Education and debate

When can a risk factor be used as a worthwhile screening test?

N J Wald, A K Hackshaw, C D Frost

Department of Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, St Bartholomew's Hospital, London EC1M 6BQ
N J Wald
professor
A K Hackshaw
lecturer

Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT
C D Frost
senior lecturer

Correspondence to:
N J Wald
n.j.wald@mds.qmw.ac.uk

One of the most important areas of medical inquiry is the identification of risk factors for specific disorders. Such research is usually aimed at discovering new causes of a disease, but risk factors can also be used as screening tests. The fact that a risk factor must be very strongly associated with a disorder if it is to be a worthwhile screening test is not widely recognised. If this were better understood, fewer risk factors would be proposed unnecessarily as screening tests. Serum cholesterol measurement, for example, would probably never have been considered seriously as a screening test for ischaemic heart disease. Although a high cholesterol concentration is a strong risk factor for ischaemic heart disease in aetiological terms, the association is not sufficiently strong for it to be used as a screening test—in practice, its screening performance is poor.[1]

In this article we specify the quantitative relation between risk factors and screening tests and show how strongly a risk factor needs to be associated with a disease before it is likely to be a useful screening test. For simplicity, we consider only risk factors with a Gaussian distribution, though the general principles we present can be applied to all frequency distributions.
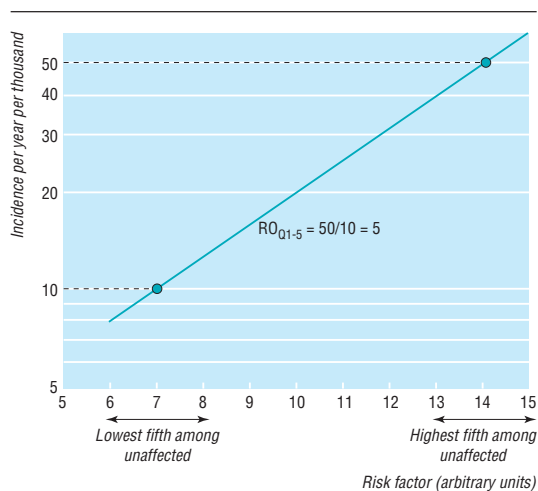

Fig 1 Dose-response relation between incidence of a disorder and levels of a risk factor where the relative odds for people in the highest fifth of the risk factor distribution compared with people in the lowest fifth (RO$_{Q1-5}$)=5 ((50:950)÷(10:990))

Summary points

To be a worthwhile screening test, a risk factor must be strongly associated with a disorder

The strength of association between a risk factor and a disorder can be quantified by the relative risk or relative odds (odds ratio)

A risk factor can also be considered as a screening test, and its association with the disorder can be quantified as the detection rate for a specified false positive rate

There is a direct numerical equivalence between the relative odds and the detection rate for specified false positive rate that does not depend on the incidence or prevalence of the disorder

A relative odds of 5 between the highest and lowest fifths of the distribution of a risk factor is equivalent to only a 14% detection rate for a 5% false positive rate if the SDs of the risk factor in people with and without the disorder are the same

Expressing risk as a relative odds usually compares risk in the tails of the distribution of the risk factor; this can give an overoptimistic impression of the value of the risk factor used as a screening test

Methods

Relative risk and relative odds

The strength of an association between a risk factor and a disorder can be quantified by a dose-response relation between the incidence of the disorder and increasing values of the risk factor. When risk (or incidence) is expressed on a log scale, a straight line relation with increasing levels of the risk factor is commonly seen. This simply means that there is a proportional, rather than absolute, change in the risk of the disease for a given change in the level of the risk factor. This is illustrated in figure 1. The slope of the line summarises the dose-response relation. This relation can be expressed quantitatively as the risk of people in the highest fifth of the distribution (for unaffected

individuals) of the risk factor developing the disorder compared with the risk of people in the lowest fifth of the distribution; $50/1000 \div 10/1000$ in figure 1. It can also be expressed as the relative odds (or odds ratio) for people in the highest and lowest fifth of the distribution, so in figure 1 it would be $(50:950) \div (10:990)$. Here this relation is described as the RO_{Q1-5}. If the disorder is rare, the relative risk and relative odds are almost the same.

Detection rate and false positive rate

The relation between a risk factor and a disorder can also be described as two overlapping relative frequency distributions of the risk factor in people with and without the disorder. This approach, in which the risk factor is considered as a screening test, is illustrated in figure 2. The same approach applies to individuals who are at risk of developing a disorder in the future—for example, people with high blood cholesterol who are at risk of developing ischaemic heart disease. In this case affected individuals are defined as those who develop the disorder over a given period of time, and unaffected individuals as those who remain free of the disorder over the same period. People with positive results have screening test values that are above a specified cut off level, and the approach yields the detection rate for a specified false positive rate. The detection rate (sensitivity) is defined as the proportion of affected individuals (or those who become affected during a given period of time) with positive results and the false positive rate (1 minus specificity) is defined similarly as the proportion of unaffected individuals with positive results. The performance of a screening test is often specified by the detection rate for a 5% false positive rate, which can be abbreviated as the DR_5. In the example shown in figure 2, the DR_5 is 25%. A similar notation can be used for detection rates equivalent to any specified false positive rate—for example, DR_1 and DR_{10} for false positive rates of 1% and 10% respectively.

DR₅ = 40%

Affected individuals

Unaffected individuals

Detection rate = 40%

False positive rate = 5%

Screening test (arbitrary units)

Equivalent $RO_{Q1-5} = 71\%/1.25\% = 57$

20%

20%

71%

1.25%

Lowest fifth

Highest fifth

Risk factor (arbitrary units)

Fig 3 Relation between detection rate for 5% false positive rate (DR_5) (top) and relative odds between lowest and highest fifth for unaffected individuals (RO_{Q1-5}) (bottom). Both distributions are identical, with equal SDs, but the means are 1.4 SDs apart

Methodological relation

The two methods of quantifying the relation between a variable and a disorder (illustrated by figs 1 and 2) are directly related if the frequency distribution of the variable is Gaussian in both affected and unaffected individuals. Thus, for any value of the RO_{Q1-5} there will be an equivalent DR_5 and vice versa. The relation depends on the means and standard deviations (SDs) of the variable in affected and unaffected individuals, but it does not depend on other factors such as the incidence or prevalence of the disorder.

Figure 3 shows how the RO_{Q1-5} and the DR_5 are linked quantitatively. In the example shown in the figure, the distributions of a variable in affected and unaffected individuals are known (each distribution having the same SD) and the distributions are separated by 1.4 SD units. The top part of the figure shows that at a cut off level chosen to give a 5% false positive rate the detection rate is 40%. The bottom part shows the same distribution yielding an RO_{Q1-5} of 57. It is derived as the ratio of the shaded areas on the right (defined by the upper fifth) divided by the ratio of the shaded areas on the left (defined by the lower fifth), which is $71\%/20\% \div 1.25\%/20\%$. The figure therefore illustrates how an RO_{Q1-5} of 57 is equivalent to a detection rate of 40% for a 5% false positive rate (DR_5). The formal mathematical relation between the two is given in the statistical appendix (see website).

In this paper we have assumed that screening variables are distributed in a Gaussian manner. Often an

Affected individuals

Unaffected individuals

Detection rate = 25%

False positive rate = 5%

5   6   7   8   9   10   11   12   13   14   15

Negative

Positive

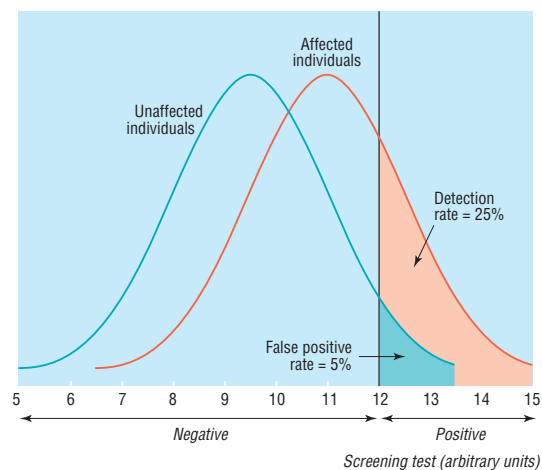*Screening test (arbitrary units)*

Fig 2 Overlapping relative frequency distributions in affected and unaffected individuals where the detection rate for a false positive rate of 5% is 25% (DR_5=25%). The implicit vertical axis describes the percentages of affected and unaffected individuals according to the screening variable
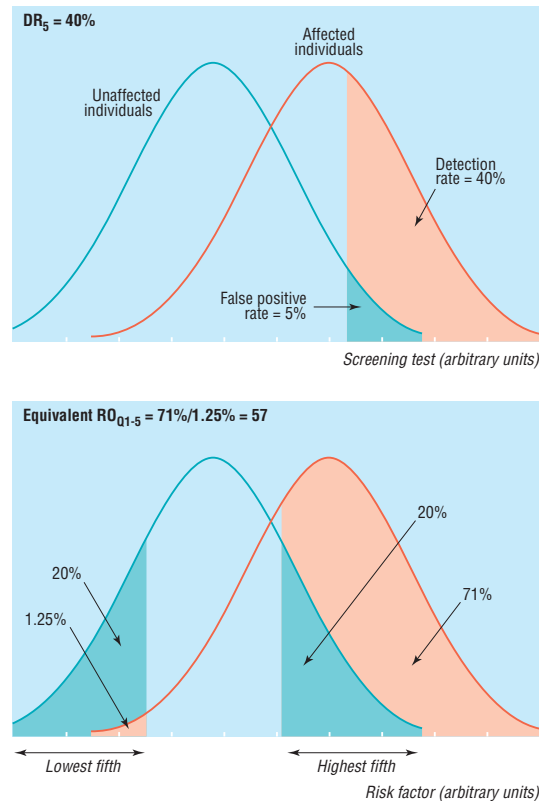
Maternal serum α fetoprotein and open spina bifida at 16-18 weeks' gestation

The measurement of α fetoprotein in pregnancy (at about 16-22 weeks of gestation) is a proved screening test for open spina bifida. The upper diagram in figure 4 shows the distribution of α fetoprotein in spina bifida and unaffected pregnancies at 17 weeks' gestation.[2 3] The RO_{Q1-5} is 246, indicating a strong association between α fetoprotein and open spina bifida. The extent of the association is perhaps not fully recognised by many people. The figure shows that the corresponding screening performance is high—91% of affected pregnancies can be detected in this way for a false positive rate of 5%. The detection rate, given the RO_{Q1-5} value, is somewhat higher than is suggested by the table because the SD in open spina bifida pregnancies is greater than that in unaffected pregnancies. This is a highly effective screening test.

adequate transformation (such as log transformation) can be found to allow data to fit a Gaussian distribution. Even if this is not the case, the approach illustrated in figure 3 whereby the distributions of a screening variable in affected and unaffected individuals are used to calculate both the detection rates and the relative odds is generally useful in linking quantitatively the screening performance to the strength of the risk factor.

Results

The table shows the relation between the RO_{Q1-5} and the DR_5 when the SD of the screening variable is the same in affected and unaffected individuals. An RO_{Q1-5} of 10, for example, which would be regarded as a high odds ratio in aetiological terms, corresponds to a DR_5 of only 20%; even an RO_{Q1-5} as high as 100 corresponds to a DR_5 of only 48%. To achieve a DR_5 of over 70% the RO_{Q1-5} would have to be about 800 or more.

If the ratio of the SDs in affected and unaffected individuals is less than 1, the detection rate is less than

Fig 4 Distribution of maternal serum α fetoprotein in pregnancies affected and unaffected by open spina bifida (derived from Wald et al[2]) and distribution of serum cholesterol in men who did and did not die of ischaemic heart disease (derived from Wald et al[1])

those shown in the table for the same RO_{Q1-5}. For example, if the ratio is 0.8, the DR_5 values are 2%, 6%, 9%, 20%, and 27% for RO_{Q1-5} values of 1, 5, 10, 50, and 100 respectively. If the ratio is greater than 1, the detection rates would be greater than those shown in the table. For example, if the ratio is 1.2, the DR_5 values are 9%, 23%, 31%, 53%, and 63% for RO_{Q1-5} values of 1, 5, 10, 50, and 100 respectively.

The boxes present two practical examples illustrating the above concepts. In each example, the risk factor is known to be associated with the particular disorder, but only in the first example is the screening performance acceptable.

Discussion

A risk factor has to be extremely strongly associated with a disease within a population before it can be considered to be a potentially useful screening test. Even a relative odds of 200 between the highest and lowest fifths will yield a detection rate of no more than about 56% for a 5% false positive rate, provided, as is commonly the case, that the distribution of the screening variable is approximately Gaussian (or log Gaussian) and shows a similar SD in affected and unaffected people.

It is not unusual for a strong risk factor of aetiological importance to be proposed as a screening test for the disorder, perhaps in the belief that strong risk factors make good screening tests. We show that

DR$_5$ percentages for specified RO_{Q1-5} values (SD in affected and unaffected individuals the same)

RO_{Q1-5}	DR_5 (%)
1	5
2	8
3	11
4	13
5	14
10	20
20	28
40	36
60	41
80	45
100	48
200	56
400	64
600	69
800	71
1 000	74
2 000	79
4 000	82
6 000	87
8 000	88
10 000	89

DR_5 is the detection rate for a 5% false positive rate; RO_{Q1-5} is the relative odds of having or developing a disease in the top fifth of the distribution of unaffected individuals compared with those in the bottom fifth.

Serum cholesterol and ischaemic heart disease

Because serum cholesterol is an established risk factor for ischaemic heart disease, it was believed that it would be a useful screening test. According to this view, individuals with high cholesterol concentrations—that is, those regarded as being "screen positive"—would be offered cholesterol lowering drugs to reduce the risk of a myocardial infarction. This belief was unfounded, as illustrated by figure 4. The lower diagram in figure 4 shows the distribution of serum cholesterol concentrations in men aged 35-65 in the United Kingdom who did or did not subsequently die from ischaemic heart disease over a period of about 10 years. The RO_{Q1-5} value is 2.7, indicating that people with a high serum cholesterol concentration (in the highest fifth) are nearly three times more likely to die from ischaemic heart disease than those with a low serum cholesterol concentration (in the lowest fifth). This is a moderately strong association, but when it is assessed as a screening test the performance is poor. For a false positive rate of 5%, only 15% of those who would later die of ischaemic heart disease would be identified.[1] Again, this screening performance is slightly better than predicted by the table because of the somewhat higher SD of serum cholesterol concentrations in affected individuals.
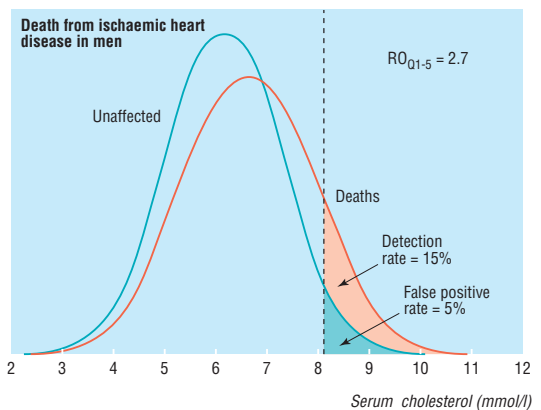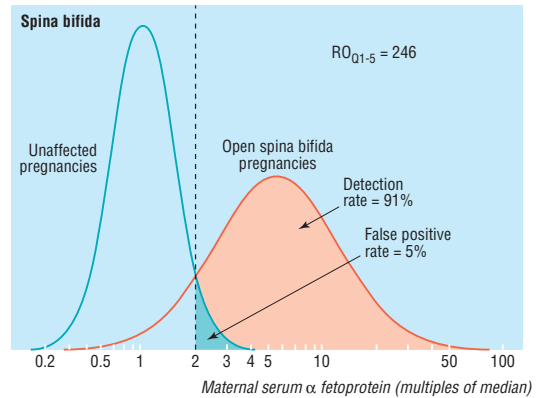
this is not necessarily the case, giving as an example serum cholesterol concentration in relation to ischaemic heart disease. A risk factor with an RO_{Q1-5} of 10 or greater is unusual. Few risk factors in epidemiology are as strongly associated with a disease as this. A risk factor with an RO_{Q1-5} would, however, perform poorly as a screening test. It would have a detection rate of about 20% for a false positive rate of 5% if the SDs were similar for affected and unaffected individuals. Even if the SD in affected individuals was 20% greater, the detection rate would only be 30% for a 5% false positive rate.

The fact that a strong risk factor can be a poor screening test may seem counterintuitive. The paradox is explained when it is recognised that the relative odds (or relative risk), usually used to evaluate risk factors as possible causes of a disease, is usually assessed by comparing the risk of disease at each end of the distribution of the risk factor. In this way the effect of being highly "exposed" to the factor is compared with being slightly "exposed." The groups being compared are mutually exclusive and most people in the middle of the distribution are ignored. When the risk factor is examined as a screening test, the likelihood of having (or developing) a disease given a positive result (say, ≥95th centile) is estimated relative to the average risk in the entire population, which not only includes all those below the cut off but also those above it. The aim in screening is to identify a group with a high risk relative to everyone.

Another reason why strong risk factors may make poor screening tests is that there may be little variation in exposure within populations. For example, we know that smoking cigarettes is a risk factor for lung cancer. However, if everyone in a certain population smoked 20 cigarettes a day, asking about cigarette consumption would not distinguish those who are more likely to develop lung cancer from those who are not.

Failure to recognise the above considerations may explain why serum cholesterol determination was proposed as a screening test for ischaemic heart disease even though it performed poorly as a screening test when cut off levels corresponding to the 95th centile were used.[1] Before a risk factor is considered as a screening test it would be worth determining the RO_{Q1-5} and then examining the table. This should help to determine which tests are potentially useful in medical screening.

1 Wald NJ, Law M, Watt H, Wu T, Bailey A, Johnson M, et al. Apolipoproteins and ischaemic heart disease: Implications for screening. *Lancet* 1994;343:75-9.
2 Wald NJ, Cuckle HS, Densem JW, Kennard A, Smith D. Maternal serum screening for Down's syndrome: the effect of routine ultrasound scan determination of gestational age and adjustment for maternal weight. *Br J Obstet Gynaecol* 1992;99:144-9.
3 Wald NJ, Kennard A. Prenatal screening for neural tube defects and Down's syndrome. In: Rimoin DL, Connor JM, Pyeritz RE eds. *Emery and Rimoin's principles and practice of medical genetics.* 3rd ed. Edinburgh: Churchill Livingstone, 1996:545-62.

(Accepted 19 July 1999)

Registering clinical trials

Alison Tonks

Randomised trials of medical interventions are the foundation of evidence based health care, but most are effectively conducted in secret. Few people—usually an elite network of investigators, funding agencies, and government regulatory bodies—know about a trial from its inception. Most trials become public knowledge only when the investigators publish their completed project in a journal—if they ever do.[1] In the meantime, others may be duplicating the effort or, worse, ignoring early warning signs that an intervention is dangerous. One trial of the class I antiarrhythmic drug lorcainide, for example, went unreported for over a decade even though the data suggested that the drug increased mortality in patients with myocardial infarction. During that time, use of such drugs continued and shortened the lives of up to 70 000 people each year in the United States alone.[2] A register of clinical trials is one way of opening up the process and avoiding these problems.

The idea was first mentioned 13 years ago[3] and has been refined since then by an international group of trialists, academics, and enthusiasts campaigning for a comprehensive, up to date, and searchable archive of ongoing and recently completed randomised trials, including trials done by the pharmaceutical industry. This article reports a recent conference on trial

Summary points

Clinical trials should be registered so that essential details are made public from a trial's inception, rather than from publication many years later

Openness about trials in progress reduces the impact of publication bias, prevents duplication of effort, promotes collaboration, and can save lives

Hundreds of trial registers already exist, but the information on them is not standardised and is incomplete; most contain only a subset of trials, often in high profile areas such as cancer or AIDS; and there are few incentives for researchers to register trials

Initiatives have begun to unify the existing web of registers, but they are at an early stage

Registration of clinical trials should be compulsory, either by law or by linking it to ethical approval of research

BMJ
Alison Tonks
papers editor
atonks@bmj.com