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 $f=0.5$ 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/(0·08×0·075)=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 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.

fractures during the period of observation followed a Poisson distribution." Perhaps they may have seen a reason for doing so, but they do not present any argument to justify this strong assumption.

The "solution" seems to kill two birds with one stone. Firstly, it seems to solve the "drop out" problem said to be inherent in clinical trials because all patients can be included in this analysis. And, secondly, as the number of events (fractures) in a certain time period is considerably higher than the number of patients with (at least) one event, the counting of events increases the power of a study compared with the counting of patients and consequently reduces the number of patients and costs. This is why the above defined fracture rate gained much interest and is found everywhere in the osteoporosis literature.

The solution of the drop out problem by the person years approach, however, is based on critical assumptions and as the approach has no "direct interpretation on an individual level"[15] it is hardly useful for the interpretation of clinical trials and medical decision making. Besides, counting events instead of patients, although having been used in papers published in journals of high scientific reputation, is suitable merely for a chapter in *Methodological Errors in Medical Research*.[16]

Statistical remarks

Statistical methods that are based on common probability distributions (normal, t, χ^2, binomial, Poisson, etc) assume independent observations. Therefore, four possibilities can be considered.

Firstly, of each patient in a clinical trial one and only one piece of information is included in a test statistic or estimator; the patient is the sampling unit. As patients can be regarded as independent of one another no problems arise with statistical procedures.

Secondly, more than one piece of information concerns one patient (for example, repetitive events, multiple fractures). In this case the assumption is needed that all events—single in some patients, multiple in others—can be regarded as independent; a second event in a certain patient is as likely as the first event in this or another patient. Under this condition the occurrence of events could be described by a simple Poisson model with the underlying likelihood of event occurrence being the same for all patients. This is the assumption that underlies the so called fracture rate. It is a critical assumption, however, which is hardly ever justified in clinical medicine.

Thirdly, the likelihood of event occurrence is different among patients but can be assumed constant within one patient; a second event in a certain patient is as likely as the first event in the same patient. Each patient can be characterised, so to speak, by his or her own Poisson model. In this setting comparisons between groups may be performed. No simple estimator is available, however, to describe a group of patients. The assumption may be justified in certain settings (infectious diseases) but not with a chronic disease like osteoporosis. And even if it were true it would not be of much use when a description, say, about the fracture risk in a group of patients is wanted.

Finally, no assumptions about independence or homogeneity are justified. No statistical "standard" procedures are available in this case. The usual calculation of statistical tests (whether parametric or non-parametric) or estimators is invalid and the use of usual confidence limits without meaning. The calculation of sample size is impossible. It is interesting that the authors who claim to have calculated a certain sample size[5] do not tell us explicitly how. Finally, the calculation of relative risks, which assumes two

binomial distributions with independent observations, does not make sense.

There are two possibilities of saving the concept. One could admit that there is a problem but argue that this violation of statistical assumptions has no practical impact; no such arguments are provided, and the comparison of results shows that the contrary is true (see below). Alternatively, one could assume that the likelihood of events is the same for all patients. Perhaps this viewpoint can be accepted in special settings, but in general it must be rejected. Without further discussion, people are more similar to themselves than to others. It seems self evident that vertebrae of the same spine do not behave like independent individuals, particularly when a systemic disease such as osteoporosis is present.

A new defensive strategy—we have not read any defence of this strategy yet—may be the argument that only few patients have repetitive events, that nearly all events come from different patients. Of course, this does not solve the principal problem at all but perhaps it again diminishes its practical consequences. If this were so, however, the question should be allowed of why such an end point was chosen. One aim of this choice—to reduce sample size—cannot be achieved in this case because the number of patients and the number of events are nearly the same. Only if the number of events is relevantly higher than the number of patients having at least one event will a reduction in sample size be possible. This, however, inherently means the violation of basic statistical assumptions and thus makes calculation of sample size and statistics a farce.

Clinical remarks

Although statistical arguments clearly show that the fracture rate approach in terms of events per person year is invalid and although one has to state that in general the sampling unit of clinical experiments is neither leg nor tooth nor vertebra nor erythrocyte but the patient, one must ask whether the concept is appreciated because of a highly useful clinical interpretation. This would require every effort to develop a statistically sound model rather than discard the concept. Clinical trials in general aim at learning something that may improve the management of patients. They are to provide information for appropriate decision making in individual patients. If that is accepted what are the implications of the statement "the fracture rate is reduced by 50%"? Does this allow a statement such as "the probability that a patient X develops a fracture within the next y years is reduced by 50%"? Certainly not. Does it allow the prognosis that "the treatment is expected to reduce the number of fractures in a patient X by 50%"? Again no. Does it allow a valuable comparison between treatments? No. Fracture rates (here events per 100 person years) are the same (20/100 person years) whether 20 patients are observed for 10 years and each has two fractures or 1000 patients are observed for half a year and 100 (10%) of them have one fracture each.

Would a patient be advised in the same way if one knew to which group he or she belonged? Would the decisions be the same? Because it is not known from looking at fracture rates what the impact of the treatment for the individual patient is expected to be; the fracture rates do not provide the information needed to decide about the application of a certain treatment to the individual patient.

The major importance of the problems and short-comings described here becomes obvious from the fact that different approaches lead to different interpretations of clinical studies. This can be seen from the results of the study of transdermal oestrogens[17] with

TABLE I—*Results provided by Lufkin et al.*[17] *Relation between patients and fractures is unknown for placebo group*

Detail	No included	No who dropped out	No analysed	No of patients with at least one new fracture	No of new fractures
Oestrogen	36	2	34	7	8
Placebo	39	5	34	12	20
P value				0·28 (Fisher's exact test)*	0·04 (Person years approach)†

*Own calculation.   †P value given by Lufkin *et al*, exact method not clear.

TABLE II—*Range of possible results with fixed number of new fractures. Calculations based on results published by Lufkin et al*[17]

No of new fractures for oestrogen *v* placebo	No of patients with at least one new fracture for oestrogen (n=34) *v* placebo (n=34)	P value (Fisher's exact test)
8 *v* 20	1 *v* 20	P<0·01 in favour of oestrogen
8 *v* 20	7 *v* 12	P=0·28
8 *v* 20	8 *v* 1	P=0·03 in favour of placebo

the fracture rate (event per 100 person years) leading to a significant result whereas the adequate comparison of patients with new fractures clearly leads to a non-significant result (table I). The relation between patients and fractures is not given in the paper. Most studies, however, do not even provide the reader with the number of patients with fractures. One can easily see that a certain result concerning the number of fractures (or fracture rate) is compatible with a wide range of possible results concerning the number of patients with at least one fracture (table II). Until complete information and an adequate analysis are available caution seems to be justified against the premature interpretation of beneficial effects.

To solve the problem

Of course, this does not mean that randomised clinical trials in osteoporosis are impossible. There are at least three possibilities for the definition of a primary end point.

Firstly, the proportion of patients with at least one new fracture at the end of the study (=observation period) could be used. This is a binary outcome measure which from experience will require quite a large number of patients. Leaving costs aside this does not seem to be a problem as osteoporosis is regarded as a "growing epidemic."[18] Efforts to lower drop out rates and an intention to treat analysis are necessary.

Secondly, we could use the mean number of fractures per patient in the study period. This is a quantitive outcome measure, which in contrast with the first includes a rough impression of the severity of the disease. There are scores (such as the spine deformity index) available which may meet many demands[19]; unfortunately they have not been sufficiently evaluated and are not widely accepted. In this case lower drop out rates than usual and an intention to treat analysis are also necessary.

Thirdly, the time until event (first fracture, worst fracture, etc) would be considered. This is also a

quantitive variable, which in contrast with the so called fracture rate makes use of the actual observed time (until event) as the outcome. The effect of a treatment is regarded as potentially delaying the onset of an event. This time will not be known for all patients because some of them will not experience the event within the study period and maybe not at all. Therefore techniques of analysing censored data such as survival analysis have to be used. One problem here as with the first suggestion is that the incidence of new events may be too low to gain valid estimates. Inclusion of high risk patients and longer observation times will contribute to overcome these difficulties. A high proportion of patients with their follow up completed is needed for a valid interpretation.

The use of the fracture rate should be abandoned. It is of doubtful use and by its existence impedes the search for and development of new approaches. To perform reliable studies with sufficient numbers of patients and to develop the methods of fracture weighting and of decreasing drop out rates seems to be much more promising.

1 Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, *et al*. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *Br J Cancer* 1976;34:585-612.
2 Newell DJ. Intention to treat analysis—implications for quantitative and qualitative research. *Int J Epidemiol* 1992;21:837-41.
3 Tilyard MW, Spears GFS, Thomson J, Dovey S. Treatment of postmenopausal osteoporosis with calcitriol or calcium. *N Engl J Med* 1992;326:357-62.
4 Storm T, Thamsborg G, Steiniche T, Genant HK, Soerensen OH. Etidronate in postmenopausal osteoporosis (reply). *N Engl J Med* 1990;323:1210.
5 Kleerekoper M, Peterson EL, Nelson DA, Phillips E, Schork MA, Tilley BC, *et al*. A randomised trial of sodium fluoride as a treatment for postmenopausal osteoporosis. *Osteoporos Int* 1991;1:155-61.
6 Storm T, Thamsborg G, Steiniche T, Genant HK, Soerensen OH. Effect of intermittent cyclical etidronate therapy on bone mass and fracture rate in women with postmenopausal osteoporosis. *N Engl J Med* 1990;322: 1265-71.
7 Palmieri GMA, Pitcock JA, Brown J, Karas JG, Roen LJ. Effect of calcitonin and vitamin D in osteoporosis. *Calcif Tissue Int* 1989;45:137-41.
8 Pak CYC, Sakhaee K, Zerwekh JE. Effect of intermittent therapy with a slow release fluoride preparation. *J Bone Min Res* 1990;5(suppl 1):149-55S.
9 Watts NB, Harris ST, Genant HK, Wasnich RD, Miller PD, Jackson RD, *et al*. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *N Engl J Med* 1990;323:73-9.
10 Riggs BL, Hodgson SF, O'Fallon WM, Chao EYS, Wahner HW, Muhs JM, *et al*. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *N Engl J Med* 1990;322:802-9.
11 Sahni R, Maniet AR, Voci G, Banka VS. Prevention of restenosis by lovastatin after successful coronary angioplasty. *Am Heart J* 1991;121:1600-8.
12 Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31-54.
13 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell, 1987.
14 Riggs BL, Seeman E, Hodgson SF, Taves DR, O'Fallon WM. Effect of the fluoride/calcium regimen of vertebral fracture occurrence in postmenopausal osteoporosis. Comparison with conventional therapy. *N Engl J Med* 1982;306:446-50.
15 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiology research. Principles and quantitative methods*. New York: Van Nostrand Reinhold, 1982.
16 Andersen B. *Methodological errors in medical research*. Oxford: Blackwell, 1990:152.
17 Lufkin EG, Wahner HW, O'Fallon WM, Hodgson SF, Kotowicz MA, Lane AW, *et al*. Treatment of postmenopausal osteoporosis with transdermal estrogen. *Ann Intern Med* 1992;117:1-9.
18 Ryan PJ, Blake GM, Fogelman I. Fracture thresholds in osteoporosis: implications for hormone replacement treatment. *Ann Rheum Dis* 1992;51: 1063-5.
19 McCloskey EV, Spector TD, Eyres KS, Fern ED, O'Rourke N, Vasikaran S, *et al*. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* 1993;3:138-47.

(Accepted 30 September 1994)

ONE HUNDRED YEARS AGO

THE FINANCES OF THE LONDON HOSPITALS. I.
On Christmas Day the *Times* took occasion to publish a seasonable article on the financial condition of the London charities, calling attention to the fact that the hospitals are passing through a serious crisis the outcome of which is altogether uncertain. Starting with the striking instance afforded by Guy's Hospital, where, in the last fifteen years, the annual income has decreased by no less a sum than £20,000, the writer goes on to say that hospitals dependent on voluntary contributions are in even worse plight than the great institution on the south side of the river. For this financial depression many causes are alleged. There is the ubiquitous agricultural depression, which has hit Guy's so hard, and other charities scarcely less severely; there is the natural tendency in bad times for the well-to-do to cut off their subscriptions to charities as the first means of retrenchment; there is the much-abused Finance Act of the last session, which renders it incumbent on the owners of land and wealth to provide for the death dues; and there is finally the ever-increasing number of those who clamorously appeal to the public to support some small special charity or to their private friends to enable them to carry out some temporary impulse of benevolence at small personal expense. (*BMJ* 1895;i:31.)