screening programme for colorectal cancer currently implemented in Norway. Screening by immunochemical faecal occult blood testing is offered biennially, whereas flexible sigmoidoscopy screening is once only. The time from start of screening until evaluation of the primary end point is 10 years



many of the large cancer screening programmes currently in place. Cervical smear screening for cervical cancer, introduced in the 1960s, has never been investigated in a randomised trial. Prostate specific antigen (PSA) screening for prostate cancer became endemic before any randomised trials were even started, with the consequence of contamination bias and controversy about the validity of the obtained results.⁴

However, the scientific evidence which screening programmes are based on should arise from randomised trials. To disentangle the different options of screening methods, intervals, thresholds for follow-up and surveillance, head to head randomised comparisons of the different screening options and tests available are needed, as recently emphasised by Baum in the *BMJ*.⁵ Screening programmes, rather than independent randomised trials, may be the natural platforms for these studies.

Colorectal cancer is one of the major causes of death from cancer worldwide.⁶ A range of screening tests exists—such as faecal occult blood testing, flexible sigmoidoscopy, and colonoscopy—but evidence on the performance of one test compared with the others is limited. Furthermore, the different tests may have different cost effectiveness profiles in different settings, populations, and cultures. Head to head comparisons are needed to evaluate the efficacy, effectiveness, and cost effectiveness of the tests. Best clinical practice changes with time. A good health service programme should integrate development and testing of new modalities as part of the programme itself. The introduction of organised screening programmes for colorectal cancer is an ideal opportunity to establish comparative effectiveness research. Randomised comparisons of different screening strategies are the best means of obtaining valid, high quality data.

The Norwegian approach

Norway has one of the highest incidences of colorectal cancer worldwide. In 2009 the Norwegian Directorate of Health established a national board of experts to advise the government about colorectal cancer screening. The

board recognised an imminent need to control the burden of colorectal cancer and the part screening might play in achieving this. However, in light of the ongoing harsh debate about other cancer screening programmes (particularly mammography) in Norway and other countries, a strong recommendation was given to evaluate the possibility of integrating high quality research into the colorectal cancer screening programme. The Norwegian Centre of Knowledge in Health Care, an independent publicly owned research institute, was asked to produce a report on the current evidence for the different colorectal cancer screening tests. As a prerequisite, the Norwegian government requested that only screening modalities that had been shown to be effective compared with no screening in randomised trials should be considered for inclusion in a future screening programme. The report, available in 2010, concluded that two screening tests have been proved to reduce colorectal cancer mortality in randomised trials (faecal occult blood test and flexible sigmoidoscopy), but that there was insufficient evidence to show which was the better of the two.⁷ Further, there were substantial differences between the studies with regard to effect sizes for mortality and incidence, and with regard to population and test performance settings.

The board therefore concluded that a future national colorectal cancer screening programme would need to establish its own evidence on a continuous basis to be able to monitor the effectiveness and adverse effects of the screening programme. The board recommended the use of clinical trial methodology, through randomised comparisons within comparative effectiveness research projects whenever possible, combined with rigorous data gathering and evaluation. On the advice of the board, the Norwegian government established a national comparative effectiveness screening programme for colorectal cancer. This programme's aim is to generate evidence on the comparative effectiveness of the screening tests and strategies available.⁸

The entire population aged 50–74 years in two geographical areas of Norway will be randomised to one of two screening options: half to biennial screening by immunochemical faecal occult blood test, and half to once only screening by flexible sigmoidoscopy (figure). The primary evaluation end point is the comparison of the two tests with regard to mortality from colorectal cancer after 10 years. Within the two primary comparison groups, additional randomised trials will be performed to evaluate other important measures such as anxiety and lifestyle changes due to screening, invitation procedures, appointment assignments, and bowel cleansing regimens. The entire screening programme will essentially be set up as a series of adaptive randomised trials, ensuring continuous evaluation of the effectiveness of the different measures compared by establishing concurrent comparison groups through randomisation.

This will result in a continuously updated programme that can rapidly take into account new, self generated evidence and integrate it into the programme. Also, when future screening tests show promising results in clinical trials (such as molecular markers), these new methods can be enrolled into the programme in a randomised fashion as

additional arms along the two established ones, and thus rigorously tested within the setting of a real life screening programme.

Although the randomised trial is the optimal method in clinical research, not all unanswered questions can be resolved by applying this method. Observational studies with case-control or cohort design or registry based observations are useful to evaluate health services and will be applied in the planned screening programme. For example, a system for continuous monitoring of all endoscopies in the programme will be set up. These observational data will be used to improve the quality of the service, and may also be used in research projects.

Comparative research in public health programmes

Cancer screening addresses large numbers of presumptively healthy individuals, who are subjected to medical interventions that are not risk-free, often cumbersome, and frequently result in false positive or false negative results. The vast majority of people invited to screening will never get the disease. For example, the lifetime risk for colorectal cancer in Western countries is around 5%. This means that 95% of the population will not get the disease irrespective of attending screening, and that screening only affects the course of the disease for the 5% who would get the disease if screening had not been an option. The others will not have any personal gain by participating in screening, but are prone to the adverse effects and the complications. The integration of randomised, head to head comparison trials within new or established screening programmes provides healthcare providers, funding bodies, and the public with population specific, updated, and reliable evidence on the effectiveness and risks of different options and strategies in cancer screening. There is no opposing interest between proponents of cancer screening programmes and those advocating randomised trials.

Some may argue that it is too late for many screening programmes to incorporate comparative research strategies. However, even in established programmes, such as for mammography or cervical cancer screening, many new concepts and innovations can be tested by means of comparative effectiveness research. Examples for comparison trials within the programmes would include evaluation of "watch and wait" versus radiotherapy for screen detected ductal carcinoma in situ (www.clinicaltrials.gov; NCT 00077168), or comparisons of different human papillomavirus tests in cervical cancer screening.

The establishment of comparative effectiveness screening programmes does not require significantly more resources compared with conventional programmes. Comparing two or more screening tools, different modes for invitation, or strategies for reminders and follow-up (such as telephone versus letters versus emails) are easily handled with most available information technology and database management systems. Many new methods to be tested are less resource demanding or improve compliance and thereby effectiveness. Therefore, incorporation of comparative effectiveness research will help to save costs and increase cost effectiveness, and is attractive also for developing countries with limited resources and a high level of uncertainty about the effect of public health programmes (due to the lack of own data).

The overarching aim of the new comparative effectiveness screening programme for colorectal cancer in Norway is to achieve continuous optimisation of the screening service. This includes comparisons of different screening tests to find the best test for the Norwegian population. The programme will produce comparative data on effectiveness, adverse events, side effects such as overdiagnosis, and costs. The programme, however, is not designed to evaluate the effectiveness of screening versus no screening because everyone is offered screening. Thus, we will not be able to find out if screening is effective, only help find the most effective of different screening options.

Contributors: MB and GH had the idea for the paper. MB drafted the first version of the manuscript. MB is a member of the Norwegian Directorate of Health national advisory board for colorectal cancer screening. GH is head of the Norwegian colorectal cancer screening programme.

Provenance and peer review: Not commissioned; peer reviewed.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

- 1 Richards M. An independent review is under way. *BMJ* 2011;343:d6843.
- 2 Kalager M, Zelen M, Langmark F, Adami HO. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 2010;363:1203-10.
- 3 Sox H. Better evidence about screening for lung cancer. *N Engl J Med* 2011;365:455-7.
- 4 Djulbegovic M, Beyth RJ, Neuberger MM, Stoffs TL, Vieweg J, Djulbegovic B, et al. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ* 2010;341:c4543.
- 5 Baum M. Screening for breast cancer: an appeal to Mike Richards. *BMJ* 2011;343:d7535.
- 6 Cancer Mondial. Globocan 2002 database. www-dep.iarc.fr (accessed 22 Aug 2011).
- 7 Fretheim A, Bretthauer M. *Expected impact of introducing screening for colorectal cancer in Norway: rapid review*. National Centre of Knowledge in Health Care, 2010.
- 8 Ministry of Health and Care Services. Press release: Kraftig satsing på helse og omsorg [in Norwegian]. www.regjeringen.no/nb/dep/hod/pressemeldinger/2010/kraftig-satsing-pa-helse-og-omsorg.html?id=620271.

Playing the odds

Twice now in about as many months she's led me, only half attending, into this trap.

First she tells me of a raffle ticket she bought for a chance to win a car. "Yes, you told me that," I say, "I hope you win." (See how I've forgotten where this story leads?)

Then she lays out the odds: "They're only selling 500 tickets... I think I have a pretty good chance, don't you?"

(Of course, this is where I should say, "No, actually, 1 in 500 may seem like good odds, but in fact it means you have almost no chance of winning the car.") I say, my mind a bit on other things, "You do. You might actually win. That would be great."

Then she tosses out part two: "And the chances that my breast cancer will come back are only 8 in 100. It's funny how that seems like it's pretty unlikely, doesn't it?"

Kathryn B Kirkland associate professor of medicine, Department of Medicine, Dartmouth Medical School; Center for Leadership and Improvement, Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School, Hanover NH, USA
Kathryn.b.kirkland@hitchcock.org
Patient consent obtained.

Cite this as: *BMJ* 2012;344:e3490