Useful addresses for patients, families, and health care professionals:

Parkinson's Disease Society, 22 Upper Woburn Place, London WC1H 0RA

International Tremor Foundation, c/o Mrs Karen Walsh, ALAC Centre, Harold Wood Hospital, Romford, Essex RM3 0BE

PSP (Europe) Association, c/o Mr Brian Fisher, 21 Church Street, Mears Ashby, Northampton.

1 Rajput AH, Rozdilsky B, Rajput A. Accuracy of clinical diagnosis in parkinsonism—a prospective study. *Can J Neurol Sci* 1991;18:275-8.
2 Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;55:181-4.
3 Findley LJ, Koller WC. Essential tremor: a review. *Neurology* 1987;37:1194-7.
4 Fahn S. Torsion dystonia: clinical spectrum and treatment. *Neurology* 1983;2:316-23.
5 Critchley M. Arteriosclerotic pseudoparkinsonism. In: Rose FC, Capildeo R, eds. *Research progress in Parkinson's disease.* London: Pitman, 1981:40-2.
6 Gibb WRG, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry* 1988;51: 745-52.
7 Bannister R, Oppenheimer DR. Degenerative diseases of the nervous system associated with autonomic failure. *Brain* 1972;95:457-74.
8 Dooneief G, Mirabello E, Bell K, Matder K, Stern Y, Mayeux R. An estimate of the incidence of depression in idiopathic Parkinson's disease. *Arch Neurol* 1992;49:305-7.
9 Biggins CA, Boyd JL, Harrop FM, Madeley P, Mindham RHS, Randall JI, *et al.* A controlled longitudinal study of dementia in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 1992;55:566-71.
10 Quinn N, Rossor MN, Marsden CD. Dementia and Parkinson's disease—pathological and neurochemical considerations. *Br Med Bull* 1986;42:86-90.
11 Byrne EJ, Lennox G, Lowe J, Godwin-Austen RB. Diffuse Lewy body disease: clinical features in 15 cases. *J Neurol Neurosurg Psychiatry* 1989;52: 709-17.
12 McKeith IG, Fairbairn AF, Bothwell RA, Moore PB, Ferrier IN, Thompson P, *et al.* An evaluation of the predictive validity and inter-rater reliability of clinical diagnostic criteria for senile dementia of Lewy body type. *Neurology* 1994;44:872-7.
13 Förstl H, Burns A, Luthert P, Cairns N, Levy R. The Lewy-body variant of Alzheimer's disease. Clinical and pathological findings. *Br J Psychiatry* 1993;162:385-92.
14 Stephen PJ, Williamson J. Drug-induced parkinsonism in the elderly. *Lancet* 1984;ii:1082-3.
15 Graham JG, Oppenheimer DR. Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy. *J Neurol Neurosurg Psychiatry* 1969;32:28-34.
16 van der Eecken H, Adams RD, van Bogaert L. Strioapallidal-nigral degeneration. An hitherto undescribed lesion in paralysis agitans. *J Neuropathol Exp Neurol* 1960;19:159-61.
17 Shy GM, Drager GA. A neurological syndrome associated with orthostatic hypotension. *Arch Neurol* 1960;2:511-27.
18 Dejerine J, Thomas AA. L'atrophie olivo-ponto-cérébelleuse. *Nouvelle Iconographie de la Salpêtrière* 1990;13:330-70.
19 Lantos PL, Papp MI. Cellular pathology of multiple system atrophy: a review. *J Neurol Neurosurg Psychiatry* 1994;57:129-33.
20 Quinn N. Multiple system atrophy—the nature of the beast. *J Neurol Neurosurg Psychiatry* 1989(suppl):78-89.
21 Wenning GK, Ben Shlomo Y, Magalhaes M, Daniel SE, Quinn NP. Clinical features and natural history of multiple system atrophy. An analysis of 100 cases. *Brain* 1994;117:835-45.
22 Quinn N. Multiple system atrophy. In: Marsden CD, Fahn S, eds. *Butterworth-Heinemann International Medical Reviews. Neurology 13, movement disorders 3.* London: Butterworth-Heinemann, 1994:262-81.
23 Beck RO, Betts CD, Fowler CJ. Genitourinary dysfunction in multiple system atrophy: clinical features and treatment in 62 cases. *J Urol* 1994;151: 1336-41.
24 Steele JC, Richardson JC, Olszewski J. Progressive supranuclear palsy. A heterogeneous degeneration involving the brainstem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. *Arch Neurol* 1964;10:333-59.
25 Lees AJ. The Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). In: Marsden CD, Fahn S, eds. *Butterworth International Medical Reviews. Neurology 7, movement disorders 2.* London: Butterworth, 1987:272-87.
26 Golbe LI, Davis PH, Schoenberg BS, Duvoisin RC. Prevalence and natural history of progressive supranuclear palsy. *Neurology* 1988;38:1031-4.
27 Ballard PA, Tetrud JW, Langston JW. Permanent human parkinsonism due to 1-methyl-4-phenyl-1-2,3,6-tetrahydropyridine (MPTP): seven cases. *Neurology* 1985;35:949-56.
28 Tyrrell PJ, Rossor MN. Extrapyramidal signs in dementia of Alzheimer type. *Lancet* 1989;ii:920.
29 Rinne JO, Lee MS, Thompson PD, Marsden CD. Corticobasal degeneration: a clinical study of 36 cases. *Brain* 1994;117:1183-96.
30 Geddes JF, Hughes AJ, Lees AJ, Daniel SE. Pathological overlap in cases of parkinsonism associated with neurofibrillary tangles. A study of recent cases of postencephalitic parkinsonism and comparison with progressive supranuclear palsy and Guamanian parkinsonism-dementia complex. *Brain* 1993;116:281-302.

The number needed to treat: a clinically useful measure of treatment effect

Richard J Cook, David L Sackett

The relative benefit of an active treatment over a control is usually expressed as the relative risk, the relative risk reduction, or the odds ratio. These measures are used extensively in both clinical and epidemiological investigations. For clinical decision making, however, it is more meaningful to use the measure "number needed to treat." This measure is calculated on the inverse of the absolute risk reduction. It has the advantage that it conveys both statistical and clinical significance to the doctor. Furthermore, it can be used to extrapolate published findings to a patient at an arbitrary specified baseline risk when the relative risk reduction associated with treatment is constant for all levels of risk.

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada N2L 3G1
Richard J Cook, *assistant professor*

Nuffield Department of Clinical Medicine (Level 5), John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU
David L Sackett, *professor of clinical epidemiology*

Correspondence to:
Professor Cook.

BMJ 1995;310:452-4

More emphasis is now being put on effective use of biomedical literature to guide clinical treatment. As a result accessing, critically appraising, and incorporating the results of clinical investigations into clinical practice are becoming higher priorities for doctors and medical students.[1]

A pivotal step in translating clinical research into practice is the summarisation of data from randomised trials in terms of measures of effect that can be readily appreciated by doctors and other carers. Various measures of the effect of treatment are used in analysing results. Each measure has its own interpretation and statistical properties that make it suitable for some applications but perhaps not for others. We describe here a new measure referred to as number needed to treat[2] and a simple method of adopting this approach to individual patients at different levels of risk.

Measures of treatment effect

Consider a parallel group study in which patients are randomised to either an active treatment or a placebo control arm, are followed for a fixed amount of time, and are observed to experience a binary response to treatment (event/no event). We assume here that the events are adverse, and the objective is therefore to prevent them.

The effect of treatment is usually measured by comparing the probabilities of events in the two groups of patients. Point estimates of these measures are obtained by substituting the observed rate of events for the probabilities. For example, the absolute risk reduction is the difference in the probabilities of an event in the control and treatment groups and is estimated as the corresponding difference in the event rates. If the event rate in the treatment group is less than that in the control group this suggests a potential benefit from the active treatment. Similarly, if the event rate is greater in the treatment group than the control group (negative absolute risk reduction) the active treatment may be harmful. Before recommendations can be made regarding the treatment more formal analyses of the treatment effect are needed to quantify the strength of evidence: this is done by tests of significance or confidence intervals.

Another approach to summarising effects of treatment is based on the relative risk. Relative risk is defined as the probability of an event in the active treatment group divided by the probability of an event in the control group. The relative risk can be conveniently estimated as the ratio of the corresponding

event rates, with beneficial treatments giving relative risks below one. A related measure, called the relative risk reduction, is derived simply by subtracting the relative risk from one. On this scale a relative risk reduction of zero indicates no benefit or harm associated with the active treatment, whereas a relative risk reduction of one could indicate a "cure." The relative risk reduction can also be expressed as the absolute risk reduction divided by the probability of an event in the control arm and hence can be thought of as a standardised measure of the absolute risk reduction.

Another measure often used to summarise effects of treatment is the odds ratio. This is defined as the odds of an event in the active treatment group divided by the odds of an event in the control group. Though this measure has several statistical advantages and is used extensively in epidemiology, we will not pursue it here as it is not helpful in clinical decision making.

For some treatments and conditions the benefit of a specific treatment, as measured by the relative risk or the relative risk reduction, remains roughly constant over patient populations at varying baseline risk. In these cases relative measures appear attractive since a single estimate of treatment effect can be provided for a broad class of patients. On the other hand, it is often clinically important to consider the baseline (control) risk of an event before recommending treatment since for a given relative risk reduction, the expected absolute benefit of treatment could vary considerably as the baseline risk changes. For example, an estimated relative risk reduction of 50% might be statistically significant and clinically important for patients at moderate to high risk of a particular adverse event. However, for patients with a low probability of an event the risk reduction might not be sufficient to warrant the toxicity and cost of active treatment. This is the main criticism of relative measures of treatment effect for the purposes of clinical decision making.

Number needed to treat

Laupacis *et al* introduced an alternative approach to summarising the effect of treatment in terms of the number of patients a clinician needs to treat with a particular therapy to expect to prevent one adverse event.[2] The "number needed to treat" can be expressed as the reciprocal of the absolute risk reduction. In addition, a 95% confidence interval for the number needed to treat can be constructed simply by inverting and exchanging the limits of a 95% confidence interval for the absolute risk reduction. Though mathematically related to risk differences, the number needed to treat formulation is becoming widely used as a tool for therapeutic decision making[3] and bedside teaching[4] as it facilitates interpretation in terms of patients treated rather than the arguably less intuitive probabilities.

We use data from a recently published overview on the benefit of antihypertensive therapy for mildly and moderately hypertensive patients[5] to show the advantages (table). We divided studies in the overview into two groups: those in which all patients had a diastolic blood pressure of less than 110 mm Hg at entry and those in which all patients had a diastolic blood pressure of less than 115 mm Hg at entry. The two groups of studies were mutually exclusive, although the second group of studies includes patients with diastolic blood pressure of less than 110 mm Hg. The table shows that for patients with moderate hypertension receiving placebo treatments about 20% would be expected to have a stroke over the next five years; this risk is reduced to 12% with antihypertensive drugs, generating an estimate of the absolute risk reduction of 0·20−0·12=0·08. The reciprocal of this number is about 13, implying that a doctor would need to treat about 13 moderately hypertensive patients for five years before he or she could expect to prevent one stroke.

The attractive feature of the number needed to treat analysis over methods based on measures of relative efficacy is seen if we compare moderately and mildly hypertensive patients. For both risk groups the relative risk reduction is 40%, suggesting that both groups should be treated with equal vigour. However, the estimate of the number needed to treat to prevent one stroke is 13 for moderately hypertensive patients and 167 for mildly hypertensive patients. The clinical recommendation is therefore likely to be different for these two groups.

Extrapolating to patients at different baseline risks

The numbers needed to treat method still presents a problem when applying the results of a published randomised trial in patients at one baseline risk to a particular patient at a different risk. For example, in the hypertension example suppose a particular patient had only half the baseline risk of stroke of the moderately hypertensive patients in the overview. Such a judgment is typically made by comparing the patient's clinical history with the characteristics of the study patients, as indicated by baseline variables and inclusion or exclusion criteria.

Until now, the published relative risk reduction has been applied to the individual patient's baseline risk. This assumes a constant relative risk reduction for varying baseline risks, as is the case in our example. The estimated relative risk reduction of 40% from the trial would be applied to the patient's hypothesised baseline risk of 0·10 (0·2×0·5), generating an estimated absolute risk reduction of 0·04. This number would then be inverted, resulting in a number needed to treat of 25. The process becomes even more laborious when it is necessary to calculate the 95% confidence interval around the relative risk reduction. This requires two additional calculations based on the confidence limits for the relative risk reduction.

When we used the number needed to treat method in decision making during ward rounds we found translating the results of published trials to individual patients at potentially different baseline risks was time consuming and that the results were sometimes incorrect. We therefore looked for a simpler method.

The process can be greatly simplified by comparing the baseline risk of an individual patient with that of the typical patient in the published trial. If the baseline risk of the individual patient is a factor f times the baseline risk of a typical study patient and the relative risk stays constant, the absolute risk reduction for the

Calculation of risk reduction and numbers needed to be treated for patients with hypertension (based on results of Collins et al[5])

Hypertension	Stroke in 5 years: Control group	Stroke in 5 years: Active treatment group	Relative risk reduction $(P_c-P_A)/P_c$	Absolute risk reduction P_c-P_A	Number needed to treat $1/(P_c-P_A)$
Moderate (diastolic ≤115 mm Hg)					
Event rate (P)	0·20	0·12	0·40	0·08	13
Total No of patients	16 778	16 898			
Mild (diastolic ≤110 mm Hg):					
Event rate (P)	0·015	0·009	0·40	0·006	167
Total No of patients	15 165	15 238			

patient is scaled according to the same factor f. The estimated number needed to treat corresponding to patients at the revised baseline risk is therefore simply the study number needed to treat divided by f. Thus, in our example if a patient was judged to be at only half the baseline risk of the moderately hypertensive patients in the published trial $f=0{\cdot}5$ and the corresponding number needed to treat is 12·5/0·5 or 25. Confidence intervals can be easily obtained by dividing the limits of the corresponding interval from the original study by the factor f. In the trial the 95% confidence interval for the absolute risk reduction in moderately hypertensive patients was (11·4 to 13·9). The corresponding interval for a patient at half the baseline risk is therefore (11·4/0·5 to 13·9/0·5)=(22·8 to 27·7).

This simplification of translating the results of published trials to individual patients allows easy and rapid consideration of questions such as "what if the patient's risk was a third or a quarter that of patients in the published trial?" The ability to perform these sensitivity analyses is important since the baseline risk is partly based on subjective clinical judgment.

In our example the assumption of a constant risk reduction is satisfied exactly. If we consider the baseline risk of mildly hypertensive patients as 0·015/0·200=0·075 times that of the moderately hypertensive patients, we obtain a number needed to treat of $1/(0{\cdot}08\times0{\cdot}075)=167$, the same value derived from the raw data. Though this is an extreme example the general approach has proved useful in a wide variety of clinical scenarios when a quick "adjusted number needed to treat" is required and departures from the assumption of constant relative risk reductions are expected to be minimal.

1 Evidence-based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practise of medicine. *JAMA* 1992;**208**:2420-5.
2 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Eng J Med* 1988;**318**:1728-33.
3 Smith GD, Egger M. Who benefits from medical interventions? *BMJ* 1992;**308**:72-4.
4 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: *a basic science for clinical medicine*. 2nd ed. Boston: Little Brown, 1991.
5 Collins R, Peto R, MacMahon S, Herbert P, Fiebach NH, Eberlein KA, *et al.* Blood pressure, stroke, and coronary heart disease. II. Short-term reductions in blood pressure: overview of randomized drug trials in their epidemiologic context. *Lancet* 1990;**335**:827-38.

Events per person year—a dubious concept

Jürgen Windeler, Stefan Lange

In 1982 a new measure was introduced in research into osteoporosis and is now used everywhere in the literature. The so called "fracture rate" relates the number of fractures (single in some patients, multiple in others) to the cumulative time of observation of all patients. This concept, however, has no sound basis. Counting events instead of patients usually violates basic statistical assumptions and invalidates the use of common statistical tests and estimators. Its clinical interpretation is rather dubious. The use of such a measure impedes the search for valid and clinically meaningful outcome criteria and should be abandoned.

The concepts of design and analysis of randomised clinical trials seem to be well known to most researchers in clinical medicine. Randomisation, double blinding, definition of a primary end point, and prior calculation of power and sample size are widely accepted criteria for the quality of a clinical trial. There are additional problems, however, one of them being the handling of drop outs and missing information about them. Despite the favoured concept of intention to treat[1,2] these problems have not gained adequate attention or—what is worse—have produced inadequate and invalid solutions.

We came to know one "solution" when we reviewed several trials concerning the treatment of osteoporosis,[3-10] but there are other topics of research in which a similar procedure can be observed.[11,12]

The problem

Suppose that a clinical trial is performed to compare a new drug versus placebo with some binary end point. This may be death, myocardial infarction, recurrence of cancer, or any other criterion of success or failure. In the case of osteoporosis this is the occurrence of new (vertebral) fractures. We will assume a three year treatment and observation period with the primary end point of the trial being the proportion of patients with new fractures after three years. We know from experience that a small or considerable number of patients will not complete the study. Reasons will not be discussed in this context. With those patients who reached the end point event before "dropping out" no problems arise. But how do we deal with patients who leave the study after one or two years without having reached an end point event?

The "solution"

The information about these patients that can be used is the actual time under observation and the occurrence of an event in this time period. The observation time of each particular patient (to the time of an event if an event occurred) is expressed in a suitable unit (days, months, years). The sum of these observation times forms the denominator of some kind of event rate. The number of patients with an event is the numerator of the event rate. This approach is known as the subject years or person years method.[13] It is widely used in epidemiology especially in the analysis of mortality or incidence of cancer. Note that such settings have in common that a certain event (death) occurs only once in each patient.

Therapeutic research in osteoporosis goes further than this. If a fracture, which is usually defined as a certain relative decrease in vertebral height identified by roentgenograms, occurs in more than one vertebra this will be counted as two or more fractures. And if in a patient a fracture is observed after the first year and an additional decrease in height of the same relative amount in the same vertebra is observed after the third year then this again is counted as two fractures. Hence, while the denominator of the rate remains the same the numerator actually does not express a number of patients but a number of events scattered in some way over the study patients. The resulting term is generally referred to as the "fracture rate."

The origin of this procedure is quite easy to discover. Several authors speak of it as "the method of Riggs" and refer to a publication of 1982.[14] In fact, it can be seen from the "statistical analysis" section of this paper that Riggs and colleagues just invented this calculation by stating that "we assumed that the numbers of

Institut für Medizinische Biometrie, Ruprecht-Karls-Universität Heidelberg, Germany
Jürgen Windeler, *lecturer*

Abteilung Medizinische Informatik und Biomathematik, Ruhr-Universität Bochum, Germany
Stefan Lange, *research assistant*

Correspondence to:
Dr Windeler, Institut für Medizinische Biometrie, Ruprecht-Karls-Universität, Im Neuenheimer Feld 305, 69120 Heidelberg, Germany.

BMJ 1995;310:454-6