Mass screening: theory and ethics

David Mant, Godfrey Fowler

The need for caution

Screening has the potential to do more harm than good. In the 1950s considerable enthusiasm was engendered for mass radiography and sputum cytology as screening tests for lung cancer. At first, screening seemed to offer great benefit: in one of the largest population based screening studies the proportion of cancers detected sufficiently early to be treated with surgery was almost 50%, and the five year survival was 35%—more than twice that in the control population. It was therefore a great disappointment that 10 years after screening began the cumulative mortality from lung cancer was identical in the screened and control populations.¹

The experience of lung cancer screening should be remembered by those who try to justify screening in terms of high yields of disease, early stage at diagnosis, or even improved survival. Screening always advances diagnosis (the “lead time”) and thereby improves survival from the time of diagnosis, but not necessarily from the time when the patient would have presented clinically. This is why screening is beguiling, and comparison of overall mortality and morbidity must be the final arbiter of efficacy.

When is mass screening justified?

In practice, screening is often promoted without due consideration of its efficacy or practical feasibility. There is strong empirical evidence that cervical cytology and (even taking into account recent trial results from Britain² and from Malmö³) breast cancer screening by mammography will save lives if properly organised.^{4,5} The effectiveness of cardiovascular screening has not been proved by clinical trial (nor by case-control studies), but advice from general practitioners to stop smoking and reduce alcohol consumption is heeded by patients,^{6,7} and stopping smoking reduces morbidity and mortality.⁸ Similarly, treatment of hypertension (particularly moderate and severe hypertension) reduces the risk of heart attack and stroke.⁹ The efficacy of urine analysis, faecal occult blood testing, and breast self examination is, however, less certain,¹⁰⁻¹² and mass screening cannot be justified on present evidence. The value of cholesterol measurement probably rests on its motivational effect rather than on its efficacy as a screening test.

*Estimated sensitivity, specificity, and predictive value of single view mammography in outpatient clinic and general practice**

Test	Outpatient clinic			General practice		
	Breast cancer			Breast cancer		
	Found	Not found	Total	Found	Not found	Total
Positive	50	90	140	5	95	100
Negative	10	850	860	1	899	900
Total	60	940	1000	6	994	1000
Sensitivity = 50/60 = 83%						
Specificity = 850/940 = 90%						
Predictive value† = 50/140 = 36%						
Sensitivity = 5/6 = 83%						
Specificity = 899/994 = 90%						
Predictive value = 5/100 = 5%						

*Actual numbers given are theoretical.

†Difference in predictive value reflects difference in prevalence of disease in the two settings.

Efficacy in clinical trials does not equal effectiveness in everyday practice. Over 20 years ago Wilson¹³ listed criteria that should be satisfied before screening is undertaken: the disease should be common and serious; its natural history should be understood; there should be a good screening test; acceptable treatment should be available; and this should favourably influence the outcome. For general practitioners these criteria can be translated into practical questions that must be answered before they commit their practice to screening (box). The rest of this brief paper is devoted to an explanation of the theory and ethical issues underlying these questions.

Minimum criteria for mass screening in general practice

Before screening anyone general practitioners should have answered the following questions to their own satisfaction:

- Can we offer effective treatment for patients positive on testing?
- How many positive tests will prove to be false alarms (and is this acceptable)?
- How many patients will need follow up over the next five years (and can we sustain this workload)?
- How are we going to audit routinely the quality of the test, of the intervention, and of follow up?

Assessing the test: false positives and false negatives

Screening tests are traditionally described in terms of their sensitivity and specificity. Calculation of these indices is shown in the table with single view mammography as an example. Essentially a “sensitive” test is one that rarely misses diseases whereas a “specific” test causes few false alarms. The concepts of sensitivity and specificity are used to characterise a screening test because they do not depend on the setting in which the test is carried out. Sensitivity is assessed on patients with disease; specificity is assessed on healthy individuals. However, in the clinical setting a screening test is applied to a mixture of people with and without disease. At this stage the practitioner is interested in the number of patients who have positive test results (for these must be followed up and further investigated) and in the number of false alarms. The proportion of patients with positive results who actually have the disease is termed the positive predictive value of the test.

The predictive value is probably the most important test characteristic for the clinician and the patient, and it varies greatly in different settings. In the theoretical examples given in the table more than one in three outpatients with a positive initial mammogram had breast cancer compared with one in 20 patients in general practice. This is because the predictive value depends on the prevalence of disease. A test (or clinical symptom or sign) may give a high predictive value in a hospital setting but may be of very little value in the community. Misunderstanding of this important point

Imperial Cancer Research Fund General Practice Research Group, University Department of Community Medicine and General Practice, Radcliffe Infirmary, Oxford OX2 6HE
David Mant, MRCP, lecturer in general practice
Godfrey Fowler, FRCP, reader

Correspondence to: Dr Mant.

Br Med J 1990;300:916-8

will lead to inappropriate screening and unfounded criticism of general practitioners by their hospital colleagues.

Trading sensitivity and specificity: the ROC curve

The other important point about sensitivity and specificity is that there is a trade off between them. This is shown by the graph in the figure, which is called a ROC (receiver operated characteristic) curve because it was first formulated to describe the decision making behaviour of radar operators. If the general practitioner adopts a high threshold he or she will seldom treat someone unnecessarily but will miss treatable disease. Conversely, doctors who adopt a low threshold will treat many patients unnecessarily. In clinical care each doctor makes this decision daily for individual patients. In mass screening (for example, urine analysis) it must be decided whether the criterion for a positive test is "trace" or "+++" This decision should be made before screening begins, and it determines the number of false positive and negative results that will occur in practice.

Frequency and age intervals

In order to decide how often to screen, it is helpful to recognise the difference between the first and subsequent rounds of screening. The first screen is always the most "successful": acceptance is high, there is more disease to find, and survival from diagnosis is longer (because the proportion of patients with aggressive disease is less than in subsequent rounds—the "length bias"). It is misleading to try to assess a screening test purely on the experience of the first round. This makes urine analysis difficult to assess because there are few data beyond those obtained from once only programmes.

It is also important to understand the course of the disease in question. The longer this is, and in particular the preclinical detectable phase, the longer can be the interval between screens. A five year interval was selected for cervical screening because the average length of the preclinical detectable phase of cervical neoplasia was thought to be at least 15 years. Breast cancer has a shorter phase (perhaps three to four years on average), and so a shorter screening interval is necessary. The recommended screening interval is a compromise that seeks to balance the number of patients who develop disease between screens (interval cases) with the economic costs to the health service and the personal costs to the patient.

Conventional ROC (receiver operated characteristic) curve showing three possible operating points¹⁴

Costs of screening

Costs to the patients

- Economic loss (lost work, increased insurance premiums)
- False reassurance to some patients
- Unnecessary anxiety and further investigation for some patients
- Psychological harm

Costs to the practice

- Time and resource cost of test
- Time and resource cost of follow up

Costs to the NHS

- Reimbursements to practice from family practitioner committee
- Prescribing costs
- Increased use of hospital facilities.

The choice of age groups to be screened depends on a number of factors. Mammography is not recommended for women under 45 because of lack of evidence of effectiveness or for women over 64 because of concern over low compliance. Screening for glycosuria in patients under 60 is not useful because it identifies few cases of treatable diabetes and a high number of patients with renal glycosuria. Cardiovascular risk is strongly related to length of exposure to most risk factors (such as smoking habit), but the optimal time to screen must also reflect the time when intervention is most likely to be effective.

Outcomes of screening: costs and benefits

The costs and benefits of screening are extremely difficult to quantify. Benefits are most easily measured in terms of improvement in morbidity and mortality, and when possible these must be established by randomised trial. There may also be a long term benefit to the health service in terms of future treatment avoided. The main costs are summarised in the box. The reported negative psychological effects merit further study.¹⁵

Economic costs are no easier to establish than personal costs. For the general practitioner it is usually the marginal cost of the test (for example, the urine dipstick) and of the follow up time that must be weighed against payments from the family practitioner committee. For the health service as a whole the average cost is more important, and the capital costs of building as well as of investigations and long term treatment must be considered. Analyses have been carried out to try to establish the cost-benefit ratio for screening. For example, it was calculated in 1985 that 40 000 smears and 200 excision biopsies were being carried out in Britain to prevent one death from cervical cancer.¹⁶ Figures given for the cost per life saved have been as high as £300 000 for cervical screening. These figures can only be accepted with qualification, but their usefulness is in allowing some assessment of opportunity costs. In 1986 the estimated cost of mammographic screening was £3000-5000 per quality adjusted life year saved, which is comparable to renal and heart transplantation but more than hip replacement (£750).¹⁷ In comparison, the screening cost per quality adjusted life year is estimated to be £180 for smoking and £1700 for hypertension.¹⁸

Ethical responsibilities

The ethical imperative is to make sure that the benefit to each patient from screening is likely to outweigh the harm. Because there is a trade off between false reassurance and false alarms, this does not mean that each patient will benefit from screening. It does mean, however, that each practice has an ethical responsibility to inform each patient of the balance of risks and benefits, to ensure that the quality of screening is sufficiently high to minimise the number of false positive and false negative tests, to ensure that the intervention offered is effective, and to audit routinely the completeness of follow up.

Many commentators have been rightly critical of the failure of screening programmes in the United Kingdom to meet these ethical responsibilities. In some cases screening has been imposed (for example, routine urine analysis for protein) even though the balance of scientific evidence is against any benefit. More often, enthusiasts embrace screening without sufficiently training their staff, without protocols for the definition of positive results and management of these patients, and without appreciation of the enormous workload generated by the need to investigate and follow up patients over a long period. These observations are made from bitter experience rather than theory: enthusiasm to do good is simply not enough.

Although screening may have been unethical in its incompetent practice, we disagree with Skrabanek that screening can be achieved only by coercion and deception and that it is invariably an infringement of the patient's personal liberty.¹⁹ Our view is that effective screening for smoking habit, blood pressure, alcohol consumption, and breast and cervical cancer is

likely to enhance the likelihood of a patient living a free and healthy life.

- 1 Fontana S. Screening for lung cancer. In: Miller AB, ed. *Screening for cancer*. Orlando: Academic Press, 1985:377-94.
- 2 UK Trial of Early Detection of Breast Cancer Group. First results on mortality reduction in the UK trial of early detection of breast cancer. *Lancet* 1988;ii:411-6.
- 3 Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *Br Med J* 1988;297:943-8.
- 4 Day NE. Screening for cancer of the cervix. *J Epidemiol Community Health* 1989;43:103-6.
- 5 Roebuck EJ. Mammography and screening for breast cancer. *Br Med J* 1986;292:223-5.
- 6 Wallace P, Cutler S, Haines A. Randomised controlled trial of general practitioner intervention in patients with excessive alcohol consumption. *Br Med J* 1988;297:663-5.
- 7 Russell M, Wilson C, Baker C. Effect of general practitioner's advice against smoking. *Br Med J* 1979;iii:231-5.
- 8 Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J* 1976;iii:1525-36.
- 9 Wilcox R, Mitchell J, Hampton J. Treatment of high blood pressure: should clinical practice be based on the results of clinical trials? *Br Med J* 1986;293:433-7.
- 10 Mant D. Breast self examination: should we discourage it? *J R Coll Gen Pract* 1989;39:180-1.
- 11 Simon JB. Occult blood screening for colo-rectal carcinoma: a critical review. *Gastroenterology* 1985;88:820-37.
- 12 Zilva JF. Is unselective biochemical urine screening cost effective? *Br Med J* 1985;291:323-5.
- 13 Wilson JMG. Some principles of early diagnosis and detection. In: Teeling-Smith G, ed. *Proceedings of a colloquium Magdalen College, Oxford, July 1965*. London: Office of Health Economics, 1976.
- 14 Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 1978;8:283-98.
- 15 Marteau T. Psychological cost of screening. *Br Med J* 1989;299:527.
- 16 Anonymous. Cancer of the cervix: death by incompetence [Editorial]. *Lancet* 1985;ii:336-7.
- 17 Department of Health and Social Security. *Breast cancer screening: report to the health ministers of England, Wales, Scotland, and Northern Ireland*. London: HMSO, 1986. (Forrest Report.)
- 18 Williams A. Screening for risk of CHD: is it a wise use of resources? In: Oliver M, Ashley-Miller M, Wood D, eds. *Screening for risk of coronary heart disease*. Chichester: Wiley, 1987:97-106.
- 19 Skrabanek P. Preventive medicine and mortality. *Lancet* 1986;i:143-4.

THE MEMOIR CLUB

One day, she marched into my room (I was junior to her, of course) and sternly laid a manuscript on my table. Her gesture said, "Take that — or else!"

Even as I thanked her for the contribution I caught sight of the title and let out a rather ill judged laugh. The title was "ATS reconditioning centres." Albertine Winner was not amused.

"What are you laughing at?" she demanded to know.

I replied that I was sure such centres would be useful and ought to be made known; but the title would surely need further consideration.

"What is wrong with it?" demanded Albertine. "It describes exactly the function of these centres; and I will not have them described as 'rehabilitation centres' as you are probably going to suggest."

I denied any such thought. I did not like "rehabilitation" as a word any more than she did; but I foresaw that I was not the only one who might give a rude giggle at the idea of ATS reconditioning centres. I tried to persuade Albertine that ribald laughter from "the boys" would be all too likely to prevent the centres from being taken seriously and understood for what they were.

She patiently described how she had become aware that many girls volunteering for army service had such health disabilities as would make them unable for the work required. "They have bad teeth, poor eyesight, feet with corns and bunions, running noses, and poor vital capacity." Until they have been "reconditioned" they would be quite unfit for army service. I could see all that, but I still could not see the article being taken at its true value because of the naughty thoughts that its title would provoke.

"Think about it as much as you like," said Albertine, "but I will not have 'rehabilitation.'"

Fortunately my distinguished predecessor as editor was due to call on me shortly, and I proposed to Albertine that his opinion should be sought. I invited her to be present when, without any comment, I showed him the manuscript. Fox's spontaneous reaction was all that was required. Albertine saw that "reconditioning centres" must find another name. After a good deal of detailed discussion about the function of the centres, Fox said: "Albertine, the girls are disabled. You want them to undergo 'reablement.'"

I do not know if this was the moment when "reablement" was born in medical literature. The word was, and is, a good one. It appeared in the

title of Albertine's article and was given a good run; but I doubt if anyone uses it now or if it will ever become current usage. I cannot think why not.

From *Portraits from Memory* by James Howie. Published under the BMJ's Memoir Club imprint. ISBN 0 7279 0243 1. Price: Inland £14.95; abroad £17.50; USA \$29.00. BMA members: Inland £13.95; abroad £16.50; USA \$27.00.

As the weeks passed and our experiments began to show results, these evening hours were turned over to working out the results and trying to interpret them by drawing innumerable graphs. Every few days, whenever he was free or we had something new to discuss with him, we would have noisy discussions with Dick. Only when I was leaving did I appreciate the right way to use these occasions. I made the mistake of approaching each session expecting conclusions and plans for further experiments. But no. It was not always easy to get Dick to pay attention; he tried but his mind kept wandering and he would miss steps in the arguments or, if he did understand something and like it, he would go off at a tangent. But his delayed action was superb. Get enough of the ideas across, give him some results and a few graphs, and he would come back next day or after the weekend with fresh ideas. Sometimes he went too far too fast. Once he presented me with a very wide ranging and precise graphical analysis of the pressures and flows in various parts of the airways and lungs using our findings. I protested that our equipment was not good enough for us to trust the data he had built into his edifice but he was so excited that he swept that quibble aside and went on explaining the merit of looking at things his way. I interrupted him with a pompous remark I have never regretted, "I was not brought up to speculate wildly on the basis of very few experimental facts obtained with equipment of questionable accuracy."

"Then you've come to the wrong place. I've spent my life doing just that."

The other, romantic and truer voice of science was speaking.

From *Not Always on the Level* by E J Moran Campbell. Published under the BMJ's Memoir Club imprint. ISBN 0 7279 0184 2. Price: Inland £14.95; abroad £18.50; USA \$30.00. BMA members: Inland £13.95; abroad £17.50; USA \$28.00.