Editorial

The number needed to screen—an adaptation of the number needed to treat

The report of the Medical Research Council (MRC) trial of treatment of mild hypertension, published in 1985, ended with the conclusion that "the trial has shown that if 850 mildly hypertensive patients are given active antihypertensive drugs for one year about one stroke will be prevented—an important but an infrequent benefit".[1] This "number needed to treat" (NNT) to avoid one adverse event is simply the reciprocal of the absolute reduction in risk and was initially presented as a self evident way of summarising results clearly. Subsequently, however, it was given perhaps unjustified status,[2] became widely used, had treatises written on it,[3] and statistical techniques developed for its use.[4]

In 1998, Rembold proposed that the number needed to treat concept be adapted for screening as the "number needed to screen" (NNS).[5] This was defined as the number of people who need to be screened for a given duration to prevent one death or adverse event. In the above example, about 10 000 people in the specified age range (35–64) needed to be screened to generate the 850 mildly hypertensive patients who needed to be treated for one year to prevent one stroke.[1] Examples of some of Rembold's estimates are that the number of people who needed to be screened to prevent one death over five years was 1374 for haemoccult screening for colon cancer, 1251 for mammographic screening for breast cancer (age 60–69), 418 for "dyslipidaemia" screening and 274 for "hypertension" screening for cardiovascular disease, 354 identified as having coronary heart disease (a simple screening test) to be treated with aspirin, and so on.[5] In this issue of the journal, Ann Richardson presents a refinement of this number needed to be screened concept, which allows for the fact that not all people offered screening accept it, and for the selection effects whereby the risk in those people who accept screening may differ from the risk in those who decline screening (see page 125).[6]

The "number needed" concept appears simple and clear and is attractive in that respect. But it has limitations that can make it misleading.

The major limitation, which applies equally to the "number needed to treat" as to the "number needed to screen", is that they are expressions of absolute risk, not relative risk. Only proportional effects (relative risk or the proportional reduction in risk) can be generalised from the group of people in whom they were measured to the population at large. Absolute risk and related absolute measures are not generalisable. The 850 patients with mild hypertension (or the 10 000 screened to identify them) were comparatively young (their average age was 52);[1] less than half these numbers would be required to "prevent a stroke" if they were ten years older. The same comment applies to other estimates of the "number needed". If age is standardised, the "number needed to screen" still depends on the underlying incidence of the disease in a population and this may vary widely. Breast cancer and ischaemic heart disease are much less common in Japan than Britain for example, while stroke is more common. The underlying risk in people who have had a heart attack who could be identified and treated with aspirin would depend on the treatment that these people were already receiving.

Yet expressions like "1374 people need to be screened over five years" or "850 need to be treated for one year" to prevent one event imply generalisability. Rembold, in his paper advocating the "number needed to screen" concept, listed his estimates of 1374 people given haemoccult screening, 418 "dyslipidaemia" screening, 274 hypertension screening with other such estimates in a table that implied comparisons as to which was the most effective screening test.[5] But the ages and underlying disease incidence in the trial populations that yielded these estimates were not listed and even small differences would have affected the comparisons.

Expressions of proportional benefit usually avoid these problems. Aspirin reduces heart disease mortality by about a third, mammography (in women over 50) reduces breast cancer mortality by about a third, and a reduction in diastolic blood pressure 5 mmHg reduces the incidence of stroke by about a third, and these statements can be generalised. They apply equally to Britain and Japan, to men and women, and are similar at different ages.

A second problem connected with generalising estimates of the number needed to screen is that they can be unstable; small differences in factors that influence them can have large effects on the estimate. Rembold published three estimates for "hypertension" screening of 1961, 1307, and 274 according to the intensity of treatment—the size of the blood pressure reduction.[5] The estimate from the MRC trial cited above[1] is very different from all three of these—about 20 000. Other important factors in this respect include the definition of screen positives—in this example how high blood pressure must be for a person to be "hypertensive" and therefore eligible to receive treatment. A lower cut off, with more people receiving treatment but a lower average risk in those treated, would paradoxically increase the number needed to treat but decrease the number needed to screen. The exact definition of the outcome measure is also important; if transient ischaemic attacks are recorded in addition to completed strokes for example the proportional reduction in risk will be similar but the absolute reduction will be different—both the number needed to treat and the number needed to screen would be smaller.

"Number needed" estimates are prone to large random error. This is partly because they are inevitably derived from the results of single randomised trials rather than meta-analyses, since meta-analyses yield summary estimates of relative risk not absolute reduction in risk which is required. A further reason is that the estimate is commonly the "number needed" to prevent a death rather than an event. In deriving estimates of the proportional reduction in incidence from a randomised trial, it is customary to reduce random error by combining fatal and non-fatal

events—the proportional reduction will be similar for both. But expressions of absolute reduction in risk will be very different ("numbers needed" are greater for a death than an event). Deaths from cardiovascular disease are usually less common than non-fatal events in trials of healthy people, and if the estimate of the number needed to screen to prevent one death is taken from the results of a single randomised trial it may be subject to very large random error—it was this that made the above MRC estimate of 20 000 so large for example. The confidence intervals in the table in Richardson's paper in this issue (especially those for colorectal cancer screening) illustrate that the confidence interval may be wide.[6] But those who use the number needed to screen may ignore the wide confidence interval.

For all these reasons, estimates of the "number needed" may be misleading. But there may be a more fundamental problem. The concept of "preventing one event" may be wrong. Rather than the one patient out of 850 having a vascular event prevented by treatment, it is more likely that several benefit by having their event delayed by a few years. Our thinking on this issue has been inappropriately conditioned by experiences of the early diagnosis of specific cancers. Surgery cures some patients (a near normal life expectancy is restored) but makes little or no difference in the rest, and no intermediate position is conceivable. This is not necessarily the case with circulatory diseases. Most people with coronary artery or cerebrovascular disease eventually die of myocardial infarction or stroke, treatments tend to delay events rather than curing patients. A trial of a blood pressure lowering drug showing a 30% reduction in the incidence of stroke in treated patients might be interpreted as indicating that treatment prevented 30% of the strokes and had no effect on the other 70%, but the same result would have been recorded in the trial if all the strokes in the intervention group were delayed (by about three years on average). It is not possible to determine where the balance between the two interpretations lies. It makes relatively little difference which interpretation is correct when citing a proportional reduction in risk, only the absolute number of patients who benefit.

Estimates of "number needed" to treat or screen should carry a health warning. It may be that with this warning and the necessary caveats they would lose their simplicity and people would tend to revert to the more generalisable and reliable proportional expressions of benefit, and link these to a specified intervention and to incidence or prevalence data from a specified population to calculate an estimate of the absolute reduction in risk that is specific to that population and intervention.

M R LAW

Department of Environmental and Preventive Medicine,
Wolfson Institute of Preventive Medicine,
St Barts and The Royal London School of Medicine and Dentistry,
Charterhouse Square,
London EC1M 6BQ
m.r.law@mds.qmw.ac.uk

1 Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *BMJ* 1985;**291**:97–104.
2 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;**318**:1728–33.
3 Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ* 1996;**312**:426–9.
4 Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;**317**:1309–12.
5 Rembold CM. Number needed to screen: development of a statistic for disease screening. *BMJ* 1998;**317**:307–12.
6 Richardson A. Screening and the number needed to treat. *J Med Screen* 2001;**8**:125–7.