New concepts in screening

J A Muir Gray

Summary

All screening programmes do harm; some do good as well

ALTHOUGH screening is often delivered by clinicians, screening programmes are public health services that need to be managed at the level of a large population to monitor quality effectively. In the United Kingdom (UK) this is carried out by the National Screening Committee.

The first task of the National Screening Committee is to use research evidence to identify programmes that do more good than harm. The second is to make policy recommendations about those programmes that will do more good than harm at reasonable cost, focusing on opportunity cost; that is, the professional time involved, as well as the financial cost.

When it has been decided to introduce screening, using criteria based on the Wilson and Jungner WHO criteria,[1] care has to be taken to ensure that the programme is set up in a way that will minimise harm and maximise benefit.

In policy making the evidence for screening is often limited because of the rarity of the conditions being screened for, which minimises the contribution that the randomised controlled trial can make, but emphasises the need for a systematic review of the evidence.

Screening is a programme, not a test

Screening programmes, formerly called secondary prevention, should be set in the context of primary prevention and good diagnostic and treatment services; screening programmes are part of disease control programmes. This principle is particularly important when communicable disease is the focus of screening. Increasingly the emphasis is on identifying people at high risk rather than single risk factors.

Screening is managed on the principles of total quality management which involves the need to continually strive to prevent errors, identify those errors that do occur for effective action, help those involved in screening improve their performance, and set and re-set standards at a national level.

The term 'screening', although deeply embedded in professional and public consciousness, is not particularly helpful in the 21st century. The original meaning of the word 'screen' was a sieve, something with holes in it, and all screening programmes have false positives and false negatives. In the 20th century, however, the word 'screen' came to mean something without any holes in it and public expectation has evolved to such a degree that false positives and false negatives are automatically assumed to be due to clinician or programme error and to therefore provide a basis for litigation.

J A M Gray CBE, DSc, MD, FRCP, FFPHM, director of Screening Programmes, UK National Screening Committee, Institute of Health Sciences, Oxford. This text is based on a chapter in the forthcoming RCGP book due to be published in 2004:
Charlton R and Lakhani M (eds). *Recent advances in primary care.*

Address for correspondence

Dr J A Muir Gray, NHS UK National Screening Committee, Institute of Health Sciences, Old Road, Headington, Oxford OX3 7LF.
E-mail: muir.gray@ihs.ox.ac.uk

Submitted: 7 November 2003; Editor's response: 22 December 2003; final acceptance: 24 February 2004.

- The condition sought should be an important health problem.
- There should be an accepted treatment for patients with recognised disease.
- Facilities for diagnosis and treatment should be available.
- There should be a recognisable latent or early symptomatic stage.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- The natural history of the condition, including development from latent to declared disease, should be adequately understood.
- There should be an agreed policy on whom to treat as patients.
- The cost of case finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case finding should be a continuing process and not a 'once and for all' project.

Box 1. Criteria for appraising screening developed in the 1960s.[1]

Introduction

The management of screening is a public health service and consists of a limited number of tasks, notably:

- identifying programmes that do more good than harm;
- estimating which programmes do more good than harm at reasonable cost;
- ensuring that each programme is introduced and delivered at a sufficient level of quality to reproduce the levels of benefit and harm that were found in the research setting in ordinary service settings.

These are the key responsibilities of the UK National Screening Committee, which has been developing a new approach to screening since its foundation in 1997.

Criteria for screening policy

The criteria developed by Wilson and Jungner[1] have stood the test of time very well and were adopted by the National Screening Committee even though they were published about 30 years earlier (Box 1).

There was concern, however, about the use of these criteria because of changing thoughts on evidence and the evaluation of programmes. The Wilson and Jungner criteria were very welcome following the hubris of the early 1960s,

The condition
1. The condition should be an important health problem.
2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage.
3. All the cost-effective primary prevention interventions should have been implemented as far as practicable.
4. If the carriers of a mutation are identified as a result of screening the natural history of people with this status should be understood, including the psychological implications.

The test
5. There should be a simple, safe, precise and validated screening test.
6. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.
7. The test should be acceptable to the population.
8. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.
9. If the test is for mutations the criteria used to select the subset of mutations to be covered by screening, if all possible mutations are not being tested for, should be clearly set out.

The treatment
10. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
11. There should be agreed evidence-based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.
12. Clinical management of the condition and patient outcomes should be optimised in all healthcare providers prior to participation in a screening programme.

The screening programme
13. There should be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an 'informed choice' (for example, Down's syndrome and cystic fibrosis carrier screening), there must be evidence from high-quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.
14. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/intervention) is clinically, socially, and ethically acceptable to health professionals and the public.
15. The benefit from the screening programme should outweigh the physical and psychological harm (caused by the test, diagnostic procedures and treatment).
16. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (ie value for money).
17. There should be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards.
18. Adequate staffing and facilities for testing, diagnosis, treatment, and programme management should be available prior to the commencement of the screening programme.
19. All other options for managing the condition should have been considered (for example, improving treatment and providing other services), to ensure that no more cost-effective intervention could be introduced or current interventions increased within the resources available.
20. Evidence-based information, explaining the consequences of testing, investigation, and treatment, should be made available to potential participants to assist them in making an informed choice.
21. Public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process, should be anticipated. Decisions about these parameters should be scientifically justifiable to the public.
22. If screening is for a mutation, the programme should be acceptable to people identified as carriers and to other family members.

Box 2. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme — 2003.[1-7]

during which multiphasic screening had been introduced in many countries without adequate evaluation. Useful though they were, however, these criteria did not meet the needs of the late 20th century, in particular because:

- insufficient emphasis was given to the harm caused by screening in an era in which concern about harm, and the possibility of litigation, was much higher;
- the strength of the evidence required to support the claim of benefit was not sufficiently clear;
- the opportunity costs of screening rather than the availability of resources had become an increasing issue as all societies were faced by the challenges posed by an ageing population, new technology, and rising demand.

Because of this the National Screening Committee developed a new set of criteria focusing on four main questions:

- Do we understand the natural history of the disease?
- Is there a good screening test?
- Is there an effective treatment?
- Is the programme acceptable to the population.

These criteria are set out in Box 2.

Appraising evidence

The development of new thinking about evidence is of central importance and, for this, the National Screening Committee was able to look to the broader debates about evidence stimulated by the evidence-based decision-making movement and the Cochrane Collaboration. In a hierarchy of evidence, a systematic review of randomised controlled trials is usually placed at the top. Even this, however, does not resolve disputes, because value judgements are involved in the selection or rejection of trials to be included in a system-

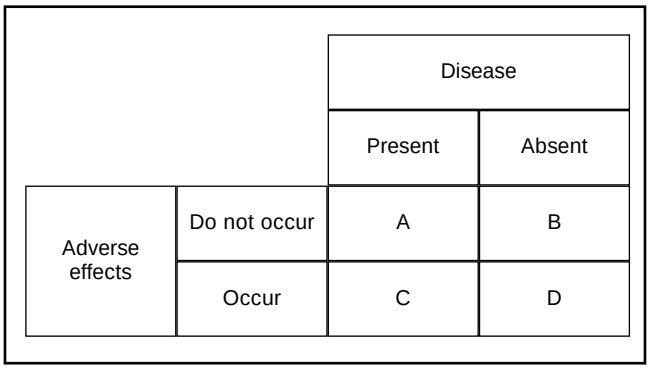

Figure 1. Benefits and side effects in screening.

atic review. This topic was most fiercely argued in the debate about breast cancer screening. In this debate, a review in the *Lancet* suggested that the evidence for screening had been biased by the inclusion of trials of low quality.[8] A vigorous exchange of letters took place in the *Lancet*'s correspondence column and the issue was reviewed by the International Agency for Research on Cancer, which published a report 2 years later.[9] The conclusion of this report was that:

> *'… the trials have provided sufficient evidence for the efficacy of mammography screening of women between 50 and 69 years. The reduction in mortality from breast cancer among women who chose to participate in screening programmes was estimated to be about 35%. For women aged 40–49 years, there is only limited evidence for a reduction.'*

The limitations of randomised controlled trials

In many screening programmes; for example, those that may lead to termination of pregnancy, such as screening for Down's syndrome, the randomised controlled trial is not an appropriate type of study design and other forms of evidence, including modelling, are used. The most striking example of this has been the serum, urine and ultrasound screening study, which reviewed the outcome of antenatal testing for Down's syndrome screening and then used modelling to estimate the impact of different screening policies.[10]

Furthermore, many studies have focused on the benefits of screening, and the need to focus on harm was also emphasised by the National Screening Committee, aware of the fact that controlled trials with enough power to identify the benefit of screening may not have sufficient power to identify the harm.

Limited evidence about harm

Evidence about harm is now required by the National Screening Committee, but it is not always clearly stated in research studies carried out in the 1980s and 1990s.

The fact that screening is offered to people who are apparently healthy is obviously one hallmark screening. Another important hallmark is the fact that side effects can occur in people for whom there is no possibility of benefit. In clinical practice, a person with a health problem who seeks the help of a clinician accepts the risks that treatment might involve, once they have been properly explained to them. In the case of screening, some people experience adverse effects but do not have the disease for which they have been screened (group D in Figure 1). For example, it is theoretically possible that a person without colorectal cancer could have their colon perforated during colonoscopy and die (Figure 1).

For this reason, it is important to gather evidence about harm, both common psychological harm, which is usually transient, and rarer and more serious forms of harm that may occur.

The need to calculate opportunity costs

The question, 'Would you spend £35 million on breast cancer screening?' may bring about a different answer to that from the question, 'If you had £35 million to spend on breast cancer control, would you spend it on screening?'.

The new criteria of the National Screening Committee seek to address the opportunity costs by considering whether or not the resources that might be made available for screening would be better invested in either primary prevention or treatment services. Furthermore, as the scarcest resource in health services becomes staff and not money, it is important to consider whether or not the investment of resources would not be better channelled into some other intervention designed to reduce the burden of the particular disease, or to reduce the burden of disease in the particular population that might be served by a screening programme.

The appreciation of opportunity costs forces screening programmes to identify new ways of delivering services. The newborn hearing screening programme has, for example, developed a completely new type of health worker, without traditional professional qualifications, to carry out screening tests. In primary care, receptionists could be trained to take blood samples or measure blood pressure.

New concepts in population disease control

One of the aims of screening is to control a disease at the population level. When screening techniques were first developed, health care largely consisted of individual consultations by clinicians who worked either in primary care or in hospitals, and who were isolated and remote from other hospitals. Since that time a number of changes have taken place.

The development of clinical networks has introduced 'systems thinking' to clinical care, setting the consultation in the context of other levels of care. The influential report from the Institute of Medicine called *Crossing the quality chasm* describes four levels of care[11]:

- Level 1 — the consultation.
- Level 2 — the microsystem; for example, the morning surgery.
- Level 3 — the organisation; for example, the hospital.
- Level 4 — the world of regulations, policies and guidelines.

Screening, therefore, can now fit more clearly into a systematic approach to disease control, and the concept of disease control is one that developed in poorer countries

All individuals aged 40 years and over

Have you had: a heart attack, angina, stroke, transient ischaemic attack, pain in your calves on walking that is relieved at rest, an operation to bypass the arteries of your heart, neck or legs?	Do you have diabetes? Yes	Do you have diabetes? No
Yes	A	B
No	C	D

A: prevent further manifestations of cardiovascular disease (CVD) and diabetes
B: prevent further manifestations of CVD, screen diabetes
C: prevent CVD, prevent complications of diabetes
D: screen for diabetes in selected people and prevent CVD

Figure 2. Clinical detection and screening scenario. Amended from Adler et al*, 2003.*[17]

with examples such as the schistosomiasis control programme, and the malaria control programme. This type of planning, called vertical planning, was often disparaged by people, particularly in the UK, where it was felt that horizontal planning; namely, organising primary care and secondary care, was more important. The truth, of course, lies somewhere in the middle. Many countries have now adopted a disease control model where, with the basis of a good platform of high quality primary care, disease control programmes are introduced. These are called National Service Frameworks in England. However, systems thinking is also influencing screening within the context of disease control programmes in a number of different ways.

Reducing population risk by primary prevention

The National Service Framework for coronary heart disease in England marked a shift away from the medicalisation of cardiac risk factors by emphasising that the top priority was to promote primary prevention among people who did not have heart disease.

This reflects the strategy advocated by the late Professor Geoffrey Rose in his influential book *Strategy of preventive medicine*.[12] Rose's approach would not be to seek out individuals with high blood pressure, but to try and reduce the body mass index of the whole population, or even the mean weight of the population by 2 or 3 kg, because most strokes occur in people who have not been identified in high-risk screening strategies. The health benefits that would result if the mean weight of the population was reduced by 10% would be dramatic.[13]

This led to another important concept — the focus on person-based rather than risk-based screening.

Identifying people at high risk

The National Service Framework emphasised the need to identify people who already had heart disease, picking up on the intriguing statement by Professor Nicholas Wald who said that: 'The best screening test for heart disease is the question "Do you have heart disease?"'.

Epidemiological evidence has been gathered to support this approach. The heart protection study, for example, demonstrated that the group who benefited most from cholesterol reduction were people with heart disease, whatever their initial level of cholesterol.[14-15] The issue, therefore, was not to identify people with a single risk factor and judge intervention on the basis of that risk factor, but to identify people who would benefit from a wide range of measures to reduce risk.

This was complemented by the development of risk factor calculators, which would allow the overall risk of an individual to be measured, and by suggestions that what was needed was a 'polypill' — a pill that would assume that everyone over a defined age was at risk of vascular disease and seek to reduce that risk by small amounts of active pharmacological treatments, each delivered at a dose high enough to have a beneficial effect but low enough to minimise side effects.[16] This also led the National Screening Committee to introduce a pilot project to reduce the risk of diabetes and vascular disease by using two questions as screening tests:

- Do you have diabetes?
- Do you have heart disease or any other form of vascular disease?

On the basis of the answers to these questions, the population could be divided into four groups: those who had heart disease and those who had diabetes, both of which already received risk reduction, and those who did not have heart disease or diabetes, which received advice to try and change their lifestyle, although the contribution of professional advice was relatively limited. The use of these screening tests, and the 2 ∞ 2 box that resulted from it, allowed the identification of individuals at particularly high risk to be screened for diabetes, and in the pilot project, the presence of a family history or a body mass index of over 25 is being used (Figure 2).[17]

Thus, this concept of screening focuses more closely on populations at risk rather than on risk factors. This concept for disease control is also relevant when screening for infectious diseases.

Chlamydia control

A proposal was made to screen for chlamydia, and the initial proposal suggested that a screening test be offered every 2 years. Communicable disease control is different from the control of non-communicable disease, however, because if an individual does not have the disease on testing, there is only one way in which that person can subsequently test positive — by infection from a person who already has the disease. Thus, the model of screening for non-communicable disease is not directly applicable to screening for communicable disease. The testing of asymptomatic people, piloted by the National Screening Committee in its chlamydia pilot programme, which took place in two populations in England, included asymptomatic testing, but this was part of a broader set of measures to control the communicable disease, including the promotion of primary prevention and the effective follow-up and treatment of those who tested positive.[18]

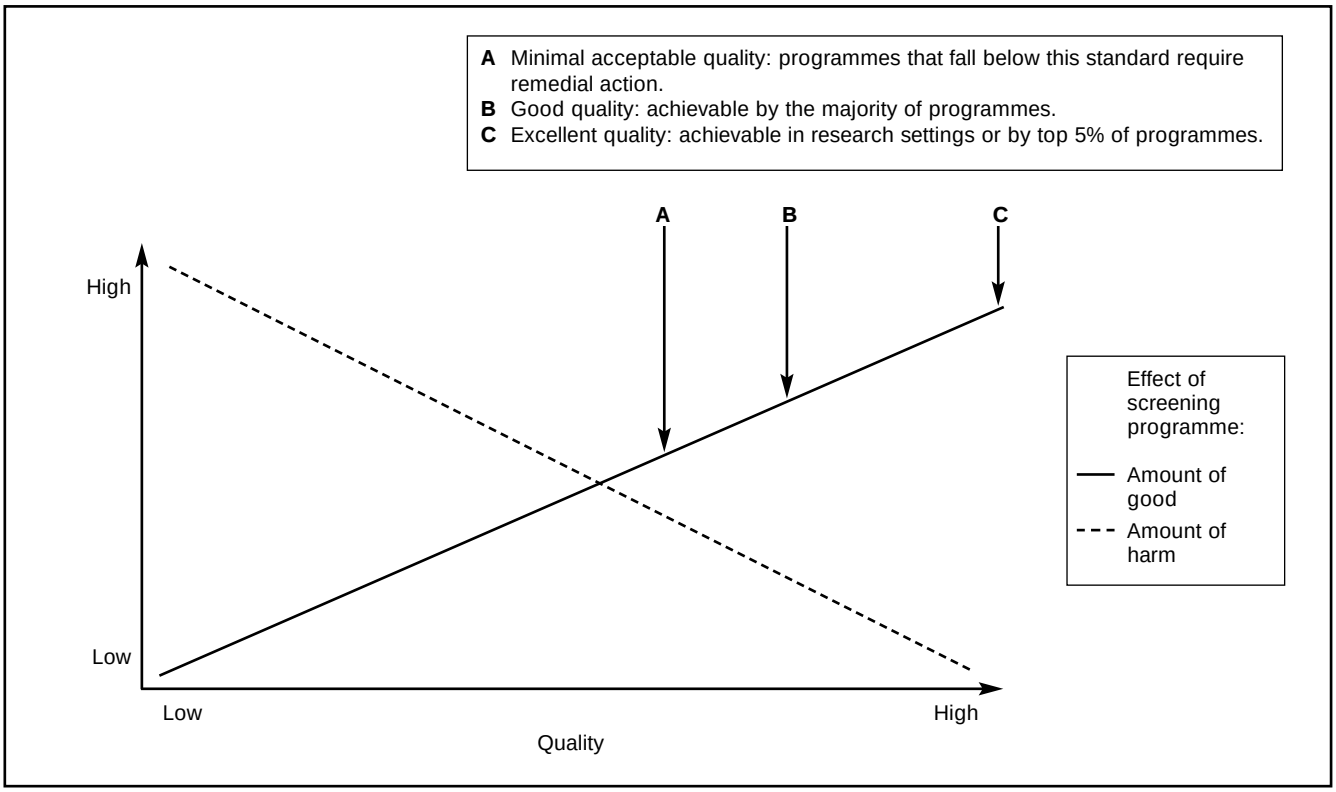

Figure 3. The balance of benefits, harm and quality.

New concepts in management

Screening is a programme, not a test; this is one of the problems that the health service has with private sector screening, where all too often an individual is offered a test, but the investigation and reassurance of those who test positive is then transferred to the National Health Service.

Programme management is essential to ensure quality, and quality is essential to ensure more good than harm is achieved. Screening policy is based on research findings, but there is growing evidence that the level of care delivered in a research study is better than the level of care delivered in an ordinary health service setting, for the control group as well as for the intervention group. For this reason, it cannot be assumed that the levels of benefit and harm achieved in a research setting will be reproduced in the ordinary service setting. Similarly, it cannot be assumed that the level of quality that was achieved by the research workers can be reproduced in service. In a typical research study covering half a million people, a number of highly dedicated professionals achieve levels of organisation and technical skill that will have to be reproduced by a hundred times as many staff to cover a population of 50 million. The relationship between quality and harm, and the need for explicit national standards is illustrated in Figure 3.

It is very difficult for any individual programme to work in this way. The balance between good and harm is very fine in screening and it is therefore essential for programmes to be able to compare their performance with the performance of other programmes. This requires national management.

National screening programmes:

- cover a defined population;
- have a simple set of objectives;
- develop valid and reliable criteria to measure performance and produce an annual report;
- relate performance to explicit quality standards;
- organise quality assurance systems to help professionals and organisations prevent errors and improve performance;
- communicate clearly and efficiently with all interested individuals and organisations;
- coordinate the management of these activities, clarifying the responsibilities of all individuals and organisations involved.

Box 3. Features common to all national screening programmes.

Nationally managed programmes have a number of common characteristics, set out in Box 3.

The national management of a programme involving millions of people being screened and thousands of healthcare professionals is complicated, but a core principle of the work of the National Screening Committee has been to 'simplify, simplify, simplify'. Almost all screening programmes follow the pattern shown in Figure 4.

Screening programme managers are not directly responsible for the quality of treatment services, but it is irresponsible to set up a screening programme without considering the quality of treatment and taking such action as is possible to make it more systematic and to improve the quality of treatment offered to those identified as having the disease.

Some screening programmes are even simpler; for example, the identification of the blood pressure level leads to treatment being introduced on the basis of blood pressure measurement alone; there is no explicit diagnostic test for high blood pressure (Figure 5).

A nationally managed programme has management arrangements for programme monitoring and quality assurance. A number of other programmes take place across the country as a whole, but are managed locally without national systems of quality assurance.

Quality assurance in screening

Quality assurance in screening is kept relatively simple and embraces four inter-related activities:

- the prevention of error by good staff training and selection, and the purchase of good quality equipment;
- the identification of errors, with rapid action to minimise their adverse effects;
- continual performance improvement on the part of professionals and screening programme teams;
- the setting and re-setting of national standards for screening.

This systems-based approach to screening is effective in assuring quality, but has been criticised for its production-line approach. It is true that the basic principles and practices are based on the techniques of quality assurance developed by the Japanese car industry, but the development of this systems-based approach has been accompanied by a radical rethinking of the philosophy of screening, with a move away from a utilitarian population-based approach to one that focuses on the needs, fears, and desires of the individual.

Screening or risk management?

Screening is a public health service delivered by clinicians to individuals. From its earliest days screening has been promoted as part of the public health service, and decision making has been based largely on utilitarian grounds; that is, on the needs of populations and the benefits to populations. If, for example, a programme would benefit a 1000 people and harm 20, the net benefit of 980 people would be regarded as evidence on which a programme could be considered for introduction.

An alternative scenario, based on utilitarianism, is that if a programme was of benefit to only a very small number of people it would not be offered, even if those individuals felt very strongly that they would benefit. The difference between these two approaches is well described in an analysis of decision making about breast cancer in the United States and Canada.[19] Screening for rare diseases does take place and is increasing, but screening for rare diseases is based on evidence of benefit for the small number of people that would be detected. More subtle problems arise when the evidence is not so strong, as is the case for screening women under the age of 50 years for breast cancer or screening for prostate cancer.

The fact that there is no evidence of benefit for these populations, based on our conventional definitions of effectiveness, does not mean that an individual might not benefit. The fact that the effect on a population is too small to measure indicates the magnitude of the benefit for the population, but does not exclude the possibility of benefit for the individual. This has led people to call for the introduction of screening programmes, such as screening for prostate cancer, because of the possibility of benefit. The people who call for prostate cancer screening are unconvinced by the epidemiological arguments based on concepts such as lead time bias. Even when there is clear evidence of benefit, the problem posed by the fact that some people would be harmed without a possibility of benefit has caused rethinking about ways in which the programmes should be presented.

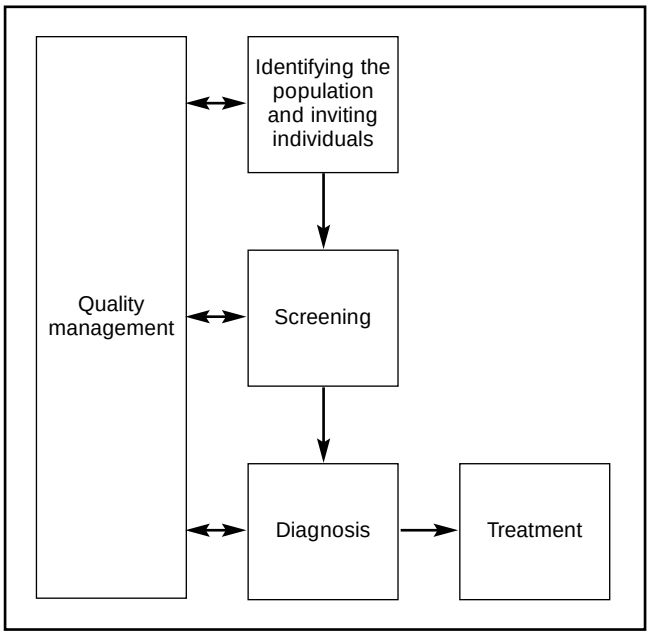

Figure 4. The basic elements of a screening programme.

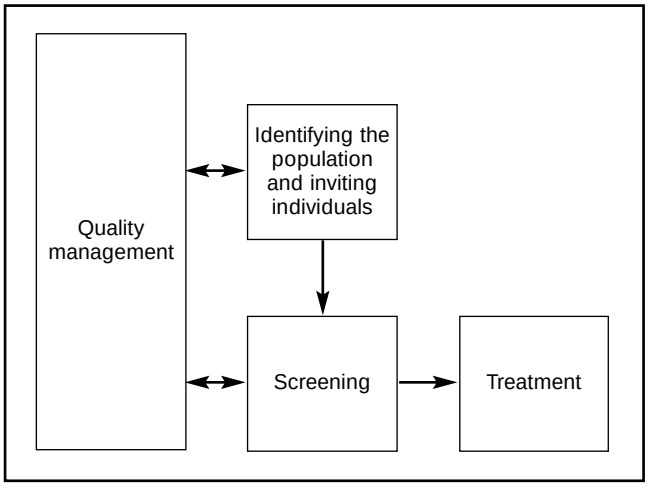

Figure 5. The basic elements of a screening programme when the screening test is diagnostic.

References

1. Wilson JMG, Jungner JJ. *Principles and practice of screening for disease*. Geneva: World Health Organisation, 1968.
2. Department of Health. *Screening of pregnant women for hepatitis B and immunisation of babies at risk*. London: Department of Health, 1998 (Health Service Circular: HSC 1998/127).
3. Cochrane AL. Holland WW. Validation of screening procedures. *Br Med Bull* 1971; **27:** 3.
4. Sackett DL, Holland WW. Controversy in the detection of disease. *Lancet* 1975; **2:** 357-359.
5. Wald NJ (ed). *Antenatal and neonatal screening*. Oxford: Oxford University Press, 1984.
6. Holland WW, Stewart S. *Screening in healthcare*. London: The Nuffield Provincial Hospitals Trust, 1990.
7. Gray JAM. *Dimensions and definitions of screening*. Milton Keynes: NHS Executive Anglia and Oxford, Research and Development Directorate, 1996.
8. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000; **355(9198):** 129-134.
9. World Health Organisation, International Agency for Research on Cancer. *Mammography screening can reduce deaths from breast cancer.* Lyon: International Agency for Research on Cancer, press release 139. http://www.iarc.fr/pageroot/PRELEASES/pr139a.html (accessed 5 Mar 2004).
10. Wald N, Rodeck C, Hackshaw A, *et al*. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Technol Assess* 2003; **7(11):** 1-77.
11. Institute of Medicine, Committee on Quality of Health Care in America. *Crossing the quality chasm: a new health system for the 21st century*. Washington DC: The National Academies Press, 2001.
12. Rose G. *The strategy of preventive medicine.* Oxford: Oxford University Press, 1993.
13. Mulvihill C, Quigley R. *Evidence briefing: The management of obesity and overweight. An analysis of reviews of diet, physical activity and behavioural approaches*. London: NHS Health Development Agency, 2003.
14. Heart protection study collaborative group. MRC/BHF heart protection study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; **360(9326):** 7-22.
15. Heart protection study collaborative group. MRC/BHF heart protection study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. *Lancet* 2003; **361(9374):** 2005-2016.
16. Smith R. Polypill may be available in two years [Editor's choice]. *BMJ* 2003; **327:** http://dx.doi.org/10.1136/bmj.327.7418.0-g (accessed 5 Mar 2004).
17. Adler A, Nicholl L, Bexon N, et al. *Diabetes, heart disease and stroke (DHDS) prevention pilot project (project description, version 3.0)*. Oxford: UK National Screening Committee, 2003.
18. Department of Health. *The national strategy for sexual health and HIV.* London: Department of Health, 2001.
19. Tanenbaum SJ. 'Medical effectiveness' in Canadian and U.S. health policy: the comparative politics of inferential ambiguity. *Health Serv Res* 1996; **31(5):** 517-532.