ACP Presidential address: Screening for presymptomatic disease

F V Flynn

Introduction

Intuition suggests that treating disease at a stage before symptoms have appeared and complications have had time to develop should offer the best chance of arresting or even reversing the pathological process. The concept of screening apparently healthy populations to detect asymptomatic disease, which follows from such thinking, has widely been regarded as an ideal which health care systems should seek to attain. Indeed, in the United States the "annual check-up" is accepted as a hallmark of good medical care.

Historical development of screening

Screening for the early detection of disease has been practised as a form of preventive medicine in Britain ever since the regular examination of schoolchildren was introduced on a national basis some 80 years ago. Provision for routine antenatal care and medical examination of infants soon followed, and early detection techniques were introduced to deal with endemic industrial diseases, such as chest x rays for detecting silicosis, and blood film and urinary analyses for picking up lead poisoning. In 1943 screening for pulmonary tuberculosis by mass radiography was introduced in the United Kingdom on a national basis and proved very successful. The purpose of this screening was not only to provide treatment at an early stage but also, by removing the affected person from the community, to prevent the spread of infection. The control of the spread of communicable diseases was likewise the major objective of mass screening campaigns for parasitic infections in tropical and subtropical areas of the world.

As communicable diseases and the diseases of poverty declined in importance in the developed countries of the world screening trials became directed towards the detection of non-communicable conditions, such as diabetes mellitus, hypertension, ischaemic heart disease, iron deficiency anaemia and glaucoma.^{1,2} Screening for phenylketonuria was introduced in the National Health Service in the mid 1950s, for carcinoma in situ of the cervix in the mid 1960s, and for congenital hypothyroidism in 1982. This change of target was accompanied by a change in objectives, which are no longer to prevent the spread of disease and to bring benefits to the economy as a whole. Screening now is directed towards benefitting the individual by curing, or at least alleviating, his or her ill-health by less radical

Professor FV Flynn, MD, FRCP, FRCPath (President, Association of Clinical Pathologists, 1989-90).

treatment, by increasing life expectancy, or by providing reassurance should the screening test prove negative.

Differences between opportunistic and invitational screening

Activities such as the ordering of haematological and biochemical "profiles" on out-patients or patients just admitted to hospital, or the requesting of blood cholesterol concentrations or cervical cytology by general practitioners on patients attending their surgeries for unrelated health problems, are now classified under the term "opportunistic screening". Sackett and Holland pointed out that there are important differences between this type of activity and that of classic screening³; (1) the subjects are patients rather than well volunteers; (2) the process is initiated by the patient who sought health care, rather than a community health doctor or health authority; (3) the tests are performed for the purpose of making a comprehensive assessment of the patient's health rather than with the prime objective of detecting a specific disease at an early stage when treatment is either easier or more effective; (4) there is no implied guarantee that the patient will benefit from having such tests performed.

This address was presented to the Association of Clinical Pathologists on 9 November 1990 at the Barbican Centre, London.

Department of
Chemical Pathology,
University College and
Middlesex School of
Medicine, Windeyer
Building, Cleveland
Street, London
W1P 6DB
F V Flynn

Correspondence to:
Professor F V Flynn

Accepted for publication
30 November 1990

McKeown was one of the first to draw attention to the important difference in the ethical obligations implicit in ordinary medical practice and when screening is undertaken by invitation.⁴ When a patient seeks medical advice, the doctor's ethical position is relatively simple: he or she undertakes to do his or her best with the knowledge and resources that are available at that time. He or she cannot fairly be criticised when the state of medical knowledge or limitation of resources does not permit effective treatment of the patient's condition. The position is quite different in classic screening, when a doctor or health authority takes the initiative and seeks out people with no overt symptoms of disease and invites them to undergo investigation to see whether a particular condition is present. In this circumstance there is a presumptive undertaking not merely to identify the abnormality if it is present, but also to ensure that those affected will derive benefit from subsequent treatment or care.

Requirements for successful screening

The implicit undertaking to diagnose and provide effective treatment for those responding to an invitation to undergo screening profoundly affects the criteria which should be met before a screening procedure is introduced into medical practice. These must obviously include proof that the participants will benefit and that the total expenditure, including that required for diagnosis and long term care, can be justified. As long ago as 1966 Wilson proposed that to achieve successful screening for presymptomatic disease 10 criteria should be satisfied.⁵ His 10 guiding principles, which have not been improved on since, were as follows:

- 1 The condition sought should constitute an important health problem. The importance of the problem needs to be considered from the point of view of the individual and of the community, and weighting will need to be given according to the prevalence of the condition and the prognosis if early treatment is not instituted. Thus phenylketonuria with a birth incidence of only 10–15 per 100 000 warrants relatively uneconomic screening measures because of its serious consequences for the individual and the family.
- 2 There should be an accepted treatment for patients with recognised disease. This is a vital requirement, but there is also a need to know if treatment given at an earlier stage than normal will favourably affect the course of the disease and its prognosis.
- 3 Facilities for diagnosis and treatment must be available. It is very important to know in advance what the commitment is likely to be and that it can be met.
- 4 There should be a recognisable latent or early symptomatic stage. In other words there must be a reasonable period in the development of the condition during which symptoms are absent or minimal.
- 5 There must be a suitable test or examina-

tion. This is obviously a fundamental requirement.

- 6 The test or examination should be acceptable to the population. Acceptability is related to the nature of the investigation, the seriousness of the disease being screened for, and to the way in which the ground is prepared by health education. There seems to be no way, however, of gauging the public response before screening is attempted.
- 7 The natural history of the condition, including its development from latent to declared disease, should be adequately known. The pathological changes in the precursor stage should be recognisable and known to be progressive and responsive to treatment.
- 8 There should be an agreed policy on whom to treat as patients. Only those almost certain to need treatment should be classified as positive with the screening test, to avoid a risk of harm resulting to some participants.
- 9 The cost of the screening programme, including diagnosis and subsequent treatment of patients, should be economically balanced in relation to possible expenditure on medical care as a whole.
- 10 Screening should be a continuing process and not a once-and-for-all project. Only a small proportion of the chosen population is likely to be examined at a one-off screening and the procedure will only pick up those persons who happen at that particular time to have the abnormality sought. Continuing schemes are needed to cover complete populations and to establish the work as an accepted regular part of preventive medicine.

With such clear guidelines, it might be thought that the setting up of a screening programme would be straightforward, but I shall now turn to consider the many problems that have been encountered in the development of this form of preventive medicine.

Problems with screening tests

TEST SELECTION

An ideal screening test should detect only those who actually have the condition screened for and should not miss any cases, and the reliability of such a test in distinguishing diseased from healthy persons is best defined by its sensitivity and specificity. The sensitivity of a test describes its ability to give a positive result when the person tested truly has the disease, and is expressed as a percentage determined as follows:

$$\text{Sensitivity} = \frac{\text{Number of persons with the disease giving a positive test}}{\text{Number of persons with the disease who were tested}} \times 100$$

It provides a measure of the number of false negative results that will be given by the test; the lower the sensitivity the larger the number of false negative results. The specificity of a test describes its ability to give a negative finding when the person tested is free of the disease,

and is also expressed as a percentage figure determined as follows:

$$\text{Specificity} = \frac{\text{Number of persons without the disease giving a negative test}}{\text{Number of persons without the disease who were tested}} \times 100$$

It provides a measure of the false positive rate that will be encountered.

The sensitivity and specificity of each test proposed for screening purposes should ideally be established before its application, by means of a pilot programme carried out on a representative sample of the population.⁶ The whole of this sample needs to be investigated by the best available diagnostic procedures for the disease which the screening test is designed to detect. Having validated the test in this pilot survey, its subsequent application in the screening programme must be checked to ensure that its reliability is being maintained in routine operation. Because of the effort required and the difficulties encountered in establishing the sensitivity and specificity of screening tests, it is not surprising that the values are often not determined. The consequence for screening programmes of not having such information is very undesirable. For example, the inadequate sensitivity and specificity of urine testing for screening for diabetes mellitus was only discovered when it became apparent that cases were being missed and many others were being referred for glucose tolerance tests that proved normal.^{7,8}

Apart from reliably detecting persons with the disease and producing few false positive results, the ideal test should be safe, simple to administer, capable of use by paramedical and other personnel, quick to perform, relatively inexpensive, acceptable to the population, and it should give reproducible and reasonably accurate results.^{3,6} Unfortunately it is not always possible to satisfy all these requirements; for instance, tests providing greater accuracy may be more expensive, and ease of use is not always paralleled by reliability of performance, as has been shown with simple urine tests.⁹ Thus the Phenistix test that was used for screening for phenylketonuria had the advantage of being simple, easy to perform, readily accepted and cheap, but it missed between 25% and 50% of cases,^{10,11} and serum cholesterol measurements made with certain easy to use desk-top analysers, while precise, have been shown to be significantly lower than the values obtained with reference methods.¹² Reliability and acceptability are overriding requirements, but the final choice of test must often be based on compromise.

When there is a choice between a quantitative and qualitative test, the former is preferable, because performance can be monitored much better by quality control procedures, and where the screening programme is repeated the results can be compared, enabling small but significant trends to be detected.

ENSURING RELIABILITY OF TEST RESULTS

High standards of practice are essential throughout the preanalytical, analytical, and post-analytical phases of the testing procedure,

because a small deviation from the true value may be of much greater importance in the screening situation than during the routine investigation of patients. Standardisation of the conditions of specimen collection may therefore become of greater importance than in routine work. In this context the highly significant effects of venous occlusion, posture, and prior exercise on analyte concentrations in plasma must not be forgotten. In cervical cytology screening many smears are claimed to be technically unsatisfactory.¹³

Errors arising from mislabelling of specimens in the preanalytical phase or from clerical errors at the reporting stage cannot be picked up so readily in the screening situation because previous results will seldom be available to enable a delta check to be made. Consequently extra care has to be taken by those involved in the labelling of specimens and all other stages of the documentation.

Normal laboratory quality control measures and participation in external quality assurance schemes should detect poor performance arising during the analytical phase, but they need to be stringently applied to pick up the lesser deviations that can be tolerated in screening results. One of the problems arising in this context is the relative infrequency with which quality assurance specimens are available for analysis. This caused delay in the detection of a standardisation problem which was encountered in my own laboratory during the performance of serum α fetoprotein assays for a screening programme for Down's syndrome. In cytology there has recently been severe criticism of the lack of quality control, and schemes for periodically assessing the competence of individual screeners are now being introduced, but a serious difficulty is that agreement among competent observers is not good.¹⁴

Analytical imprecision when the test is performed by non-laboratory personnel outside the laboratory is a cause for concern.¹⁵ Recently Broughton *et al* reported on the scatter of cholesterol results obtained on quality control specimens by 69 general practices which were equipped with a sophisticated desk-top analyser¹⁶; on one serum specimen the values varied between 4.8 and 8.7 mmol/l and 8.6% of the results were 1 mmol/l or more adrift from the target value. This performance is worrying enough, but what will happen when high street testing gets under way is grim to contemplate.

Changes of method during screening programmes extending over prolonged periods of time can only be tolerated if comparability of the results has been unequivocally shown. Observer variation in the assessment of qualitative tests should, likewise, be kept to a minimum by avoiding changes in personnel.

Organisational problems

CHOICE OF POPULATION FOR SCREENING

The effectiveness of a screening procedure will be affected by the choice of population for scrutiny. In 1974 Holland concluded, after reviewing the evidence in a series of articles on

Table 1 Effect of disease prevalence on predictive value of positive result (% of true positives/all positives) for a test of 95% sensitivity and 95% specificity¹⁹

Disease prevalence (%)	Findings among a population of 10000					Predictive value of a positive result
	No with disease	No of true positive results	No without disease	No of false positive results	Total No of positive results	
1	100	95	9900	495	590	16.1
2	200	190	9800	490	680	27.9
5	500	475	9500	475	950	50.0
10	1000	950	9000	450	1400	67.9
20	2000	1900	8000	400	2300	82.6
50	5000	4750	5000	250	5000	95.0
75	7500	7125	2500	125	7250	98.3
100	10000	9500	0	0	9500	100.0

Table 2 Effect of disease prevalence on predictive value of a negative result (% of true negatives/all negatives) for a test of 95% sensitivity and 95% specificity¹⁹

Disease prevalence (%)	Findings among a population of 10000					Predictive value of a negative result
	No with disease	No of false negative results	No without disease	No of true negative results	Total No of negative results	
1	100	5	9900	9405	9410	99.9
2	200	10	9800	9310	9320	99.9
5	500	25	9500	9025	9050	99.7
10	1000	50	9000	8550	8600	99.4
20	2000	100	8000	7600	7700	98.7
50	5000	250	5000	4750	5000	95.0
75	7500	375	2500	2375	2750	86.4
100	10000	500	0	0	500	0.0

screening, that total population screening seemed hard to justify for any condition save phenylketonuria,¹⁷ but congenital hypothyroidism would be added now.¹⁸ The advantages of selecting a population at risk in terms of increasing the predictive value of a positive screening test result—that is, the likelihood that a subject found to have a positive test result will actually be found to have the disease—are shown in table 1. This quotes data from Vecchio¹⁹ relating to a test of 95% sensitivity and 95% specificity, and shows that the predictive value rises steeply as the prevalence increases. By contrast, the predictive value of a negative test result—that is, the percentage of all negative test results that will turn out to be true negative results—is not significantly affected by the prevalence of the disease in the population until it becomes very high (table 2). There is much to be said, therefore, for selective screening, inviting only those with recognised risk factors to take part, and the lower the specificity or sensitivity of a screening test the more important it becomes to select a high risk population. For instance, if screening for diabetes is contemplated it would be most profitable to concentrate on those over 50 years, those with a family history of the disease, and women who have had more than six children or infants weighing over 4.5 kg at birth.²⁰ Similarly, if cholesterol testing is to be considered in relation to the prevention of coronary disease it would be best to select those with a family history of cardiovascular disease under the age of 50, those with xanthomas, or xanthelasmata, or a corneal arcus at an early age, those who are 30% above their ideal weight, or who are diabetic.²¹ Examples of selective screening programmes applied to high risk groups include the screening of pregnant women for symptomless bacteriuria,²² the

regular cytological examination of urine specimens from workers in the rubber and dyestuffs industries for signs of malignancy,²³ and the haematological investigation of those working with ionising radiations for signs of marrow depression.

PUBLIC APATHY AND RETICENCE

Poor participation adversely affects the effectiveness of screening programmes and limited public interest seems to be common even when intensive publicity campaigns have accompanied the operation. Among screening trials in the United Kingdom, only 67% were persuaded to participate in the Bedford diabetic screening campaign,²⁴ 50% in a cervical screening programme in Aberdeen,²⁵ and, according to age group, between 36% and 53% in a screening project in a general practice in Hertfordshire.²⁶

In Britain between a third and a half of all women at risk of developing cervical cancer have never been screened.¹⁴ Factors that have been recognised as contributing to non-attendance for screening include ignorance of the availability of the test, fear of the test and the disease, lack of female staff to take the smear, inconvenient location and time of testing, problems of communication with those who cannot read or do not speak English, and failure to receive an invitation to have the test.^{14 27}

The failure of cervical screening to appeal to the women who need it most²⁸ has led to various attempts to persuade older women of high parity and low socioeconomic status to take part. Unfortunately high risk groups often embrace the socially inept who are the hardest to reach. Carruthers *et al* offered them a “do-it-yourself” vaginal pipette,²⁹ and Osborn *et al* nurses to visit them to take the smears at home,³⁰ but the advantage of the small extra yield of participants achieved by these approaches was offset by the decreased sensitivity of the testing and the loss of opportunity for diagnosing other diseases.³¹ Johnston has recently suggested that better information on the function and relevance of the test needs to be provided, that all eligible women should be personally invited to attend and, bearing in mind that child minding and time off work may need to be arranged, that a choice of appointment and venue should be offered.³² She also suggested that the fact that some women prefer to have the smear taken by someone other than their own doctor needs to be taken into account.

Recruitment into screening programmes might be improved by the opportunistic approach. In the context of blood cholesterol testing, in a study involving four centres in England and Wales, Mann *et al* found that the uptake of testing was 91–94% when the approach was opportunistic, and 36–63% when by invitation.³³ Other attractions of opportunistic screening are that the patient incurs no extra time or travel costs and that the general practitioner is saved the expense and trouble of identifying those who are eligible and issuing invitations. Opportunistic screening has the disadvantage, however, that

it will miss those who consult their doctor infrequently.

CHOICE OF SINGLE OR MULTIPLE SCREENING

From the participants' and organisers' point of view, and in the interests of economy, there are obvious advantages in screening for several diseases at the same time. There are drawbacks in this approach, however, when each screening procedure only gives a reasonable yield in a selected population group of particular sex, age, or occupation. It must also be recognised that the ratio of benefits to hazards offered by each test is not altered in multiple screening, and when several tests are combined, the benefits of the appropriate test for a particular participant may be offset by the hazards of the others. In many circumstances, therefore, applying several tests simultaneously may not be the best choice; indeed, after reviewing reports of multiphasic health screening procedures Knox³⁴ concluded that they seemed to be of little value, particularly in respect of those who were apparently well.

PROVIDING DIAGNOSTIC AND TREATMENT FACILITIES

The technology and expertise of the hospital service and the continuity provided by the general practitioner are essential for diagnosis, follow up, and treatment. The load on those involved in these aspects must be estimated in advance and the availability of the necessary resources assured. When necessary, the rate of screening must be adjusted to match the resources available. In early trials of screening, planning was often deficient in this respect. For instance, Keen reported that the blitzkrieg approach used in the Bedford diabetes screening trial resulted in a local epidemic of the disease and placed an enormous strain on all the facilities for many months.³⁵ If mass cholesterol testing based on current recommendations is introduced in the United Kingdom, the numbers of individuals who would require counselling regarding risk factors, or medical supervision of dietary or drug treatment, would exceed 50% of the population³⁶ and obviously existing facilities would be overwhelmed.

REPEATING SCREENING

Once started, screening should be a continuous process that progressively covers more of the chosen population, becomes organisationally more efficient, and is eventually accepted as part of normal medical care. Repeated screening is, however, critically dependent on a first class record system that ensures that the participants are reminded when they should reattend. The failure of the United Kingdom cervical cancer screening service, which aims at rescreening at three to five year intervals, is partly attributable to the deficiencies of the call and recall systems. Based on the experience in Tayside, these should improve after being computerised,²⁷ but in inner city areas the difficulties of maintaining up-to-date addresses on the register in the face of the considerable mobility of the population will remain a major problem; in Manchester and Salford up to a quarter of the addresses on

the register are incorrect.¹⁴ Unfortunately, even when participants can be contacted, it has been found that the response to recall is least among those at high risk.³¹

Interpretation problems

DECIDING WHICH RESULTS REQUIRE FOLLOW UP

The use of quantitative measurements for screening purposes has drawn attention to the difficulty of differentiating between results which require further investigation and those which do not. For most tests the results fall into a more or less continuous distribution curve, a bimodal distribution with a deep trough between two peaks that separates the results of those who are healthy from those who are diseased, rarely being seen. In ordinary medical practice the clinician can decide to ignore test values of doubtful clinical importance, but with screening there is no such fall-back position and a decision has to be made on where to draw the dividing line between further action and no follow up being required. A common solution has been to make use of a reference range.

One of the commonest approaches to setting the reference range for a test is to describe it in terms of the mean value plus or minus two standard deviations of the results obtained in a healthy population. Another approach, used when the distribution of results is not Gaussian, is to define it in terms of percentiles, the reference range then being quoted as extending between the 2.5 and 97.5 percentiles. With both methods there is a 5% chance that the results of a test on a perfectly healthy person will fall outside the reference range. In the early days of screening, more than 5% of healthy persons were liable to be misclassified as abnormal because single global reference ranges were used for all participants, the effects of physiological variables not being appreciated.³⁷ We now know that a multitude of reference ranges must be used, to allow for the influence of the participants' sex and age,³⁸ and that this only becomes practical if a computer is used to interpret the results.³⁹

Despite the refinement of allowing for sex and age effects, using a statistical basis to define "normal" ranges has dire consequences for screening programmes where multiple independent tests are used. Thus the probability that the results of two independent tests on a healthy person will both be in the reference ranges is 0.95×0.95 , and the probability that n independent tests on a healthy person will all be in the reference ranges is 0.95^n . It follows from this that if five such tests are carried out on a healthy person the chance of "abnormality" being found is 23%, with 10 tests the chances become 40%, and with 20 tests they would be 64%. The choice of using twice the standard deviation is, of course, quite arbitrary, and to lessen the burden of follow up that arises, Schoen and Brooks proposed that the "normal" range for each test should be defined as a function of the number of tests in the screening procedure, such that 95% of those who are healthy would always end up as being declared

Table 3 Action levels for blood cholesterol screening

Serum cholesterol (mmol/l)	Action
< 5.2	Reassure and repeat test in five years
5.2–6.5	Repeat test now to confirm result Counsel about diet and risk factors
> 6.5	Measure fasting serum lipids Institute dietary treatment and monitor response Consider drug treatment later if necessary

normal.⁴⁰ Thus if five tests were included the normal range for each test with a Gaussian distribution would be the mean plus or minus 2.57 times the SD, with 10 tests the mean would be plus or minus 2.8 times the SD, and with 20 tests the mean would be plus or minus 3.02 times the SD. This approach, of course, increases the chance that true abnormality will not be recognised. A more practical approach to the dilemma created by the use of statistically defined reference ranges is to routinely follow up all unexpected results falling outside the mean plus or minus 3 SD or 99.7% limits, and only to follow up unexpected results falling outside the mean plus or minus 2 SD or 95% limits if the need for such is suggested by other information. With a computer it would be quite easy to alert the screener as to which category the results fell into by flagging each accordingly.⁴¹

It is apparent that the use of reference ranges for deciding which results warrant further action leaves much to be desired. In practice the value beyond which diagnosis and intervention should automatically follow can only be decided from a detailed clinical knowledge of the natural history of both the treated and the untreated disease. The value chosen should reflect the point where currently accepted treatment begins to do more good than harm,⁴² but may also have to take into consideration the resources that are available for managing the newly diagnosed.³⁵ Sometimes two arbitrary dividing lines are drawn to reflect current uncertainties, the results then being classified into those requiring no further action, those requiring counselling and some follow up, and those definitely requiring further investigation. This was done in relation to diabetic screening by Keen,⁴³ and the European Atherosclerosis Society,²¹ and several other professional bodies have proposed action levels for cholesterol screening; a simplified version of the latter consensus proposals is given in table 3. The dividing lines should be set so that findings above the action limit have a high predictive value of abnormality, to minimise the number of false positive results, and even more importantly so that the "no action" results do not include false negative ones. The decision on the action level used ultimately rests, therefore, on a subjective judgment of many factors, including the validity of the test results, the seriousness of the disease being sought, the advantages of early treatment, and the resources available for follow up.

THE PHENOMENON OF REGRESSION TOWARDS THE MEAN

When a subgroup of the population who have been found to have abnormal results with a screening test have the investigation repeated, it will be found that the values obtained tend to redistribute themselves closer to the population mean.^{37, 42} This phenomenon of regression of extreme values towards the mean is often attributed to laboratory error,³⁷ but in fact it represents the inevitable consequence of the combined effects of inherent biological and analytical variation. The importance for screening is that some abnormal results will inevitably prove false positive, and that any treatment applied on the basis of the initial result may erroneously be judged to have been effective.

Benefit-related problems

ACHIEVING EFFECTIVE FOLLOW UP

A major problem for mass screening programmes is ensuring adequate follow up of those identified as being in need of care. After an interval of 18 months Keen reviewed what had happened to 117 unequivocal diabetic patients found in the Bedford diabetes mellitus screening programme, and concluded that about half had received inadequate attention.³⁵ Nine had not seen a doctor, 23 had been seen only once or twice, and 17 three times, and others had received inadequate treatment. More recently, Morgan reported that failure to act on abnormal urine test results found in outpatient clinics was common,⁴⁴ and Rastam *et al* reported that six months after 424 subjects had been referred for follow up of an increased plasma cholesterol, documented on two occasions, only 65% had been seen.⁴⁵ Similarly, disappointing follow up of cholesterol testing has been recently recorded by Wynder *et al*.⁴⁶ Clearly such failures of follow up are unacceptable as they negate the achievement of the promised benefits, but in some instances they may reflect negative second thoughts on the part of the participants.

KNOWING WHETHER TREATMENT OF RISK FACTORS IS EFFECTIVE

The identification and modification of a risk factor is inadequate evidence of the effectiveness of a screening programme. For instance, in cervical screening there is at present no way of determining which preinvasive lesions will become malignant, and estimates of progression and regression vary considerably.³² In coronary disease it has repeatedly been shown that the risk of developing and dying from myocardial infarction is directly proportional to the plasma cholesterol concentration,^{47–49} and it has recently been estimated that the change in coronary heart disease events following a 1% change in the blood cholesterol concentration is about 3%.⁵⁰ The apparent strength of this correlation, however, diminishes with increasing age, and the mortality from coronary heart disease increases continuously, without a threshold, over the entire range of plasma cholesterol concentrations

observed in western countries. Furthermore, most victims of myocardial infarction are not severely hypercholesterolaemic, and differences in cholesterol concentrations do not seem to account for the appreciably higher mortality from coronary heart disease in Scotland than in England.³³⁻⁵¹ The relation is complicated by the interaction of the various other risk factors, notably smoking, hypertension, and diabetes mellitus, which multiply the risk of coronary events. Recently Muldoon *et al* concluded, after evaluating six primary prevention trials, that interventions to lower cholesterol concentrations do not affect overall survival,⁵² and consequently it is perhaps not surprising that many have reservations about the adoption of a “population” cholesterol testing as opposed to an “individual high-risk” approach to prevention. The World Health Organisation,⁵³ the Committee on Medical Aspects of Food Policy,⁵⁴ and the Study Group of the European Atherosclerosis Society,²¹ however, all emphasised their complementary nature. The National Institutes of Health consensus development conference on cholesterol⁵⁵ recommended that opportunistic cholesterol testing should be the normal for all adults, and both the Working Group on Cardiovascular Disease of the Faculty of Community Medicine of the Royal College of Physicians⁵⁶ and the King’s Fund Consensus Conference on blood cholesterol measurement in the prevention of coronary heart disease⁵⁷ recently concluded that mass testing of blood cholesterol concentrations was not justified. Khaw and Rose have calculated that a national strategy aimed at reducing the mean population cholesterol by 0.5 mmol/l by changing national dietary habits should be more effective,⁵⁸ and such an approach should be more practical.⁵⁹ As drug treatments are costly and have some highly undesirable side effects⁶⁰⁻⁶² this approach has much to commend it. Genetic influences rather than dietary excesses, however, may be more important as an explanation of the universal hypercholesterolaemia found in western societies.⁶³ All seem agreed that any initiatives involving coronary heart disease prevention must attempt to control all modifiable risk factors, and give priority to those with a personal or family history of coronary disease.

OBTAINING COMPLIANCE WITH TREATMENT

Even when treatment can be counted on to delay or prevent the development of a disabling or lethal disorder, it cannot be assumed that persons who are free of symptoms will adhere to the regimen prescribed. Success at modifying cigarette consumption or food intake

generally leaves much to be desired and examples of poor compliance to therapeutic regimens abound—for example, with antibiotics for tuberculosis,⁶⁴⁻⁶⁵ and iron for anaemia.⁶⁶ Indeed, Sackett has shown that ambulatory patients are unlikely to take more than 50% of the prescribed medicines they receive, and that the amount of information which patients possess about their illness has almost no relation to their degree of compliance.⁶⁷ Such background knowledge demands a cautious approach to screening.

KNOWING WHETHER SCREENING REALLY ALTERS OUTCOMES

Often many years elapse before an answer to this question becomes available. Several evaluation studies of biochemical profiles performed routinely at the time of admission to hospital have been reported in this country, and very different success rates in the detection of unsuspected disease have been reported, varying from 0.5% to 8.3%.³⁷⁻⁶⁸ At University College Hospital we evaluated the procedure carried out on 2000 patients by comparing the clinical value of the 11903 tests that were requested with those of the 23069 tests which were added by the laboratory to complete the profile. The results are shown in table 4 where it can be seen that the yield of abnormal results with a bearing on diagnosis was almost twice as common among the requested tests than among those that were unsolicited—13.7% compared with 7.2%. In most cases such results were providing supporting evidence for an existing diagnosis, and less frequently they suggested a complication or a new diagnosis. Unsuspected disease was indicated by unsolicited tests in 3% of the patients, but only a minority of these derived any benefit because less than half the conditions found were treatable. As a result of our trial we concluded that the advantages of such opportunistic screening did not outweigh its disadvantages in the context of our teaching hospital practice.

Properly controlled trials to determine whether cervical smear programmes can reduce the incidence of invasive cervical cancer were unfortunately not performed before the service was introduced. A trial now would be unethical and consequently the only yardsticks that can be applied to measure its effectiveness are to follow the cervical cancer registration and mortality rates. Unfortunately the increased numbers of hysterectomies for non-malignant disease complicates the interpretation of the data. There is no evidence as yet that the service has been responsible for any reduction of the mortality in the United Kingdom.⁶⁹ Indeed, recently Villard, Murphy, and Vessey reported that there has been an increase in the cervical cancer mortality in England and Wales among 35 to 39 year olds.⁷⁰ Despite the failure to show any improvement in the death toll in the United Kingdom after 40 million smears,¹³ many are convinced, from the fall in the mortality in countries with well organised screening programmes that reach a high proportion of the eligible population, that an efficient system will reduce the incidence of

Table 4 Clinicians’ assessments of 11903 requested and 23069 unsolicited biochemical profile results obtained on 2000 patients at time of admission to University College Hospital

Percentage	Requested	Unsolicited
Considered abnormal	15.1	8.6
Considered abnormal with a bearing on diagnosis	13.7	7.2
Considered abnormal with a bearing on management	5.9	1.6
Considered abnormal and requiring further investigation	4.4	4.3
Considered abnormal and can be disregarded	2.8	2.5

Unsuspected disease was revealed by unsolicited tests in 3% of the patients

cancer appreciably.^{69 71} The failure to attract high risk women to attend for smear tests would certainly be expected to contribute to the apparent ineffectiveness of screening in this country,^{14 69} but uncertainties as to when and if treatment of precancerous changes is called for may also be important.^{28 32}

In the context of the current debate about whether population blood cholesterol testing should be undertaken, it is interesting to note that had the NHS adopted such a policy in 1975, the 25% fall in the mortality from coronary disease in men and women under 60 years that has taken place since then,⁵⁰ could mistakenly be assumed to be a direct consequence of such screening. Carefully controlled trials are essential before screening programmes are introduced.

The usual methods of measuring survival are unable to answer the question of whether treatment begun as a result of early diagnosis through periodic health screening is effective or worthless. An increased five year survival rate cannot be assumed to imply a benefit from earlier treatment, because as Feinleib and Zelen point out, it may merely reflect the fact that the starting point for the survival measurement has been put back from the usual time of diagnosis to the point at which early diagnosis can be achieved.⁷² Selecting an unusually early starting point for measuring survival guarantees an increased survival rate, even if the treatment does no good, and producing earlier diagnosis without delaying death is only acceptable if other adverse effects of the disease can be mitigated.

Studies of certain malignant diseases, such as cancer of the breast, colon, and rectum, have shown that those who have a long preclinical stage also tend to have a long clinical course, despite late treatment; conversely, patients with short preclinical stages tend to have a relatively short clinical course.⁷³ This correlation also complicates attempts at analysing the effectiveness of screening.⁷² Inevitably laboratory screening programmes will be more likely to pick up those whose disease has a long preclinical course, and consequently those diagnosed as a result of screening are virtually guaranteed longer clinical stages and better short time survivals than those diagnosed in the usual manner.

HARM RESULTING FROM SCREENING

Enquiry into the possibility of disease, when no symptoms are present, may arouse anxiety or depression. This is tolerable when treatable disease is actually found but what of those in whom the sequence of follow up tests, all negative, leads to needless worry that may never be fully allayed?⁷⁴ This should not happen if care is taken to explain the purpose of the exercise clearly beforehand, and the participants are informed that most of them will be found to be free of disease. An interview at an early stage with a doctor, and preferably the participant's general practitioner, and supportive follow up, seem to provide the best safeguard in this respect. In any event all testing should be under medical supervision.

False positive results may occasionally be responsible for physical morbidity among perfectly healthy participants as a result of referrals for diagnostic procedures, and false negative results may cause harm by causing some participants to ignore the symptoms they subsequently develop, thereby delaying diagnosis because of the false reassurance given.⁷¹ The lack of a clear dividing line between health and disease may be responsible for some participants with questionable abnormalities, who would not ultimately become ill, suffering as a result of being labelled as diseased or at high risk, or because of being subjected to unnecessary treatment with its attendant side-effects. Finally the screening test itself may constitute a hazard; this applies in the case of screening involving amniocentesis or repeated mammography.

Ethically there is an obligation to ensure that a screening programme does more good than harm to the participants. Obviously everything possible must be done to minimise iatrogenic side-effects from the therapeutic regimen prescribed, and to ensure that they and any detrimental psychological effects do not offset the benefits of screening.

Financial justification problem

It is important that expenditure on screening is not incurred to the detriment of provision of resources for the treatment of overt disease, as experience to date has shown that periodic screening programmes can have at best only a limited effect in the control of non-communicable disease. Because of the difficulties,⁷⁵ full cost-benefit analysis has seldom been undertaken and only financial costs and benefits are usually considered. Generally the cost seems to be high,^{50 76} especially when screening requires to be repeated regularly; for instance, cervical cytology screening costs about £50 million a year in the United Kingdom.⁷⁷ It must also be appreciated that an inevitable consequence of improving the survival of people with chronic disabilities is that the demands subsequently made on the health service will be increased. Only in the realm of severe untreatable congenital disorders has screening brought important financial savings,^{78 79} but such prenatal screening differs from that aimed at the detection of presymptomatic disease, in that far from being concerned with the prevention of disability and untimely death, it is undertaken with the intention of offering termination of pregnancy.

The recent report of the Standing Medical Advisory Committee to the Secretary of State for Health on blood cholesterol testing,⁵⁰ which provides a detailed economic appraisal of a variety of options for identifying and treating people at increased risk from coronary heart disease, does give grounds for believing that savings for the NHS might be realised if one of the more selective approaches to screening for coronary heart disease is followed. Some of the assumptions made for costing purposes could be challenged, however, and therefore no more than a large scale and carefully controlled trial to verify the estimates would seem justified in

the first instance. The report, which constitutes a good model for the future, provides estimates of the annual costs of the various programmes as well as the cost per "Quality adjusted life year" (QALY) gained, and of the annual number of lives that should be saved. Despite reservations about the validity of the QALY⁸⁰ and some of the costings, armed with the data of the cost-effectiveness of the alternative strategies and a knowledge of the resources available, less empirical and more rational decisions as to how to proceed should become possible in the future.

Conclusions and recommendations

Many formidable problems may be encountered when implementing screening for presymptomatic disease. The most fundamental of these relates to the difficulty of persuading the public that the procedure may be relevant and valuable to them. As a consequence the potential benefits are seldom fully realised in practice, and I would like to make the following recommendations:

- 1 Screening should only be undertaken when Wilson's 10 guiding principles can be satisfied.
- 2 Screening should not be adopted until carefully controlled trials have been fully evaluated, and its effectiveness has been established both in medical and economic terms.
- 3 Screening should only be undertaken under medical supervision.
- 4 Screening test procedures should be carried out to the highest standards and with meticulous quality control.
- 5 Selective screening of high risk subgroups of the population should be preferred.

- 1 Wilson JMG, Jungner G. *The principles and practice of screening for disease*. Geneva: World Health Organisation, 1968.
- 2 Cochrane AL, Holland WW. Validation of screening procedures. *Br Med Bull* 1971;27:3-8.
- 3 Sackett DL, Holland WW. Controversy in the detection of disease. *Lancet* 1975;ii:357-9.
- 4 McKeown T. Validation of screening procedures. In: *Screening in medical care: reviewing the evidence*. London: Oxford University Press, 1968:1-13.
- 5 Wilson JMG. Some principles of early diagnosis and detection. In: Teeling-Smith G, ed. *Surveillance and early diagnosis in general practice*. London: Office of Health Economics, 1966:5-10.
- 6 Whitby LG. Well-population screening. *Br J Hosp Med* 1968;1:79-91.
- 7 Keen H. The Bedford survey: a critique of methods and findings. *Proc Roy Soc Med* 1964;57:200-2.
- 8 Malins JM. The presymptomatic diagnosis of diabetes. *Proc Roy Soc Med* 1966;59:1175-7.
- 9 Kirkland JA, Morgan HG. An assessment of routine hospital urine testing for protein and glucose. *Scott Med J* 1961;6:513-9.
- 10 Wilson JMG. Evaluation of prescriptive screening for phenylketonuria. In: *Screening in medical care: reviewing the evidence*. London: Oxford University Press, 1968: 97-115.
- 11 Raine DN. Screening for disease: inherited metabolic disease. *Lancet* 1974;ii:996-8.
- 12 Kock TR, Mehta U, Lee H, et al. Bias and precision of cholesterol analysis by physician's office analyzers. *Clin Chem* 1987;33:2262-7.
- 13 McCormick JS. Cervical smears: a questionable practice? *Lancet* 1989;ii:207-9.
- 14 Smith A, Elkind A, Eardley A. Making cervical screening work. *Br Med J* 1989;298:1662-4.
- 15 Belsey R, Vandenbark M, Goitein RK, Baer DM. Evaluation of a laboratory system intended for use in physicians' offices. II. Reliability of results produced by health care workers without formal or professional laboratory training. *JAMA* 1987;258:357-61.
- 16 Broughton PMG, Bullock DG, Cramb R. Quality of plasma cholesterol measurements in primary care. *Br Med J* 1989;298:297-8.
- 17 Holland WW. Screening for disease: taking stock. *Lancet* 1974;ii:1494-7.
- 18 Grant DB, Smith I. Survey of neonatal screening for primary hypothyroidism in England, Wales and Northern Ireland. *Br Med J* 1988;296:1355-8.
- 19 Vecchio TJ. Predictive value of a single diagnostic test in unselected populations. *N Engl J Med* 1966;274:1171-3.
- 20 Malins JM. Screening for disease: diabetes. *Lancet* 1974; ii:1367-9.
- 21 Study group, European Atherosclerosis Society. Strategies for the prevention of coronary heart disease: a policy statement of the European Atherosclerosis Society. *Eur Heart J* 1987;8:77-88.
- 22 Asscher AW. Screening for disease: urinary-tract infection. *Lancet* 1974;ii:1365-7.
- 23 Crabbe JGS, Cresdee WC, Scott TS, Williams MHC. The cytological diagnosis of bladder tumours amongst dyestuff workers. *Br J Ind Med* 1956;13:270-6.
- 24 Sharp CL. Diabetes survey in Bedford 1962. *Proc Roy Soc Med* 1964;57:193-5.
- 25 MacGregor E, Baird D. Detection of cervical carcinoma in the general population. *Br Med J* 1963;ii:1631-6.
- 26 Hodes C. Screening in general practice. *Lancet* 1968;ii: 1304-6.
- 27 Robertson AJ, Reid GS, Stoker CA, et al. Evaluation of a call programme for cervical cytology screening in women aged 50-60. *Br Med J* 1989;299:163-6.
- 28 Randall KJ. Screening for disease; cancer screening by cytology. *Lancet* 1974;ii:1303-4.
- 29 Carruthers J, Wilson JMG, Chamberlain J, et al. Acceptability of the cytopipette in screening for cervical cancer. *Br J Prev Soc Med* 1975;29:239-48.
- 30 Osborn GR, Leyshon VN. Domiciliary testing of cervical smears by home nurses. *Lancet* 1966;ii:256-7.
- 31 Chamberlain J. Screening for the early detection of diseases in Great Britain. *Prev Med* 1975;4:268-81.
- 32 Johnston K. Screening for cervical cancer: a review of the literature. *Health Economics Research Unit discussion paper 04/89*. Aberdeen: University of Aberdeen, 1989.
- 33 Mann JI, Lewis B, Shepherd J, et al. Blood lipid concentrations and other cardiovascular risk factors: distribution, prevalence, and detection in Britain. *Br Med J* 1988; 296:1702-6.
- 34 Knox EG. Screening for disease: multiphase screening. *Lancet* 1974;ii:1434-6.
- 35 Keen H. Diabetes detection. In: Teeling-Smith G, ed. *Surveillance and early diagnosis in general practice*. London: Office of Health Economics, 1966:19-24.
- 36 Tunstall-Pedoe H, Smith WCS, Tavendale R. How-often-that-high graphs of serum cholesterol. *Lancet* 1989;ii: 540-2.
- 37 Bradwell AR, Carmalt MHB, Whitehead TP. Explaining the unexpected abnormal results of biochemical profile investigations. *Lancet* 1974;ii:1071-4.
- 38 McPherson K, Healy MJR, Flynn FV, Piper KAJ. The effect of age, sex, and other factors on blood chemistry. *Clin Chim Acta* 1978;84:373-97.
- 39 Flynn FV, Ball SG. Comprehensive computerized data management in a chemical pathology laboratory with SOCRATES. *Med Inf* 1982;7:275-305.
- 40 Schoen I, Brooks SH. Judgment based on 95% confidence limits. *Am J Clin Pathol* 1970;53:190-3.
- 41 Flynn FV. Biochemical screening—how? In: Williams R, ed. *Fifth symposium on advanced medicine*. London: Pitman Medical, 1969:259-67.
- 42 Sackett DL. Evaluation of health screening programmes by laboratory tests. Proceedings of the 8th International Congress on Clinical Chemistry. *Scand J Clin Lab Invest* 1972;29(Suppl 126):21-2.
- 43 Keen H. The presymptomatic diagnosis of diabetes. *Proc Roy Soc Med* 1966;59:1169-74.
- 44 Morgan AG. Is routine urine testing in outpatient clinics useful? *Br Med J* 1988;297:1173.
- 45 Rastam L, Luepker RV, Pirie PL. Effect of screening and referral on follow-up and treatment of high blood cholesterol levels. *Am J Prev Med* 1988;4:244-8.
- 46 Wynder EL, Field F, Haley NJ. Population screening for cholesterol determination. A pilot study. *JAMA* 1986;256:2939-42.
- 47 Martin MJ, Hulley SB, Browner WS, Kuller LH, Wentworth D. Serum cholesterol, blood pressure, and mortality: implications from a cohort of 361,662 men. *Lancet* 1986;ii:933-6.
- 48 Kromhout D, Bosschieter EB, Drijver M, Coulander CdeL. Serum cholesterol and 25-year incidence of and mortality from myocardial infarction and cancer. *Arch Int Med* 1988;148:1051-5.
- 49 Pekkanen J, Nissinen A, Puska P, Punsar S, Karvonen MJ. Risk factors and 25 year risk of coronary heart disease in a male population with a high incidence of the disease: the Finnish cohorts of the seven countries study. *Br Med J* 1989;299:81-5.
- 50 Department of Health. Report of a Working party of the Standing Medical Advisory Committee. *Blood Cholesterol testing. The cost effectiveness of opportunistic cholesterol testing*. London: Department of Health, 1990.
- 51 Jones IG. Prevention of coronary heart disease in Scotland. *Br Med J* 1990;301:180.
- 52 Muldoon MF, Manuck SB, Mathews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *Br Med J* 1990;301:309-14.
- 53 World Health Organisation. Prevention of coronary heart disease. *Report of a WHO expert committee. Technical*

- report series 678. Geneva: WHO, 1982.
- 54 Department of Health and Social Security. Diet and cardiovascular disease. Committee on Medical Aspects of Food Policy. *Report of the panel on diet in relation to cardiovascular disease*. London: HMSO, 1984.
 - 55 Consensus development conference. Lowering blood cholesterol to prevent heart disease. *JAMA* 1985; 253:2080-6.
 - 56 Smith WCS, Kenicer MB, Davis AM, Evans AE, Yarnell J. Blood cholesterol: is population screening warranted in the UK? *Lancet* 1989;ii:372-3.
 - 57 A King's Fund consensus conference. Blood cholesterol measurement in the prevention of coronary heart disease. *Lancet* 1989;iii:115-6.
 - 58 Khaw K-T, Rose G. Cholesterol screening programmes: how much potential benefit? *Br Med J* 1989;299:606-7.
 - 59 Tunstall-Pedoe H. Who is for cholesterol testing? *Br Med J* 1989;298:1593-4.
 - 60 Oliver MF. Hypercholesterolaemia and coronary heart disease: an answer. *Br Med J* 1984;288:423-4.
 - 61 Anonymous. Primary prevention of ischaemic heart disease with lipid-lowering drugs [Editorial]. *Lancet* 1988;i:333-4.
 - 62 Shetty HGM, Routledge PA. Adverse effects of hypolipidaemia drugs. *Adverse Drug React Bull* 1990;142:532-5.
 - 63 Grundy SM, Vega GL. Causes of high blood cholesterol. *Circulation* 1990;81:412-27.
 - 64 Dixon WM, Stradling P, Wootton IDP. Outpatient PAS therapy. *Lancet* 1957;ii:871-2.
 - 65 Fox W. The problem of self-administration of drugs; with particular reference to pulmonary tuberculosis. *Tubercle* 1958;39:269-74.
 - 66 Bonnar J, Goldberg A, Smith JA. Do pregnant women take their iron? *Lancet* 1969;i:457-8.
 - 67 Sackett DL. The magnitude of compliance and non-compliance. In: Sackett DL, Haynes RB, eds. *Compliance with therapeutic regimens*. Baltimore: The Johns Hopkins University Press, 1976:9-25.
 - 68 Leonard JV, Clayton BE, Colley JRT. Use of biochemical profile in children's hospital: results of two controlled trials. *Br Med J* 1975;iii:662-5.
 - 69 Anonymous. Cancer of the cervix: death by incompetence. [Editorial]. *Lancet* 1985;ii:363-4.
 - 70 Villard L, Murphy M, Vessey MP. Cervical cancer deaths in young women. *Lancet* 1989;ii:377.
 - 71 Chamberlain JM. Which prescriptive screening programmes are worth while? *J Epidemiol Community Health* 1984;38:270-7.
 - 72 Feinleib M, Zelen M. Some pitfalls in the evaluation of screening programs. *Arch Environ Health* 1969;19:412-5.
 - 73 Feinstein AR. Symptoms as an index of biological behaviour and prognosis in human cancer. *Nature* 1966;209:241-5.
 - 74 Marteau TM. Psychological costs of screening. *Br Med J* 1989;299:527.
 - 75 Pole JD. Economic aspects of screening for disease. In: *Screening in medical care: reviewing the evidence*. London: Oxford University Press, 1968:141-58.
 - 76 Pearson TA, Bowlin S, Sigmund WR. Screening for hypercholesterolaemia. *Ann Rev Med* 1990;41:177-86.
 - 77 Gray JAM, Chamberlain J. Cervical screening management. *Lancet* 1989;ii:280.
 - 78 Gill M, Murday V, Slack J. An economic appraisal of screening for Down's syndrome in pregnancy using maternal age and serum alpha-fetoprotein concentration. *Soc Sci Med* 1986;24:725-31.
 - 79 A report of the Royal College of Physicians. *Prenatal diagnosis and genetic screening*. London: RCP, 1989.
 - 80 Klein R. The role of health economics. *Br Med J* 1989; 299:275-6.