Maximising benefit and minimising harm of screening

J A M Gray, J Patnick and R G Blanks

Updated information and services can be found at:

http://bmj.com/cgi/content/full/336/7642/480

These include:

References	This article cites 9 articles, 2 of which can be accessed free at: **http://bmj.com/cgi/content/full/336/7642/480#BIBL**
Rapid responses	You can respond to this article at: **http://bmj.com/cgi/eletter-submit/336/7642/480**
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top left of the article

Topic collections	Articles on similar topics can be found in the following collections Cancer: breast (659 articles) Quality improvement (including CQI and TQM) (592 articles) Screening (730 articles)

Notes

To order reprints follow the "Request Permissions" link in the navigation box

To subscribe to *BMJ* go to:
http://resources.bmj.com/bmj/subscribers

Maximising benefit and minimising harm of screening

Gordon Brown has pledged to increase screening services in the NHS. **Muir Gray, Julietta Patnick,** and **Roger Blanks** show how experience with the UK breast screening programme can help ensure that they are effective

All screening programmes do harm; some do good as well, and, of these, some do more good than harm at reasonable cost. The first task of any public health service is to identify beneficial programmes by appraising the evidence. However, evidence of a favourable balance of benefit to harm in a research setting does not guarantee that a similar balance will be reproduced in practice, so screening programmes need to be introduced in a way that allows their quality to be measured and continuously improved.

The policy decision

The decision to implement a breast cancer screening programme in the United Kingdom in 1987 was based on evidence of efficacy from other countries, especially Sweden. The Department of Health set up a committee to review the evidence, chaired by Pat Forrest, a leading researcher in treating and screening for breast cancer. His committee's analysis of evidence and recommendations[1] has been criticised,[2] but the government accepted the review's conclusions and decided not only to introduce screening for breast cancer for women aged 50-64 but also to fund a project to ensure that screening was introduced in a way most likely to ensure that the benefits found in the Swedish trial[3] would be reproduced throughout the UK. Four crucial decisions were taken:
- To introduce a standardised screening system across the whole country
- To recommend a minimum size of screening unit to optimise quality and enable evaluation
- To fund quality assurance and training centrally and develop continuous quality improvement
- To fund an implementation project.

Implementation

A small implementation team was set up, with its director held accountable both to the Department of Health and to an advisory committee with representatives of

J A Muir Gray director, National Knowledge Service, Oxford OX3 7LG
J Patnick director, NHS Cancer Screening Programmes, Sheffield
R G Blanks epidemiologist, Institute of Cancer Research, Sutton, Surrey
Correspondence to: J A Muir Gray
muir.gray@medknox.net
Accepted: 13 September 2007

the professions and the women to be offered screening, to deliver the programme within a specific time and budget and to set out performance standards. To achieve its objectives the implementation team was given a discrete budget sufficient to fund the programme; authority to centralise certain aspects of screening, notably the multidisciplinary assessment of women with abnormal mammography results; and separate funds to set up four training centres, procure an information system, and prepare clear information for the women offered screening.

Each of the 14 regional authorities then in England appointed a lead professional to manage the implementation; similar arrangements were made in Scotland, Wales, and Northern Ireland. Strong links were forged with the private sector and coordination between regions was ensured by establishing groups, called Big 18s, for radiology, pathology, surgery, and all the other disciplines involved in screening. The groups comprised representatives from all 17 NHS organisations, with the 18th partner being the private sector.

Standardising screening systems

A national screening programme needs more than a common policy. Screening is a programme, not a test, and a programme is a set of activities with a common set of objectives. These have to be complemented by nationally agreed criteria and standards if benefits are to be maximised and harm minimised.

The original aim was to reduce mortality from breast cancer in the target population, but this was broken down into a set of specific objectives.[4] Each objective is best understood by defining the data that would need to be collected to measure progress towards it.[5] The criteria for each objective were selected with respect to their validity and their feasibility of collection—two variables that are usually inversely related.[6] Because no UK national data were available to set standards for screening, the first national standards were based on the data from Sweden tempered by the experience of leading professionals in the UK (table 1).

Criteria are objective, but standards are subjective, arbitrarily chosen to describe what Avedis Donabedian called the degree of "goodness" of a service or product—namely, its quality.[7] We could have adopted the Swedish standards, but these were too unattainable in the short term. For this reason two types of standards were chosen:

Table 1 | Examples of initial standards set for the prevalent (first) round of screening for women aged 50-64

Objective	Measurement	Minimum acceptable standard	Target standard
Maximise the number of cancers detected	No of cancers detected in women invited and screened	›3.5 in 1000	›5 in 1000
Minimise the number of women referred unnecessarily for further tests	No of women referred for assessment	‹10% of women screened	‹7% of women screened

PARTNERS/ALAMY

- The minimum acceptable standard, below which no programme should fall
- The target standard, to which all programmes could reasonably aspire.

The target standard was set at a cut-off point between the top quartile of programmes and the rest: if a quarter of programmes could achieve a certain performance then all programmes should be able to. Meanwhile the programmes in the top quartile could look to the Swedish benchmark for their quality improvement targets—and the aim was that the whole programme should continuously improve.

Ensuring continuous improvement

Screening has many analogies with production line engineering. The tasks are repetitive and potentially boring, the margin between success and failure is fine, consumer expectations keep rising, and the potential of errors to destroy public confidence is great. The challenge was not only to get the programme up and running but also to improve it each year. Continuous improvement was needed not only to reach Swedish standards but also to keep staff motivated; constant investment in research and development, and in innovation and evaluation, was essential.

Quality assurance was built into the programme from its inception. It was funded centrally and each region identified a quality assurance manager, who established a quality assurance reference centre to collate and analyse data as well as provide local support. A crucial early decision was that the individual screening units had to cover a population of at least 500 000 (leading to around 100 individual units) so that expertise could be developed and maintained and so that quality assurance measures derived from annual data were statistically meaningful.

An information system was designed and procured to enable the continuous measurement and feedback of performance. Units had to send returns annually to the centre for analysis of each programme's and each region's performance. Anomalous results were always investigated. Sometimes the problem was a lack of clarity in data definition or poor data collection. But when true underperformance was found, the local quality assurance team could diagnose the cause and organise remedial action, perhaps training of staff or replacing equipment.

Some of these performance measures have subsequently been found to be wanting. The original target detection rate of five cancers per 1000 women was found to be unreliable after the publication of interval cancer rates.[8] The target was replaced by the more sophisticated standardised detection ratio, which is an age standardised (and background incidence corrected) measure that compares the observed number of invasive cancers with a calculated expected number giving a target standard of 1.0.[9] The expected number was based on a screening programme with an estimated sensitivity the same as that in the Swedish randomised controlled trial.[3] Each screening unit could therefore be compared with the trial. Units that achieved a ratio of 1.0 were performing at the same quality as the Swedish programme and thus should achieve the same reduction in mortality. No unit was expected to achieve a ratio below 0.75. The ratio was found to be both simple and robust. The policy of investigating units with a ratio less than 0.75 enabled several underperforming units to be identified, and after investigation the performance of these units improved greatly.

Results for individual units are reported to each relevant national professional group. These professional groups can review the standard and reset them, identify training needs, and where practice is changing. In this way each professional working in the programme is linked into the standard and target setting mechanism of the programme and the targets and standards reflect professional concerns and progress.

Performance

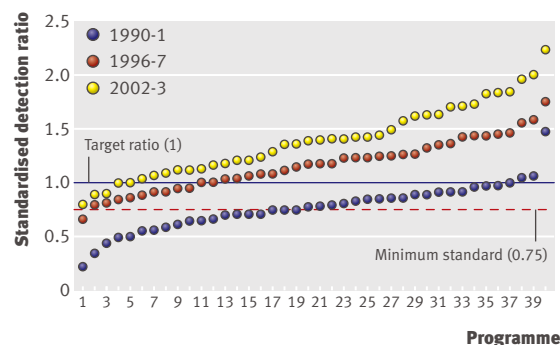As the average standard improved, the target increased so that there was always something to

Fig 1 | Prevalent screen standardised detection ratio for the 40 largest screening units in England ranked in ascending order for 1990-1, 1996-7, and 2002-3

481

strive for and standards climbed upwards. This policy has resulted in major improvements in quality over time. For example, the standardised detection ratio for invasive cancers for the programme as a whole has improved steadily (fig 1). The ratio was not developed until 1995, but had it been available in 1990-1 there would have been more concern over the cancer detection rates because few units reached the target of 1, and half were below the minimum standard of 0.75. In the early years of the screening programme a high detection rate for non-invasive cancers masked a poor sensitivity for the detection of invasive cancers and the lack of age standardisation and statistical stability were major problems in determining the performance of individual units.

Although some of the increase in the standardised detection ratio is now known to be the result of increasing background incidence from the use of hormone replacement therapy, a substantial amount—and probably the largest—is due to improvements in the programme. **These include introducing two view mammography for prevalent screens in 1995 and incident screens in 2003, setting the range for film density towards darker films, and greater use of double reading.** All these changes are the result of evidence gained from the screening programme through randomised controlled trials or observational studies (table 2).
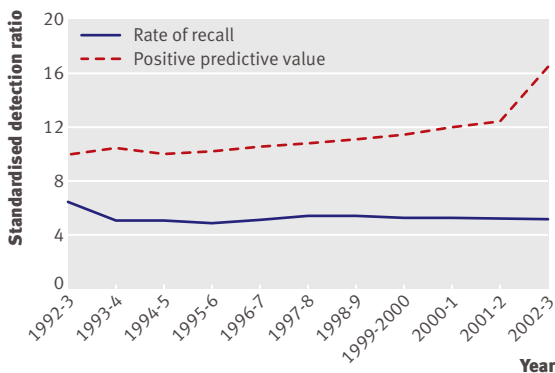
Table 2 | Effect of different protocols on standardised detection ratio (SDR) for small invasive breast cancers (<15 mm), 1996-7*[10]

Protocol	SDR	Rate ratio (95%CI)
One view/single reading	0.68	1.00
One view/double reading (recall if one reader suggests)	0.93	1.37 (1.15 to 1.62)
Two views/single reading	0.97	1.43 (1.15 to 1.77)
One view/double reading (consensus)	1.00	1.47 (1.21 to 1.78)
Two views/double reading (recall if one reader suggests)	1.05	1.54 (1.26 to 1.87)
Two views/double reading (consensus)	1.12	1.64 (1.31 to 2.06)
One view/double reading with arbitration)	1.18	1.73 (1.40 to 2.13)
Two views/double reading with arbitration	1.28	1.88 (1.49 to 2.37)

*Study may be subject to biases and should not be considered a substitute for a randomised controlled trial.

While cancer detection rates were improving, an attempt was made to reduce the rate of recalls for assessment. In the second (incident) round of screening, rates of recall were around 5% and the national advisory group thought that they could be reduced. By 1995-6 rates had fallen to 4.9%, but cancer detection rates had fallen and the pressure to reduce assessment rates was eased. Figure 2 shows how assessment rates have remained at just over 5%, while the positive predictive value of a referral has risen year on year. The steep increase in 2002-3 is the result of the introduction of two view mammography at all screening rounds.[11] Although a low recall rate is desirable, if the rate is too low it can compromise the cancer detection rate (fig 3).

In fact figure 3 shows the information that allows radiologists and quality assurance staff to compare the performance of individual units with that of others. The degree of simplicity is possible only in a national programme where the age range of women screened and the screening interval are uniform and the programme is in steady state, all units having completed initial screens for the whole of their geographical area.

Finally, table 3 shows the improvement in non-operative diagnosis rates. The minimum standard for non-operative diagnosis has now been raised to 80% from 70% because all but a handful of units are achieving the minimum standard. The target has remained the same.

Effects of screening

The aim of the programme was to reduce mortality, and it was recognised at the outset that it would take about 15 years for the programme to have its full effect. With the quality improvements and inclusion of older women (65-70), screening is likely to continue to

Fig 2 | Rate of recall for assessment at incident screening and positive predictive value of recall
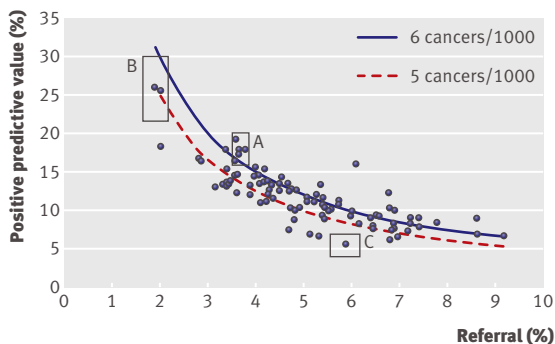

Fig 3 | Positive predictive value of recall versus recall for assessment for all 95 UK screening units 1999-2000 (women aged 50-64). Boxes A-C highlight three example units plus 90% confidence intervals, with box A showing a unit with optimal qualities of high positive predictive value and cancer detection rates but low referral rate

Secrets of successful screening

Use of "total quality" approach from outset

Clear objectives and standards

Single national protocols to enable comparison of results and identification of trends

Single national dataset with well understood definitions

Data must be complete and good quality (small differences can affect many people in population based programmes)

Availability of a clear reference point, such as a randomised controlled trial

reduce mortality until at least 2020. Recently announced extensions to the programme could produce reductions beyond this date.[12]

Many factors other than screening affect mortality from breast cancer, and while most of these factors decrease death rates (such as screening, treatment, and cohort effects in younger women), some are causing increases (such as cohort effects in older women), and some have caused both (changes in the rules for death certificates).[13] Separating these effects is difficult in the absence of control populations. Nevertheless, the fall in mortality since screening was introduced in 1990 means that breast cancer is no longer the most common cause of death from cancer in women.

The breast screening programme has also transformed diagnostic and treatment services for breast cancer. The programme forced the replacement of 200 clinics with widely varying staff and skills with 100 multidisciplinary centres focusing on the assessment of abnormal mammographic results and the diagnosis of breast cancer. Many of these assessment centres became the focus for diagnosing and managing breast disease in their area, bringing together not only radiologists and radiographers but also pathologists, surgeons, radiotherapists, and oncologists. The programme can probably therefore claim some of the credit for the swift take-up of tamoxifen as the drug of choice for women with breast cancer.

The ethos of quality assurance expanded well beyond imaging and was enthusiastically embraced by pathologists, who refined definitions and developed "slide exchange" into external quality assurance.[14 15] Consistency remains higher among recognised breast pathologists than among general pathologists, but this difference has lessened over time. Surgeons' annual audit of treatment is now a mine of useful information about breast care in the UK.[16] The box summarises the lessons from the screening programme.[17]

The concept of total quality assurance has been successfully applied to the NHS breast screening programme. The information systems and quality assurance systems are intertwined and inextricably linked to the supporting quality assurance structure. This has led to a breast screening programme that initially struggled to meet the required standard because of lack of experience, exceeding that standard and continuing to improve. As the information systems in the NHS improve and new screening programmes are brought in, this is an experience and a philosophy the rest of the NHS is drawing on to improve the entire service.

Table 3 | Rates of non-operative diagnosis for screening programme (minimum standard 70%, target standard ≥90%)

Year	Women with non-operative diagnosis (%)	Regions meeting minimum standard (%)	No (%) regions meeting target
1997-8	71	68	0
1998-9*	81	100	1 (7)
1999-2000*	85	100	1 (10)
2000-1	87	100	2 (15)
2001-2	89	100	6 (45)

*Data from Scotland not available.

SUMMARY POINTS

The NHS screening programme for breast cancer has been running for 20 years

Quality assurance was built into the programme from implementation

Standards have improved steadily and mortality has fallen

The programme has had wider effects on breast cancer services

1   Forrest APM. *Report to the health ministers of England, Scotland, Wales and Northern Ireland on breast cancer screening*. London: HMSO, 1986.
2   Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129-34.
3   Tabár L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Gröntoft O, 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:829-32.
4   Gray JAM. *General principles of quality assurance in breast cancer screening*. London: Department of Health, 1990.
5   Ayer AJ. *Language, truth and logic*. London: Penguin, 1936.
6   Pritchard J. *Quality assurance guidelines for mammography*. Oxford: NHS Breast Screening Programme, 1989.
7   Donabedian A. *Explorations in quality assessment and monitoring: the definition of quality and approaches to its assessments*, Vol 1-3. Ann Arbor, MI: Health Administration Press, 1980.
8   Prior P, Woodman CB, Wilson S, Threlfall AG. Reliability of underlying incidence rates for estimating the effect and efficiency of screening for breast cancer. *J Med Screen* 1996;3:119-22.
9   Blanks RG, Day NE, Moss SM. Monitoring the performance of breast screening programmes: use of indirect standardisation in evaluating the invasive cancer detection rate. *J Med Screen* 1996;3:79-81.
10   Blanks RG, Day NE, Moss SM. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service breast screening programme. *J Med Screen* 1998;5:195-201.
11   Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ* 2000;321:665-9.
12   Department of Health. Cancer screening to be expanded and waiting times to be further reduced. Press release 0273, September 2007. www.independenthealthcare.org.uk/downloads/briefing_papers/IHAS_Press_Releases_Sep_07.pdf.
13   Sloane JP, Ellman R, Anderson TJ, Brown CL, Coyne J, Dallimore NS, et al. Consistency of histopathological reporting of breast lesions detected by screening: findings of the UK National External Quality Assessment (EQA) Scheme. UK National Coordinating Group for Breast Screening Pathology. *Eur J Cancer* 1994;30A:1414-9.
14   Ellis IO, Coleman D, Wells C, Kodikara S, Paish EM, Moss S, et al. Impact of a national external quality assurance scheme for breast pathology in the United Kingdom. *J Clin Pathol* 2006;59:138-45.
15   Sauven P, Bishop H, Patnick J, Walton J, Wheeler E, Lawrence G, et al. The National Health Service breast screening programme and British Association of Surgical Oncology audit of quality assurance in breast screening 1996-2001. *Br J Surg* 2003;90:82-7.
16   Sloane JP, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, et al. Consistency achieved by 23 European pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas. European Commission Working Group on Breast Screening Pathology. *Virchows Arch* 1999;434:3-10.
17   Blanks RG, Bennett RL, Wallis MG, Moss SM. Does individual programme size affect screening performance? Results from the United Kingdom NHS breast screening programme. *J Med Screen* 2002;9:11-4.