The ethics of screening: Is ‘screeningitis’ an incurable disease?

Darren Shickle and Ruth Chadwick University of Wales College of Medicine and University of Cardiff, respectively

Authors’ abstract
Screening programmes are becoming increasingly popular since prevention is considered ‘better than cure’. While earlier diagnosis may result in more effective treatment for some, there will be consequent harm for others due to anxiety, stigma, side-effects etc. A screening test cannot guarantee the detection of all ‘abnormal’ cases, therefore there will be false reassurance for some. A proper consideration of the potential benefit and harm arising from screening may lead to the conclusion that the programme should not be offered. A modified utilitarian approach may be used for allocation of scarce resources in health care. Ethics has an important role in this evaluation.

Most Western societies are concerned about health care expenditure in view of an ageing population and a growth in the number of effective health care interventions. Scarcity of resources means that health care systems are expected to produce more health gain with relatively fewer resources. It is assumed that if ‘prevention is better than cure’, then it will also be ‘cheaper’. Cost benefit analysis tends to consider tangible costs and benefits which accrue directly to the health service since these are easier to measure and are the only ones that will appear on balance sheets at the end of the financial year. However, there are ethical implications that arise from ignoring the intangible and indirect aspects of this equation.

Screening may be defined as a selection procedure for further investigation, applied to a population of asymptomatic individuals, with no personal or family history to suggest that they are at a higher risk of the disease than the rest of the population. It may be a form of ‘secondary prevention’, which aims to detect disease in presymptomatic individuals in order to provide more effective treatment in the early stages of disease, or ‘primary prevention’ which aims to identify risk factors or carrier states. Individuals offered screening would usually consider themselves to be ‘healthy’ (unless they have coincidental illness). We tend not to consider our health until either we or someone close to us loses it. The offer of screening implies that we may not be healthy after all. This may result in a rise in anxiety, even if small and short-lived, until the result comes back ‘negative’. This anxiety-state may be described as that of the ‘worried well’ (figure 1).

Consider a population, some of whom have a disease, while others do not. The diagnosis of ‘disease’ may require extensive investigation (or may only be made at postmortem). Such a ‘gold standard’ test is important prior to embarking on treatment. Screening should be a simpler, cheaper and quicker test to select for further investigation. However, cases with a positive screening-test finding may not have the disease (falsely labelled as ‘positive’) and similarly, disease may be present in individuals who have a negative screening result i.e. ‘false negatives’ (figure 2).

It is not always possible to make an absolute distinction between those individuals in whom the disease is present or absent even with a ‘gold standard’. For some conditions there will be a

Key words
Screening; ethics; sensitivity; specificity; non-maleficence, beneficence.
continuous distribution of variables, such that at one extreme, individuals could be considered ‘diseased’, while those at the other are ‘healthy’, for example, in the context of hearing or visual acuity. For other biological variables, for example blood pressure or haemoglobin, the extremes at both ends of the distribution would be ‘unhealthy’, while those in the middle would be ‘normal’.

For the purposes of screening there should be a threshold to trigger further investigations or treatment. The position of this cut-off in the distribution should be based on the associated risk of morbidity or mortality that warrants further intervention. The choice of threshold may therefore be arbitrary or depend on the resources available. For example, a blood cholesterol of 6.5 mmol/l or above, is considered to be associated with a significantly high risk of coronary heart disease (1). However, between 25 and 36 per cent of the UK population (1) have blood cholesterol above this threshold, and it would be unrealistic to treat this number of people.

A screening test with high specificity would tend to detect only people with the disease, i.e. there would be relatively few false positives. Raising the threshold would raise specificity, and hence target resources at those with the highest risk of morbidity and potentially the most to gain from treatment. In contrast, a test with high sensitivity would tend to maximise identification of diseased people in the screened population, i.e. there would be relatively few false negatives, but there would also be unnecessary investigations or treatment for others. Thus an increase in sensitivity will be at the expense of specificity and vice versa. There are costs and benefits associated with each of the four permutations of true/false positive/negative (figure 3).

When a screening programme is evaluated, the benefits of being a true positive tend to be considered, and sometimes the reassurance associated with a negative result. However, associated costs tend to be ignored. If costs outweigh benefits then either the balance between sensitivity and specificity should be altered or else the screening test should not be used at all.

The preference for specificity or sensitivity should depend on the consequences of making the diagnosis or not. High specificity should be desirable if:

1. the diagnosis is associated with anxiety or stigma;
2. further investigations are time-consuming, painful and/or expensive;
3. cases are likely to be detected by other means before it is ‘too late’ for effective treatment, and
4. treatment, especially if painful or expensive, is to be offered without further investigations.

In contrast, high sensitivity would be desirable if:

1. adverse consequences of missed diagnosis for the individual, for example late treatment, might be significantly worse than early;
2. adverse consequences of missed diagnosis for society, for example with a serious communicable disease, and
3. diagnosis is to be confirmed by other investigations, so that the period of anxiety is short, or the correct diagnosis is given before treatment is started.

For some screening tests the choice between sensitivity and specificity is difficult. Consider HIV (Human Immunodeficiency Virus) screening. The diagnosis is associated with high levels of anxiety and stigma, there may be no other investigations to refute the diagnosis, treatment is expensive, with many side-effects, and may be ineffective. However, there are important consequences of a missed diagnosis for society.

The ethicist has a role in resolving dilemmas arising from the conflict of sensitivity and specificity. Moral conflict can arise in circumstances where an intervention is associated with both good and harm for an individual. Utilitarianism is an attractive theory in such circumstances, since the correct solution would depend on whether the utility arising from the sum of the benefits outweighs the disutility associated with harm. The
Figure 3  Costs and benefits associated with being a true positive, false positive, false negative or true negative

<table>
<thead>
<tr>
<th>True positive</th>
<th>False positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs</td>
<td>Benefits</td>
</tr>
<tr>
<td>May be no increase in life expectancy if treatment is of no effect but patient may have to live longer with diagnosis because it is made earlier (lead time bias)</td>
<td>Earlier treatment (cheaper; more pleasant, better prognosis)</td>
</tr>
<tr>
<td>Anxiety: ‘Worried ill’</td>
<td>Benefits of ‘sick role’ (excused social responsibilities provided seek and comply with treatment)</td>
</tr>
<tr>
<td>Stigma</td>
<td>Explanation of symptoms: knowing the cause of symptoms so that they can be ‘treated’ can relieve anxiety</td>
</tr>
<tr>
<td>Reduced quality of life, for example due to side-effects or complications of treatment</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>False negative</th>
<th>True negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs</td>
<td>Benefits</td>
</tr>
<tr>
<td>False reassurance (‘Unworried ill’)</td>
<td>Spared anxiety if treatment of no benefit</td>
</tr>
<tr>
<td>May legitimise ‘unhealthy lifestyle’</td>
<td></td>
</tr>
<tr>
<td>Later treatment (unpleasant; more expensive; worse prognosis)</td>
<td></td>
</tr>
<tr>
<td>Increased cost per desired outcome</td>
<td></td>
</tr>
</tbody>
</table>

There are frequent appeals for funds to allow a child to receive a ‘life-saving’ treatment that is not available from the UK National Health Service. It is difficult for a health planner to withdraw funding from a programme that can save life because of the consequent publicity that would be generated through ‘shroud waving’. Vested interest groups would argue that the benefits of screening to detect one individual who may otherwise have died without early treatment, would justify any harm caused.

The principle of double effect permits harm to others provided the harm is indirect and was not intended, although it may have been foreseen. However, this principle has been widely criticised (3, 4, 5), and is unhelpful in the resolution of the dilemma of whether to screen or not. It can apparently excuse the harm caused both by providing and withholding screening. For example, it is accepted that all screening procedures are associated with false positives and negatives. This...
harm, however, although foreseen, can be described as an unintended side-effect. On the other hand, if screening is withheld, individuals who could have benefited from screening, via earlier treatment of their disease, will be denied these advantages. Thus the decision to withhold arguably causes harm which is foreseen but not intended.

The principle of acts and omissions might be invoked to escape responsibility for the harm that results from a decision to withhold screening. According to this principle, failing to prevent some harm is less bad, morally speaking, than actively bringing it about (6), and the withholder has not, it might be claimed, actively caused the harm. Like the principle of double effect, this principle has been heavily criticised (7). In the light of the evidence that haemophiliacs can be infected with the HIV virus as a result of transfusions of infected blood, for example, it is arguably just as bad to fail to screen the blood, with the result that, say, 10 per cent will be infected, as it would be to inject those 10 per cent with the virus deliberately. If we fail to prevent something bad from happening where we are in a position to do so, we are responsible.

Harris (8) identifies ‘positive’ moral responsibility where something happens because someone did something, and ‘negative’ responsibility because someone allowed or permitted or let it happen. He concludes that there is equal responsibility for what happens because of our actions and what happens because of our inaction or non-action. Thus, it would be an inadequate defence to argue against a screening procedure on the grounds of the need for a needle-prick or anxiety that may be caused to the ‘worried well’, since there is ‘negative’ responsibility for the ‘true positives’ who would otherwise have been identified.

**Unlimited moral demands**

The problem that arises from the abandonment of the principle of acts and omissions is that it seems to place unlimited moral demands upon us. It might indicate a moral requirement to carry out any screening programme that could conceivably prevent harm. This has led to the search for ways of placing limits on what we are required to do, by attempting to draw moral distinctions between different kinds of cases.

In the screening context it might be argued that it is morally significant that those who are harmed through screening are identifiable, whereas those that are harmed through failure to screen are not. For it is not possible to know which members of a population will be ‘true positives’ before screening, and whether these individuals will benefit from the treatment offered. Glover has shown that this consideration is irrelevant, morally speaking: whether the harmed individual can be identified prospectively or retrospectively makes no difference to the wrong of harming them (9). Harris maintains that an individual remains ‘negatively’ responsible even if he was not aware that he could have prevented harm, although he recognises that the individual should not be blamed.

Another attempt to limit the moral demands upon us is to try to establish priorities between different kinds of moral requirements. Frankena, for example, proposed that beneficence includes four elements (10):

1. One ought not to inflict evil or harm (what is bad).
2. One ought to prevent evil or harm.
3. One ought to remove evil or harm.
4. One ought to do or promote good.

Frankena ranked the four elements such that in cases of conflict, non-maleficence took moral precedence over beneficence. Foot’s proposed priority of what she calls ‘negative’ duties to refrain from injuring someone and ‘positive’ duties to help them is similar (11).

There are at least two problems in applying these distinctions to the decision concerning whether or not to carry out a screening programme. First, the primacy of non-maleficence is not absolute. For example, a surgical incision will cause injury, but the consequence of the operation may be that the patient’s life is saved. Thus, the injury caused by taking a blood sample as part of a screening programme may be considered trivial in comparison to the potential benefits. Second, it is not always clear whether a decision not to do something, such as carry out a screening procedure, should be described as a maleficient infliction of injury or a failure to do good. It is easy to imagine contexts in which the failure to screen blood before transfusion to haemophiliacs could accurately be described as a maleficient infliction of injury.

It is difficult, therefore, to accept any of these proposals for drawing moral distinctions. There must be room for a trade-off between the harm brought about by screening and the harm brought about by failure to screen. This suggests that cases should be considered on their merits. Even if we accept, as we should, that there is negative responsibility for the harm we fail to prevent by withholding screening, this does not mean that we should carry out every screening programme that could conceivably prevent harm. A useful test is that proposed by Singer: ‘If we can prevent something bad without sacrificing anything of comparable moral significance, we ought to do it’ (12).

There is a danger that advances in medical technology will be translated into screening services and offered by ‘enthusiastic’ scientists to an ‘unsuspecting’ public, without consideration of the consequences. Even if the society is consulted, there is a general expectation among the public that medical advance is generally ‘a good thing’.
The harm consequent from a screening test for any individual will usually be trivial, in comparison with the potential harm from not offering screening and saving a life by providing effective treatment. However, the way in which individuals perceive risk is complex: some are ‘risk lovers’ while others are ‘risk averse’ or ‘risk neutral’ (13). For example, the risk of an HIV-positive doctor infecting one of his/her patients is very small, yet there has been demand within the media for regular, universal HIV testing of health-care professionals. It has been argued that patients are more concerned with the nature of risk than with its precise numerical value: ‘let us not get so preoccupied with the risks that we lose sight of the stakes’ (14). However, individuals may place differing amounts of disutility on these states. For example, some patients may be happy to receive treatment for mild hypertension, even though there is only a small decrease in risk of a ‘stroke’ or ‘heart attack’ following treatment, other individuals may be less willing to accept the label of ‘ill-health’.

An individual may choose screening even though there is a high risk of relatively trivial harm and only a small chance that he/she will benefit. Using utilitarian principles, the net utility or disutility for society arising from an intervention is equivalent to the sum of the change in utility to its individual members. However, the sum of trivial disutility for the many may exceed the large utility gain for the few. Others have raised similar concerns about the potential net disutility associated with health promotion. For example, Skrabanek pointed out that ‘the proverb, “a stitch in time saves nine”, may be sound advice for mending socks but it makes little sense if a thousand people need one stitch (in its medical equivalent) to save one person from nine stitches’ (15). It is, however, unreasonable to expect the individual to include ‘opportunity costs’ in his/her consideration of providing consent, ie resources can only be spent once and hence benefits may have to be foregone by others.

Counselling should not be prescriptive

There are further problems in placing the responsibility of whether benefits outweigh costs, upon the individual. Counselling should, as far as possible, facilitate the autonomous decisions of the individual, for example allowing informed consent to be obtained for screening. It is unlikely, however, that any individual could absorb all the necessary information during a few counselling sessions, especially if this is during a time of anxiety. While counselling should not be prescriptive, the fact that a provider of health care places a positive value on screening such that the service is offered may give the counsellor the impression that the ‘correct choice’ of action would be to accept screening. Thus arguably, even if a counsellor attempts to give information in a neutral fashion, counselling cannot be truly ‘non-directive’ (16).

We would not wish to argue that screening should not be offered because information about risk, utility, opportunity costs etc is too complex, nor that we should not encourage individuals to include such considerations in their decision-making. Unless specified in law, we believe that screening should not be overtly directive, although some may disagree if the consequences of the condition in question are sufficiently serious. However, it should be recognised that the obtaining of apparently informed consent does not remove any responsibility for harm from those who offer screening. There is a duty on health-service planners however, to offer screening tests that the majority of the target population would wish to accept.

The principles proposed by Wilson and Junger (17) in the mid-1960s, have attracted much criticism. However, these criteria may be useful guidelines for the development of screening programmes:

1. The condition sought should be an important problem.
2. There should be an acceptable treatment for patients with recognised disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognised latent or early symptomatic stage.
5. The natural history of the condition, including its development from latent to declared disease, should be adequately understood.
6. There should be a suitable test or examination.
7. The test or examination should be acceptable to the population.
8. There should be agreed policy on whom to treat as patients.
9. The cost of case-finding (including diagnosis and subsequent treatment of patients) should be economically balanced in relation to the possible expenditure as a whole.
10. Case finding should be a continuous process and not a ‘once and for all’ project.

Wilson and Junger used the term ‘principles’ for ‘ease of description rather than from dogma’. It is unlikely that any screening programme will be able to fulfil all of these criteria to everyone’s satisfaction. The question therefore arises as to whether each criterion has equal merit, or whether there is a hierarchy of importance. Wilson and Junger felt that ‘of all the criteria that a screening test should fulfill, the ability to treat the condition adequately, when discovered, is perhaps the most important’. However, it is important not to be too restrictive in our thinking about what represents an ‘adequate’ treatment. For example, with some cases of prenatal diagnosis, couples may benefit from genetic...
counselling such that they may ‘prepare themselves’ for the birth of an affected child. In this circumstance, the counselling could be considered to be a form of treatment. An analogous question is whether termination of pregnancy can be considered a treatment, and if so for whom. However, the debate on that question has been explored elsewhere (18).

The requirement for an acceptable treatment is compatible with our argument for a proper consideration of all costs and benefits. The advocates of screening have a duty to demonstrate potential benefit prior to the introduction of screening. Improved prognosis following the use of an ‘effective’ treatment will usually be the most important of these benefits.

The importance of availability of facilities for diagnosis and treatment is self-evident. Informing patients that they have a treatable disease, but withholding any intervention is likely to cause more harm than not performing screening.

**Helpful pointers**

The remaining ‘principles’ are, in our view, helpful pointers as to which screening programmes are desirable and, if fulfilled, will describe situations where net utility is likely.

Wilson and Junger recognise the difficulty in defining ‘importance’ and include conditions with high prevalence, and those which have very serious consequences for the individual or society as a whole. Screening for a condition which is not ‘important’ is unlikely to result in net utility, or accrue more utility than other interventions competing for scarce resources. However, it may be worth considering screening for relatively trivial conditions, providing screening is cheap, does not cause anxiety, and there is an effective treatment with few side-effects.

Wilson and Junger do not, however, specify whether patients, a third party, or society as a whole, should prioritise importance. John Stuart Mill, in his essay *On Liberty*, argues that ‘there is no one so fit to conduct business, or to determine how or by whom it shall be conducted, as those who are personally interested in it’ (19). We have reservations about applying this principle of utilitarianism in the context of health-care resource allocation. It is difficult for an individual patient to be objective as to whether his/her health problem is more important than that of another patient or whether he/she deserves scarce resources in preference to others.

Attempts have been made to involve society in the prioritisation process, for example in Oregon, USA (20). However, such consultation was expensive and time-consuming, and the complexity of the exercise meant that there were inconsistencies in the initial priority-list. For example, treatment for crooked teeth was ranked higher than treatment for Hodgkin’s lymphoma. In view of these difficulties, we see little alternative to the existing decision-making process, provided that this is in consultation with experts and lay organisations. We recognise that this will be biased by the social values of those involved.

The criteria for suitability and acceptability of a test for screening will vary from one condition to another. While the test should have high validity and repeatability, a compromise between sensitivity and specificity may be required. A test associated with high costs (both tangible and intangible) may be acceptable to an individual or society as a whole, if the condition is severe and an effective treatment is available. For example, most women will endure the indignity and embarrassment associated with taking a cervical cytology smear, if it means that carcinoma of the cervix may be prevented.

The requirement for an understanding of a condition’s natural history and the recognition of a latent stage, are important in the development of techniques for diagnosis and treatment, but in themselves are not essential. Screening for HIV is performed, yet the natural history of acquired immunodeficiency syndrome (AIDS) is still not ‘adequately understood’.

The balancing of economic costs resulting from case-finding with those due to possible expenditure as a whole, is compatible with our modified utilitarian approach and with the principle of justice.

If screening is available for one individual, then justice would expect screening to be offered for another individual who has a similar risk of having the disease and a similar potential to benefit from treatment. If a screening programme is considered to be desirable, why then should future patients not benefit from screening? Wilson and Junger argue that screening should be a continuous process, since the ‘start-up’ costs associated with screening (providing accommodation, purchasing equipment, training staff etc) are usually large in comparison with the marginal costs of each test performed. While these points are true, screening for prevalent (existing) cases requires different considerations from screening for incident (new) cases. Thus, the case-yield and the consequent utility may be high in the first wave of screening if the natural history of a condition is slow and hence there may be a large number of undetected cases in a population. However, the number of new cases that develop during a screening cycle may be small and hence the consequent utility for these few cases may not outweigh disutility or opportunity costs.

**Wilson and Junger criteria endorsed**

An assessment of net utility will assist in choosing service options which will result in the ‘most happiness’ for the population. We recognise that assigning utility values within the cost/benefit equation is difficult. Obtaining a consensus as to the
objective of a screening programme is an essential first step. For example, what is the purpose of genetic-carrier screening – empowerment of the individual in making reproductive choice or reducing the number of births of handicapped individuals?

We endorse the Wilson and Junger criteria, since if the majority are fulfilled (especially those relating to effectiveness of treatment) then the conclusions will approximate to those of a utilitarian assessment.

The principle of justice might suggest that access to health care is a moral right for all. However, scarcity of resources means that there must be prioritisation. There is no reason why philosophers should not be interested in the resolution of moral dilemmas in order to maximise utility from limited health care resources. Such choices have to be made, and it is better that such decisions are made with an explicit, ethical mechanism rather than them depending on the values of a few health service planners.

Darren Shickle, MB, BCh, is MRC Special Training Fellow in Health Service Research, at the Centre for Applied Public Health Medicine, the University of Wales College of Medicine, Cardiff. Ruth Chadwick, DPhil, LLB, was Director of the Centre for Applied Ethics at the University of Cardiff at the time the paper was written. She is now Professor of Moral Philosophy at the Centre for Professional Ethics, University of Central Lancashire.

References


News and notes

Genetic Screening Symposium

The Centre of Medical Law and Ethics jointly with the Nuffield Council is holding a symposium on genetic screening in the Council Room at King's College London, Strand Campus, on 22nd March. The symposium will consider the implications of the Nuffield Report on Genetic Screening. The programme promises an informed and prestigious array of speakers, which includes: Professor Dame J Lloyd, Professor Ian Kennedy, Sir Patrick Nairne and Professor Margot Stacey.

Those wishing to attend should write to the centre together with the appropriate fee. They may also phone or fax the Administrator: Tel: 071-873 2382; fax: 071-873 2575. Please note: all cheques are to be made payable to King's College London.

The cost is £50, lunch and refreshments included.