LECTURE 9: PREVENTION STRATEGIES

Learning objectives

By the end of this session, participants should be able to

i. Utilise the concepts of relative risk (rate ratio, risk ratio, or odds ratio) and attributable risk and be able to anticipate the potential benefit of interventions at the individual and population levels.

ii. Appreciate the three levels of prevention: primary, secondary and tertiary

iii. Have an appreciation of two different strategies, directed at the `high risk individual' or the population as a whole, which may form the basis of primary prevention activity.

iv. Describe the implications for public health policy of undertaking screening.

v. Define sensitivity, specificity and predictive value of a test.

vi. Explain the relative importance of these values (i.e. sensitivity, specificity, predictive value) in the context of screening programmes and epidemiological studies.

vii. Describe possible biases in ascertainment of screening benefit.

viii. Understand the advantages and disadvantages of different study designs in evaluating a screening programme.

1. Introduction

Up to now, much of this course has focused on determining how important a particular exposure is in relation to an outcome of interest (the effect of the risk factor on those exposed to it). In addition we have examined how one might assess whether a particular treatment is effective by comparing the incidence of a particular outcome (e.g. death) among those who do or do not receive the treatment in question. The focus has, therefore, been on measures of **relative risk**. *Note: The term 'relative risk' is a general term which refers to a risk ratio, rate ratio, prevalence ratio or odds ratio.*

In this lecture, we examine the impact of an exposure on the whole population, and its public health importance, which is primarily determined by the **attributable risk** of the exposure and its frequency in the population.

Following a discussion of **attributable risk** and related issues, primary prevention and the importance of determining whether to focus attention on 'high risk' individuals or on populations will be discussed. We also consider the value of tackling diseases at different stages of the disease process.

2. Attributable risk

Much of epidemiology is concerned with identifying the **risk factors** for a disease, health problem, or state of health. In comparing the strength of an association with a particular outcome, relative to the absence of this exposure, we identify the measure of effect known as the relative risk (see Lecture 1, ratio measures of effect). The **relative risk** (RR) measures the relative potency of the risk factor while the **attributable risk** is a measure of the public health effect attributable to the risk factor.

Table 1. Key definitions in attributable risk and attributable risk fractions

EXPRESSION	QUESTION	DEFINITION	CALCULATION
Relative risk (risk ratio)	How many times more likely are exposed persons to become diseased, relative to those not exposed?	Relative risk represents the likelihood of disease in exposed individuals relative to those who are non-exposed.	RR is the risk in the exposed (R_e) divided by the risk in the unexposed (R_o) $RR = R_e/R_o$
Attributable risk (risk difference) *	What is the risk of disease attributable to the exposure in the exposed group	The risk of disease in the exposed group that can be considered attributable to the exposure, after taking account of the risk of the disease which would have occurred anyway.	AR is the risk in the exposed (Re) minus risk in the unexposed (R_o) $AR = R_e - R_o$
Attributable risk percent (exposed) or aetiologic fraction*	What proportion of the disease risk in the exposed group is due to the exposure?	The proportion of disease among the exposed that is attributable to the exposure.	AR% is the attributable risk divided by the risk of disease among the exposed (multiplied by 100). $AR\% = [(R_e - R_o)/R_e] \times 100$
Population attributable risk*	What is the risk of disease in a population which is associated with the occurrence of the risk factor?	An estimate of the excess risk of disease in the total study population of exposed and unexposed individuals which is attributable to the exposure.	PAR is the risk of the disease in the whole population (R_t) minus the risk in the unexposed group (R_o). $PAR = R_t - R_o$
Population attributable risk percent (or population attributable fraction)*	What fraction (proportion) of disease in a population is attributable to exposure to a risk factor?	The proportion of disease in the study population that is attributable to the exposure (and potentially could be eliminated if exposure were eliminated)	PAR% (or PAF) is the population attributable risk divided by the risk of the disease in the population (R_t). $PAR\% = (PAR/R_t) \times 100$

[* **Note**: Risks, rates or prevalences can be used for these calculation but they need to be consistent across the different measurements and need to be clearly specified. Here we have used risk of disease and have labelled the items in the formulae appropriately i.e R_e (risk in the exposed population), R_o (risk in the unexposed population), and R_t (risk in the total population).]

Relative risk answers the question: How many times more likely are exposed persons to get the disease relative to non-exposed persons? Relative risk tells us about the **strength** of the association between exposure and disease, and is an important consideration in establishing whether a particular exposure is a cause of the outcome of interest. However, it tells us nothing about the **magnitude** of absolute risk (incidence associated with a particular exposure). Even for factors with large relative risks the absolute risk may be small if the disease is uncommon.

If we assume that a factor is causal, it is of value to express the magnitude of its influence on the occurrence of disease in a given population. If we have information on the usual **risk** (or **rate**, if we have that available) of a particular disease in the absence of the exposure, as well as in its presence, we will be able to determine the **attributable risk** (also known as excess risk or risk difference) associated with the exposure (see Table 1).

Attributable risk (AR) is, therefore, the *additional* risk of disease related to the exposure in question. For instance, what is the additional risk of lung cancer that someone has because he/she is a smoker. It acknowledges the background risk of the disease which is present in the population and due to other causes, some of which may be unknown to us. **Attributable risk** answers the question: What is the risk of disease attributable to exposure? It measures the difference between the risk among those exposed and those who are not exposed. Because of the way it is calculated, attributable risk is also called the **risk difference**. It is a useful expression of risk in the clinical setting because it represents the actual, additional probability of disease given the exposure concerned. A related concept is the **attributable risk percent** (AR%) or attributable fraction (also known as the aetiologic fraction). This describes the proportion of disease in the exposed group, which is attributable to the exposure.

The **population attributable risk** is a measure of the excess risk of disease in a total study population which is attributable to an exposure. This measure is useful for determining the relative importance of exposures for the population and is calculated by subtracting the risk of the disease in the unexposed group (R_o) from the risk in the whole population (R_t). It can also be calculated by multiplying the risk difference by the proportion of the population exposed. Population attributable risk answers the question: What is the risk of disease in the whole population which is associated with an exposure? For instance, what is the increased risk of lung cancer in the population because the population includes smokers?

Finally, we can indicate what proportion of the total risk of the condition is associated with the exposure: this is expressed as the **population attributable risk percent (**PAR% or PAF**)** or the population attributable fraction**.** This is the population attributable risk (the risk of disease in the whole population minus the risk of disease in the unexposed group) divided by the risk in the population as a whole. Population attributable risk percent answers the question: What proportion (fraction) of the disease observed in the whole population is attributable to exposure to a risk factor? It indicates what proportion of the disease experience in the population could be prevented if exposure to the risk factor was eliminated.

Consider the following hypothetical example (Table 2):

Table 2 Risk of backache in female workers aged 20-44 years by work posture

Work posture	Risk of backache per 100 female workers over two year period
Standing female workers	12.3
Other female workers	7.7
All female workers	8.3

Assuming that a causal relationship exists between standing and backache, we may wish to establish how important this factor is in relation to all cases of backache in women. We will assume that the women who stood while working would have had a risk of backache of 7.7 per 100 persons over the two year period (i.e. the same as other workers) if they had not been standing at work, instead of 12.3 per 100 persons over the two year period which we observe. The attributable risk (or excess risk or risk difference as it is also known) resulting from working in the standing position, is 4.6 per 100 persons (12.3 - 7.7 = 4.6) over the two year period. Expressed as a proportion of the total risk in women who stand at work, this is 4.6/12.3 = 37.4%. In other words, 37.4% of the risk of backache in standing women workers, over the two year period for which they were observed, can be attributed to their standing. This is the attributable risk percent (AR%) and tells us what proportion of disease in the exposed population is due specifically to the exposure in question (standing).

Similarly, the risk of backache in all women workers would have been 7.7 per 100 persons over the two year period instead of 8.3 per 100 if nobody had been standing at work. The population attributable risk, therefore, is 8.3 - 7.7, which is 0.6 per 100 persons over the two year period: this is the risk of backache which is related to standing at work. If we want to determine how important this is to the whole population of working women, we can consider the population attributable risk percent (PAR% or PAF): this is (8.3 - 7.7)/8.3, or 7%. In other words, in all female workers, the proportion of risk of backache which results from standing posture is 7%. This also tells us what proportion of backache in working women may be possible to avoid if we stopped women having to stand at work. We may also speak of this as the **preventable fraction** which is the proportion of the observed incidence which could be prevented by removal of the risk factor.

There are a number of assumptions upon which these measures of attributable risk are made: the main one is that the calculation of measures of impact assume that all of the association between the risk factor and the disease is causal; i.e. that there was control of all confounding. Other assumptions are also necessary: that both risk factor and frequency of disease were measured accurately, that removal of the risk factor actually removes the risk (i.e. that those previously exposed and now unexposed have the same risk as those never exposed) and that the risk factor is actually removable.

Summary risk relationships are generally derived from **case control, cohort** or **intervention** studies in which deaths or new cases of disease contribute (directly or indirectly) to the relevant calculations of RR or AR. These data are often derived from studies of sub-sets of a particular population (eg. manual workers, immigrant populations) and may be quite different from the populations to which the data are applied. Both levels of exposure and intrinsic susceptibility may be different within and between populations: it is therefore important that the extrapolation of data from study in a particular population to another population is undertaken with great caution.

The attributable risk percent (AR%) can also be calculated from knowledge of the relative risk (RR) among those exposed and can be expressed as (RR-1)/RR multiplied by 100. [You can work this out for Table 2 to confirm this relationship; use this as an opportunity to remind yourself of the definition of relative risk]. Similarly, the population attributable risk percent (PAR% or PAF) can be expressed as [p(RR-1)] / [p(RR-1) +1] or (RR-1)/ (RR +1/p -1) where p is the proportion of the total population that is exposed.

3. Approaches to prevention

Prevention includes a wide range of activities aimed at reducing risks or threats to health. In this section we are introducing three categories of prevention (primary, secondary and tertiary), the prevention paradox, the importance of the distribution of exposures/risk factors in the population and how this is important when adopting about high-risk vs population approaches.

3.1. Levels of prevention

Primary prevention aims to prevent disease or injury before it even occurs, including by controlling exposure to risk factors, for instance by preventing exposure to lead in the environment, or by reducing smoking through taxation or banning smoking in public spaces. An example in relation to COVID-19 could include encouraging hand washing and mask wearing, reducing close, in-person contact or vaccination (which helps prevent severe COVID-19 symptoms). Some strategies were deployed population wide by governments, such as mandating stay-at-home orders, lockdowns, and school closures, while others were undertaken at the personal level, e.g., frequent hand-washing with soap and water (or hand sanitiser) or maintaining a physical distance with other people

Secondary prevention seeks to reduce the impact of a disease or injury that has already occurred. This is done by detecting early departures from health in order to introduce timely and appropriate treatment/interventions to halt or slow the disease's progress, to encourage personal strategies to prevent re-injury or recurrence, or to implement programmes to return people to their original health and function.

For example, general practitioners could measure the lung function of patients who smoke and advise them to stop smoking as soon as there is a departure from normal function. Other examples of secondary prevention include: regular examination and screening tests to detect disease in its earliest stages (e.g. mammograms to detect breast cancer), daily, anti-hypertensive or lipid lowering medication to prevent heart attacks or strokes, and suitably modified work so injured or ill workers can return safely to their jobs. There can be prerequisites for an appropriate screening programme (further discussion below).

Finally, **tertiary prevention** seeks to treat and rehabilitate those with the disease, toward reducing complications, restoring functions, and ensuring a better quality of life even though disease has occurred. It can also aim to reduce risk of death from the disease. Some examples are: cardiac or stroke rehabilitation programs, chronic disease management programs (e.g. for diabetes, arthritis, depression, etc.), support groups that allow members to share strategies for living well. Tertiary prevention may entail regular monitoring of someone with, for example, chronic obstructive airways disease, and ensuring prompt treatment of any respiratory infections which arise. In relation to insulin-dependent diabetes, providing appropriate care will greatly reduce the complications and mortality from the disease. Secondary and tertiary prevention facilitate modification of the disease process after the process has already begun, and seek to ensure a better quality of life even though disease onset has occurred.

Table 3. Levels of prevention

Primary Prevention	Protection of health by personal and community-wide efforts to prevent disease before it occur and through limiting exposure to risk factors.
Secondary prevention	Measures available to individuals and populations for the early detection and prompt and effective intervention to detect departures from good health (e.g. screening).
Tertiary Prevention	The application of measures to reduce or eliminate long-term impairments, death, and disabilities, minimizing suffering caused by existing departures from good health and to promote the patient's adjustment to his/her condition.

3.2 The prevention paradox

The essential paradox of disease prevention strategies is that since most diseases are relatively concentrated in certain high-risk individuals, and are much rarer in the rest of the population, **when most individuals adopt a behaviour designed to lower their risk of disease (or the more severe consequences of the disease), substantial benefits are seen only in a few high-risk individuals and the rest of the population will not get direct benefit**. Thus, for example, any person's decision to lose some excess weight may have only a small impact on that person's risk of serious disease in the near future. **In the case where many people each lose a little amount of weight, shifting the whole distribution of risk/exposure to risk in a favourable direction, this may have a substantial impact upon the community's experience of obesity-related disorders as a whole.** The fact that individuals may benefit little by undertaking a preventive measure, while the populations from which they come may benefit greatly, in terms of risk reduction, is known as the **prevention paradox** (Rose, 1992). In essence, if a large number of people each reduce their risk slightly, the entire population may show a large reduction in risk.

3.2.1 Distribution of the risk factors/exposures in the population

How can a preventive strategy have little effect on the risk experienced by an average individual, but have large benefits to a community?

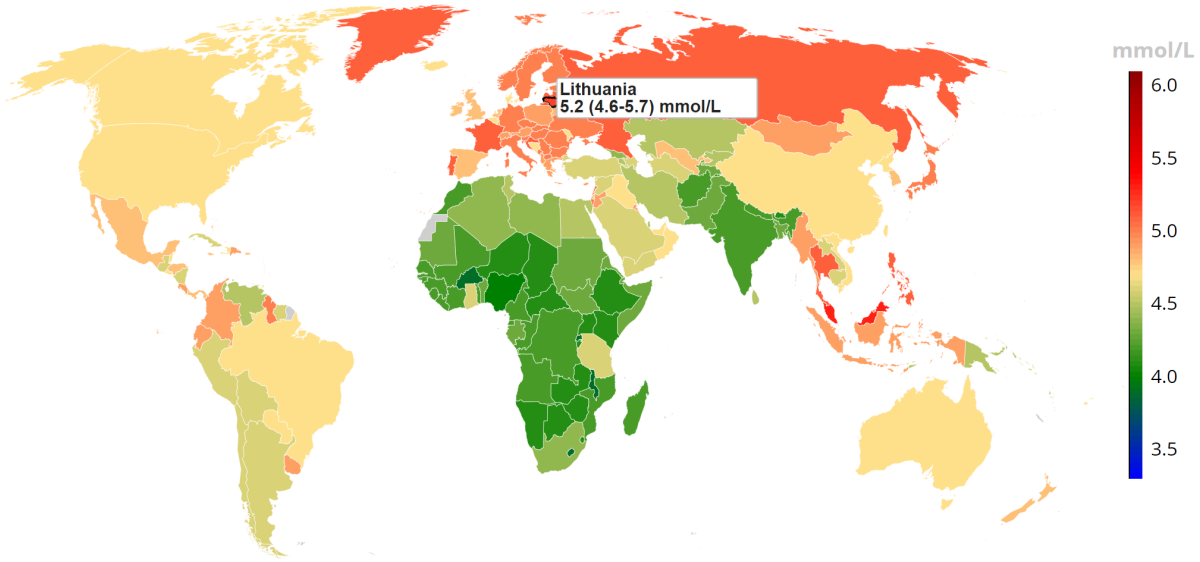
Before answering this question, Rose argued that we should first make the distinction between the determinants of disease in individuals, and the determinants of population incidence rate of a disease (Rose, 1985). That is, “Why do some individuals have condition X?” is fundamentally different from “Why do some populations have more cases of condition X, whilst in others it is rare?”.

The answer to the first question may lie in genetic variation, or individuals having a greater or lesser exposure to environmental and behavioural risk factors). In any case, one might achieve a complete understanding of why individuals vary (and tend more to have the condition), and still be missing the public health question – “Why is condition X common in population A, and rare in population B?”. The answer to which has to do with the characteristics that distinguish the two populations. According to Rose, in order to prevent and control disease in a population, we need to find and act on the determinants of prevalence and incidence rates overall, rather

than simply focus upon factors which explain why risk varies between individuals within a single population.

Figure 1 shows the national mean total cholesterol level (in mmol/L) among male adults aged 18 years or older in 2018. In each individual country, we can expect there to be some males to have hypercholesterolaemia – commonly defined as elevated levels of cholesterol. One could investigate the genetic and other causes of high cholesterol levels (e.g., physical activities, drinking or smoking) using the variation in exposure among individuals; but if we want to discover why, for example in Lithuania, men had higher mean total cholesterol than all other countries, we can look for national characteristics such as the national diet (or exercise patterns or genetics among other things) which shift the whole distribution of the exposures.

Figure 1. National mean total cholesterol (mmol/L) of male adults aged 18 years or older



Source: https://ncdrisc.org/cholesterol-tc-distribution.html

In a situation where everyone in the population has a genetic predisposition (yes vs. no) for hypercholesterolaemia, we can do very little to understand the influence of this "risk factor" as exposures is locally uniform - we cannot measure any variability in exposure status and thus its effect to the population. Water supply quality and its effect on health may be another example as exposure to it tends to be locally uniform in a country or subnational region. in such a situation, one would need to compare populations with different levels of the exposure instead.

The consequences of prevention strategies may differ depending on whether the focus is on the individual or at population level. The approach to disease prevention targeted mainly high risk, seeks to protect those people. To prevent disease in the population, by contrast, we control the determinant of incidence, lower the level of risk factors, and shift the whole distribution to a more favourable direction.

3.2.2 Examples of high risk and population prevention strategies

"Population strategies" seek to control the determinant of incidence, but may either offer little benefit to each individual, or at least appear to be less directly beneficial. As Rose has argued

an intervention aimed at reducing mean blood pressure in a population by reducing the salt in processed food in shops, and hence that may reduce mean blood pressure by a relatively modest amount (2-3 mmHg), might result in a larger reduction in the population incidence of strokes and other cardiovascular events than using drugs to reduce the blood pressure of the small fraction of people who have extremely high blood pressure. This is despite the fact that a person in the high-risk group that is targeted with drugs for high blood pressure will have their own probability of a cardiovascular event reduced very substantially.

Mass immunization against infectious disease provides an example of this. Measles is a highly contagious virus that lives in the nose and throat mucus of an infected person. It can spread to others through coughing and sneezing. Measles vaccination resulted in a 73% drop in measles deaths between 2000 and 2018 worldwide, according to WHO figures[1]. In some areas in the world, widespread use of measles vaccine has drastically reduced the rates of cases and deaths[2]. As a result, people in these places may rarely ever meet someone with measles, and are therefore at low risk of contracting the disease even if they are not immunized themselves. In other words, the mass vaccination changed the distribution of exposure in the population by limiting contact to contagious people. When there are *no* or *very few* new cases, some people may ask: Why do we need vaccinations still/again? Therein lies the prevention paradox. The prevention strategy has worked so well, that there are no or very few people around with measles to "demonstrate" the alternative.

Population strategies, however, are not necessarily in competition with targeted, high-risk, strategies. The two have different focuses and different characteristics. One positive approach to targeting high risk individuals is that only those targeted need to bear the costs of the prevention strategy, and not everybody. By contrast when adopting a population level approach, it is even more important to monitor side effects and risks of the prevention strategy (e.g. the negative effects of false positive diagnosis of breast cancer screening has be weighed against consequences of not catching breast cancer early enough). On the other hands, one drawback of targeting prevention at high-risk individuals is having a limited overall potential. The following numeric example of myocardial infarction illustrates this well.

A variety of risk factors have been identified for myocardial infarction (MI): these include raised blood pressure, smoking, and high cholesterol. The risk of death from MI rises the more risk factors the individual has. As indicated in Table 4, the proportion of men suffering a MI is greatly increased in the presence of risk factors, and is markedly more likely in the presence of signs of early disease. Although the relative risk is very much higher in those with elevated risk factors and early signs of disease, only 12% of all MI occur in this group, and 88% of MI occur in other men (men without both elevated risk factors and signs of disease).

[1] https://www.who.int/news-room/fact-sheets/detail/measles

[2] https://www.cdc.gov/measles/about/history.html

Table 4. Predicted risk of myocardial infarction in the next five years in relation to risk factors for heart disease.

Screening result	Percentage of men	Percentage subsequently suffering myocardial infarction	Percentage of all attacks occurring in this group
All men	100	4	100
Elevated risk factors only	15	7	32
Elevated risk factors plus early disease	2	22	12

Source: Cited in Rose, 1992.

If all prevention efforts focus on those at highest risk, the public health impact (i.e. the overall number of deaths averted in the population) would be small - only 12% of all cases would be prevented even if the prevention effort was fully successful. Even if prevention efforts extend to those who are at higher risk, but without early disease, only 32% of cases would be prevented and 68% of cases of MI would still remain. In this example, **people at the highest risk do not make up the majority of disease burden**. For this reason, population-wide strategies are often an integral part of public health approaches to prevention: if most men aged 55-64 years altered their risk factors (e.g. reduced their serum cholesterol levels by a little), a large public health effect would result. In other words, although individuals with high risk factors may benefit from interventions specifically targeted at them, the effect on the overall incidence of the disease will be limited in the absence of a population-oriented intervention.

The examples presented in this section illustrate certain strengths and weakness of high risk strategies (less overall potential) and population strategies (small benefits to individuals) in particular scenarios. In the next section, we take a more general look at the characterise of the two types of prevention strategy, and how such characteristics may come into play during policy-making.

3.3. High risk strategies vs. population strategies: what to consider in policy making

Who should be targeted for prevention measures? Should they be directed at those individuals at high risk who, we have seen, comprise only a small proportion of all cases of the diseases we have considered above, or should they be directed at the whole population? It is of great value to have some idea of what the likely effect of an intervention would be (see Section 2 above). There are also policy questions of how and when such interventions should be instituted, and whether these should take place at a community level, or within hospitals and/or GP surgeries. Since approaches maybe more or less efficient in particular circumstances, there cannot be a simple, universal answer to this question. What will be considered below are the merits of these different approaches.

3.3.1 The high risk approach

Interventions, especially those promoted by governments and clinical medical practitioners, tend to focus on the high-risk approach, targeting efforts at those who are known to be at increased risk of a particular disease. There are a number of reasons for this:

- it appears to be more cost-effective, in that interventions are directed at those whose health will improve most dramatically if the intervention achieves its objective;
- people who recognise that they are at increased risk may be more motivated to act to change their behaviour;
- the focus is on individuals rather than society - this message fits in with the ethos and organisation of medical care and is easier for doctors and health care providers to promote;
- the risk associated with the adverse exposure (e.g. alcohol and traffic 'accidents') is generally apparent;
- society (especially the politicians and media) tends to sanction concern which focuses on needy individuals rather than on changing societal structures to achieve health gains.

3.3.2 Limitations to the high risk approach

The high risk approach is limited by what it fails to address: public health problems arising from small, but widespread risks (see Table 5). In particular, the high risk approach fails to recognise that a large number of people, exposed to a small risk, may generate more cases than a small number of people exposed to a large risk.

The high-risk approach has other weaknesses (Table 5):

- *prevention tends to become medicalised* : people are labelled as well ('not at high risk') or unwell ('in a high risk group'). If they are 'well', they are not encouraged to change their behaviour or the environment; if `unwell' they are. There may also be adverse effects, such as stigmatisation, arising from such labelling and even those labelled well might be encouraged to adopt less risky behaviours.

- *It directs efforts towards approaches which are likely to be palliative and which bring only temporary success* : intervention is directed towards protection against the effects of the exposure (e.g. by using lipid-lowering drugs in the case of high cholesterol) or by reducing the levels of a few individual's exposure (e.g. encouraging those with high cholesterol to avoid red meat). No effort goes into changing the underlying circumstances which encourage people to engage in damaging behaviours.

- *the strategy takes inadequate account of the influences on behaviour*: most behaviours are shaped by societal norms, and peer pressure, which are in turn influenced by such factors as economic circumstances as well as marketing, advertising, and pricing. The high-risk approach encourages individuals to change their behaviour, but fails to recognise the determinants of such behaviour and how much of they are socially mediated.

- *high risk approaches have a poor record in predicting individual change in risk*: although those in `high risk' groups may be at increased risk collectively, it is still difficult to predict a particular individual's risk and to determine whether they are likely to be affected by the condition under discussion.

- *feasibility and costs*: it may be expensive to screen large populations to detect those at increased risk. For example, it would be very expensive, complex and perhaps undesirable to attempt to ensure that all elderly people are screened by tonometry to detect increased intra-ocular pressure, associated with an increased risk of glaucoma. Such a screening programme would require expensive equipment and staff, and the test itself is uncomfortable and has poor predictive power.

- *the contribution to overall control of disease may be disappointingly small*: it is likely that a high-risk strategy will be effective only if all, or a high proportion of the total population's risk is confined to a readily identified segment of the population and if available interventions are effective, affordable and acceptable. In general, the extent of the problem that can be addressed depends on the ways in which risk and exposure are distributed within the population.

Table 5. Advantages and disadvantages of primary (and in part secondary) prevention strategies

Advantages and disadvantages of different strategies for primary prevention	
High-risk individual strategy	**Population strategy**
Advantages:	
• Appropriate to individuals • Individual highly motivated • Relatively easy to motivate health worker/physicians • Relatively high potential benefit to targeted individuals	• Potential benefit for whole population • Recognises societal influences on individual behaviour • More fundamental, radical
Disadvantages:	
• Limited public health impact • Difficult to identify individuals at especially high risk • Effect is temporary and has to be repeated as new people enter the high risk category • Poor recognition of influences on behaviour	• Small benefit to individual ("prevention paradox") • Equal focus on those individuals at especially increased risk and those at relatively lower risk • Difficult to motivate both patients and physicians or other health workers
Adapted from Rose 1992	

3.3.3 The population approach

The benefits of the population approach are apparent from Table 5. The most important feature is that the population approach recognises that individual behaviour is greatly influenced by societal factors, and that considerable reductions in risk of disease may be achieved by influencing change at population rather than individual level. This approach is inherently more radical, in that it identifies societal change and healthy public policy as desirable; it implicitly suggests that focusing on individual behaviour, outside of a social context, is inherently limited.

The population approach also has weaknesses which needs to be recognised. It is best suited to situations where there is a clear increase in risk given an increase in 'dose' of exposure to the relevant risk factor. In such situations, it is useful to attempt to shift the entire population distribution of exposure towards lower levels. In relation to a problem like obesity however, one needs to be cautious about those people who are already very thin, further reducing their calorie and fat intakes. So, some knowledge of the dose-response curve is necessary before embarking on population-wide interventions and even if the basic intervention is directed at whole populations, it may be necessary to help certain groups of people appreciate that they may need to respond in a slightly different way.

4. Screening

Screening involves identifying unrecognised diseases or risk factors for disease by applying tests on a large scale to a population that does not have clinical symptoms. Screening tests usually seek to identify small groups at high risk of the condition. Further tests are needed to confirm the diagnosis.

Screening is most often used to select those people who are at higher risk of developing a disease and to offer them a health intervention aimed at prevention by one of two means:

1) Prevention of serious outcomes of existing disease (**secondary prevention**). *Example*: screening for breast cancer followed by confirmation of diagnosis and early surgical treatment of those with mammograms suggestive of breast cancer.

2) Prevention of the development of a disease (**primary prevention**). *Example*: screening for high blood cholesterol levels to select people at higher risk of coronary heart disease for health promotion or cholesterol-lowering drug treatment.

Screening is also used for other purposes:

3) Selection of people fit enough for a job. *Example*: routine health checks for army recruits.

4) Containment of infection. *Example*: screening new nurses or teachers for tuberculosis or food handlers for salmonella.

Screening can either involve the whole population (**mass screening**) or selected groups who are anticipated to have an increased prevalence of the condition for which screening has been instituted (**targeted screening**). An example of **mass screening** would be to measure the blood pressure of all adults in a population. Measuring blood cholesterol in relatives of people with familial hyperlipidaemia is an example of **targeted screening**.

Screening can be organised in a *systematic* way (for example: a list is kept of all the females in a population and each woman is invited routinely for a first mammography test the week after her fiftieth birthday) or it can be organised *opportunistically* (for example: a general practitioner takes and records the blood pressure of every patient, regardless of the reason for the consultation).

4.1 Ethics of Screening

A screening test is a medical intervention that is carried out when a person is not ill and usually carried out on someone who has not initiated the request for the test. For these and other reasons the ethics of carrying out screening are complicated and need careful thought.

Screening programmes can range between the policy extremes of strong incentives to participate and strong disincentives to participate, though most screening programmes are far from either extreme. There is increasing interest in considering the rights of individuals in relation to the issue of informed decision making, including both:

- Individuals not getting screened when policy makers feel the screening programme is beneficial to the public health, and

- Individuals getting screened when policy makers feel the screening programme is not beneficial to the public health.

In the context of informed decision making it is important to remember that screening programmes can do harm as well as give benefit.

4.2 Risks and benefits of screening

4.2.1 Risks to the individual of screening

- There may be health risk attached to the screening tests (for example: exposure to x-rays or the risk of miscarriage after amniocentesis).
- There may be health risk attached to further confirmatory tests for those with a positive screening result.
- A false positive test may cause unnecessary anxiety.
- There may be other unwanted and unplanned effects of positive test (for example, life insurance premiums may be increased).
- A person with a false negative test may be reassured inappropriately and fail to recognise subsequent warning signs of the disease.
- A true positive test may increase anxiety and pose a risk to mental health or quality of life.

4.2.2 Risk for the individual: Over-diagnosis

Some screen-detected cancers might never have progressed to become symptomatic in the absence of screening, and some people who have cancer detected by screening would die from another cause before the cancer became evident. These cancers cannot be distinguished from other cancers and will thus be treated. This adverse consequence (harm) of screening is called over-diagnosis or over-detection, and is defined as the detection of cancers that would never have been found were it not for the screening test.

4.2.3 Benefits for the individual

- Early detection.
- Reduced morbidity.
- Reduced chance of dying for diseased individuals.
- Reassurance for those with normal results.

4.2.4 Risks for society

- The opportunity cost of the resources put into screening. For example, health care providers can be allocated to screen cigarette smokers for lung cancer. The time and resources devoted to lung cancer screening could have been allocated to other health-promotion activities, some of which could have a greater impact on population health and possibly at lower cost.

- The costs of confirmatory tests and of treatment. For example, if all adults are screened for raised serum cholesterol levels (a known risk factor for heart disease), then there will be a large increase in the number of people prescribed lipid lowering drugs. Yet, only a minority of these adults taking lipid lowering medication would otherwise have developed heart disease.

4.2.5 Benefits for society

- Fewer premature years of life lost.
- Economic benefit from these years.
- Reduced cost of treating advanced disease.
- Less transmission (i.e. of infections).

4.3 Guidelines for Screening

There are WHO guidelines for deciding when screening is appropriate, drawn up by Wilson and Jungner in 1968:

1. The condition being screened for should be an important health problem.
2. The natural history of the condition should be well understood.
3. There should be a detectable early stage.
4. Treatment at an early stage should be of more benefit than at a later stage.
5. There should be a suitable test for the early stage.
6. The test should be acceptable.
7. Intervals for repeating the test should be determined.
8. There should be adequate health service provision for the extra clinical workload resulting from the screen.
9. The risks, both physical and psychological, should be less than the benefit.
10. The costs should be balanced against benefit.

Beaglehole *et al* provide a simpler list of criteria, which cover the same points:

Disease	-Serious -High prevalence at preclinical stage -Natural history understood -Long period between first signs and overt disease
Diagnostic Test	-Sensitive and Specific -Simple and cheap -Safe and acceptable -Reliable
Diagnosis and Treatment	-Facilities are adequate -Effective, acceptable and safe treatment available

Some of these criteria are essential. For example, there is no value in screening unless effective treatment is available. Screening for lung cancer has been debated for some time because, although it is a serious life-threatening disease, there are few effective treatments. Other criteria are more relative. For example, if a disease is very rare but very serious and easily preventable, it may still be worth screening for it. Phenylketonuria (PKU), a congenitally acquired inability to metabolize the amino-acid phenylalanine, is an example of a rare disease. If undetected, it leads to serious intellectual disabilities. A highly sensitive and specific screening test, performed on a blood sample taken from a prick in the heel of a newborn, can identify babies who have this condition. With treatment and a diet low in phenylalanine, children with PKU can expect to live a normal life.

4.4 Evaluating screening programmes

Determining whether a particular screening programme is of value in a particular community will depend on four main issues:

4.4.1 Relative burden of disease

Prevalence, incidence or mortality rates for the disease can usually be determined, but the final judgement on whether this disease is sufficiently important to institute a screening programme inevitably depends on its relative value in comparison with other health problems.

4.4.2 Feasibility

Feasibility will depend on how easy it is to organise the population to attend for screening, whether the screening test will be acceptable (having one's mouth checked for dental caries is acceptable to most people, but having a sigmoidoscopy to detect colon cancer is much less acceptable), whether facilities exist to carry out more extensive diagnostic tests on those who are found positive at screening (for example, can the hospitals cope with the expected increase in demand for breast biopsies following a mammography programme), and whether there are sufficient resources available to treat everyone confirmed as positive (for example, can the health services afford to provide cholesterol-lowering drugs to everyone found to have an elevated serum cholesterol level).

4.4.3 Effectiveness

Effectiveness is evaluated by the extent to which instituting a screening programme effects health outcomes (i.e. mortality or morbidity). Effectiveness is challenging to measure - particularly when using an observational study design - because of a number of potential biases:

Selection bias. People who participate in screening programmes often differ from those who do not. Selection bias can work both ways; people who are at high risk may be more likely to come forward (for example, women whose mothers had breast cancer are at higher risk of breast cancer, and are more likely to request screening) or people at lower risk might be more likely to participate (for example, many women at low risk of cervical cancer have a high risk perception, and are more likely request screening).

Lead-time bias. Because screening identifies disease that would otherwise be identified at a later stage, the longer ‘survival’ time observed for those who were screened may be due to an earlier date of diagnosis, and not because of improved prognosis from earlier treatment.

Length-time bias. Some conditions may be slower in developing to a stage where they threaten health than others. For example, some breast tumours are slower growing, so they spend longer in a preclinical stage. This means slow-growing tumours are more likely to be detected at screening but people with these tumours also have a more favourable prognosis. A screening trial that has detected many slow-growing tumours can incorrectly conclude that screening was beneficial in lengthening the lives of those who were found positive.

4.4.4 Cost

The cost of programmes is important. However wealthy a society is there is a finite amount of resources available for health care, so the relative cost-effectiveness of a screening programme compared with other forms of health care should always be considered. Costs to be considered will be those relating to the conduct of the screening programme, those relating to the further diagnostic tests required for those labelled positive by the screening programme, and the cost of treating those in whom the disease is confirmed. On the other hand, in the absence of screening and prevention or early treatment of disease, costs will be incurred by the treatment of patients in more advanced stages of disease.

4.5 Study designs for evaluating screening effectiveness

Because there are such powerful potential biases, it is difficult to evaluate screening. Some of the study designs which can be used are:

1. **Cohort studies** comparing length of survival in screen detected and non-screen detected cases.

   Liable to: **lead-time bias**
   length-time bias

2. **Case-control studies** comparing screening history of cases (usually cancer deaths) with age matched **controls** from the same population who have not died or developed the disease to estimate odds ratio for the reduction in risk due to screening.

   Liable to: **selection bias**
   recall bias (cases or their surrogates may have different recall of screening episodes from controls)

3. **Non-randomised trials or cohort studies** comparing mortality using historical or neighbourhood **controls**.

   Liable to: **selection bias**

4. **Randomised controlled trials** are the **best** method where possible.

Selection bias is removed by random allocation to receive screening or not. Other biases (lead time, length time and diagnosis) are removed when comparing the overall mortality in the two randomised groups but adjustments for lead time and length time still need to be made when comparing the length of survival.

Problems with randomised trials of screening may be:

Uptake. Factors influencing uptake are complex, including psychological and social components as well as the acceptability of the test procedure itself. In the Swedish Two County Study (see section 9 below) acceptance rates of up to 90% were obtained. In contrast uptake, in the Edinburgh breast cancer screening trial was only 66%. When screening uptake is low, the potential effect size is diluted.

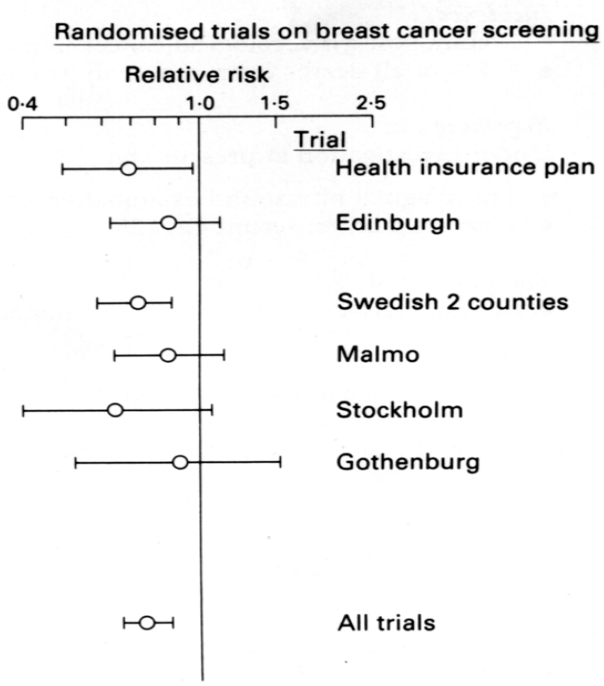
Contamination of the control group. Awareness of the screening programme may lead participants in the control non-screened group to seek out screening which leads to a further dilution of the expected benefits.

Sample size. Very large numbers of participants are often required in screening trials, with increasingly large numbers for diseases with a low incidence rate or where the trial is designed to show smaller benefits.

Equipoise. If a screening programme has already been introduced in a population despite the lack of experimental evidence, it may be considered unethical to subsequently conduct a randomised trial.

4.6 Examples of RCTs to evaluate screening effectiveness

The largest trial of breast cancer screening was the Swedish Two County Study. Over 120,000 women were randomized to either screening or control arms. The trial with the longest follow-up was the Health Insurance plan trial (over 18 years of follow up). The majority of trials randomised women in the age range 40 or 45 to 64 or above. Some trials used individual randomisation (such as Swedish Two County Study and the Edinburgh trial) while other trials randomised clusters (e.g. GP practices or municipalities). The **figure** below gives the results for the meta–analysis of the trials for women aged 50 to 74 years showing an overall 24% reduction (95% CI 13 to 33%) in breast cancer mortality in women first screened at age 50-74 years. The women in the intervention arms of these trials had an average 70% uptake of screening. The benefit for younger women was smaller (15% reduction in mortality, p=0.2). The Canadian National Breast Screening Study (not shown here) which was mainly in younger women (40 to 59 years) did not detect a reduction in mortality from the screening programme (RR= 1.08, 95% CI 0.84 to 1.27) in women aged 40 to 49 years or 50 to 59 years.



The relative risk of breast cancer mortality in women aged 50–74 years invited for mammographic screening compared with those not invited is shown for each trial together with the 95% confidence interval (CI). The combined estimate for all trials is also shown.

4.7 Validity of a Screening Test

The validity of a screening test is measured by the degree to which it accurately distinguishes between persons with the condition (that is, the preclinical disease or the risk factor for the disease) and those without. When a screening test incorrectly labels someone as having the condition when further investigation shows that they do not, it is called a '*false positive'*. When a screening test incorrectly labels someone as not having the condition when further investigation (or the subsequent development of the clinical disease) shows that the condition is present it is called a '*false negative*'.

Sensitivity of a test is defined as the proportion of people with the disease who are test positive. If the sensitivity is low, it suggests that a substantial proportion of cases will be missed. These are the 'false negatives'.

Specificity of a test is defined as the proportion of people without the disease who are found test negative. If the specificity is low it suggests that a substantial proportion of people without the disease will be labelled as having the disease. These are the 'false positives'.

Estimation of sensitivity and specificity for a given screening test will depend on the definition that is used for a true positive. In some situations this may be determined by carrying out a further 'gold standard' diagnostic test, and in others it may be determined by following up participants to see who develops clinical manifestations of disease.

Another important measure for a screening test, which is particularly important to the individual who has received the results of a screening test, is the 'predictive value'.

The **positive predictive value** is defined as the proportion of people found to be test positive who actually have the disease.

The **negative predictive value** is defined as the proportion of people found to be test negative who do not have the disease.

The predictive values depend on the prevalence of the disease as well as the sensitivity and specificity of the screening test. Sensitivity and specificity are only dependent on the characteristics of the test and the condition being tested (and are not dependent on the prevalence of disease).

Table 7 summarizes the relationship between the results of a screening test and the actual presence of disease as determined by the result of a subsequent confirmatory diagnostic test (the 'gold standard').

Table 7. Assessing screening test validity.

		True disease status	
		positive	**negative**
Result of screening test	**positive**	**a**	**b**
	negative	**c**	**d**

a = people who have the condition and are found positive by the test (true positive)
b = people who do not have the condition but are found positive by the test (false positive)
c = people who have the condition but are found negative by the test (false negative)
d = people who do not have the condition and are found negative by the test (true negative)

$$\text{Sensitivity} = \frac{a}{a+c} \qquad \text{Positive Predictive Value} = \frac{a}{a+b}$$

$$\text{Specificity} = \frac{d}{b+d} \qquad \text{Negative Predictive Value} = \frac{d}{c+d}$$

5. Conclusion

It is important to appreciate not only the relative risks associated with particular exposures, but also the absolute difference in risk conferred by exposure to a particular factor. Similarly, it is desirable to understand what proportion of cases of a disease can be attributed to a particular exposure; and to determine what proportion could potentially be prevented if the exposure could be removed or prevented. Concepts of attributable risk and attributable risk fraction assist in quantifying these features of exposures, and their public health implications.

An intervention policy designed to benefit the health of the public, must be justified by its effectiveness, its acceptability and its cost. The relationship these factors have with risk both in the short and long term are crucial determinants of appropriate public health policy. In some instances concentrating on high risk individuals will be better than a general untargeted approach. When using this approach, screening programmes may be implemented. This should be done using the criteria for suitability of screening. In terms of acceptability, however, it is often stigmatising, and therefore alienating, to have a policy directed only at particular individuals in a community characterized as being at ‘high risk’. The tradition in medicine is to identify sick individuals and treat them individually, without intervening in the community from which they come. In prevention it is quite possible that a generalised message is quite as potent and as valid, and possibly much more effective, than one concentrating on particular individuals. In practice, a combination of targeted and population-wide initiatives is often the most powerful.

References

Webb P and Bain C. *Essential Epidemiology: An introduction for Students and Health Professionals*. Chapter 15. Second Edition. Cambridge University Press. 2011.

Bailey L, Vardulaki K, Langham J and Chandramohan D, *Introduction to Epidemiology*. Chapter 12. Open University Press, 2005 (Understanding Public Health, Series editors: Nick Black and Rosalind Raine)

Rose G, The strategy of preventive medicine, Oxford University Press, 1992.

Rothman K.J, Greenland S., & Lash TL. *Modern epidemiology*, 3rd ed. Philadelphia: Lippencott, Williams, and Wilkins (2008).

Additional reading

Criteria to assess screening
https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme

Newborn screening
Therrell BL, Padilla CD, Loeber JG, Kneisser I, Saadallah A, Borrajo GJ, Adams J. Current status of newborn screening worldwide: 2015. *Semin Perinatol*. 2015 Apr;39(3):171-87.
http://www.sciencedirect.com/science/article/pii/S0146000515000191

Breast cancer
This paper includes a good account of over-diagnosis.
Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380:1778-86.

Cervical cancer
Sankaranarayanan et al HPV screening for cervical cancer in rural India. *N Engl J Med*. 2009 Apr 2;360(14):1385-94.

Catarino R, Petignat P, Dongui G, Vassilakos P. Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. *World Journal of Clinical Oncology*. 2015;6(6):281-290. doi:10.5306/wjco.v6.i6.281.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4675913/

Prostate cancer
The USA preventive services taskforce are currently updating their guidance on prostate cancer screening: their website provides a good account of their thinking and the key questions that need to be answered:
http://www.uspreventiveservicestaskforce.org/Page/Document/final-research-plan/prostate-cancer-screening1

A recent reanalysis if the two large RCTs to date and a useful commentary are available:

Tsodikov A et al. Reconciling the effects of screening on prostate cancer mortality in the ERSPC and PLCO trials. Ann Intern Med 2017 Sep 5; [e-pub]. (http://dx.doi.org/10.7326/M16-2586)
Vickers AJ. Prostate Cancer screening: Time to question how to optimize the ratio of benefits and harms. Ann Intern Med 2017 Sep 5; [e-pub]. (http://dx.doi.org/10.7326/M17-2012)