

23. Cox EV. The clinical manifestations of vitamin B₁₂ deficiency in Addisonian pernicious anemia. In: Heinrich HC, ed. *Vitamin B₁₂ and intrinsic factor*, 2: Europlisches Symposium, Hamburg. Stuttgart: Enke, 1961:590-602.
24. Reynolds EH. The neurology of vitamin B₁₂ deficiency. In: Zagalak B, Friedrich W, eds. *Vitamin B₁₂: proceedings of the Third European Symposium on Vitamin B₁₂ and Intrinsic Factor*, University of Zurich, March 5-8, 1979. Berlin: Walter de Gruyter, 1979:1001-8.
25. Greenfield JG, O'Flynn E. Subacute combined degeneration and pernicious anemia. *Lancet* 1933; 2:62-3.
26. Agamanolis DP, Chester EM, Victor M, Kark JA, Hines JD, Harris JW. Neuropathology of experimental vitamin B₁₂ deficiency in monkeys. *Neurology* 1976; 26:905-14.
27. Scott JM, Dinn JJ, Wilson P, Weir DG. Pathogenesis of subacute combined degeneration: a result of methyl group deficiency. *Lancet* 1981; 2:334-7.
28. Metz J, van der Westhuyzen J. The fruit bat as an experimental model of the neuropathy of cobalamin deficiency. *Comp Biochem Physiol [A]* 1987; 88:171-7.
29. Frenkel EP. Abnormal fatty acid metabolism in peripheral nerves of patients with pernicious anemia. *J Clin Invest* 1973; 52:1237-45.

SPECIAL ARTICLE

AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT

ANDREAS LAUPACIS, M.D., M.Sc., F.R.C.P.(C.), DAVID L. SACKETT, M.D., M.Sc., F.R.C.P.(C.),
AND ROBIN S. ROBERTS, M.Sc.

WITH neither the time nor the resources available to prevent, detect, or treat every disorder in every patient, which preventive, diagnostic, or therapeutic interventions should take priority? When a physician has a few spare minutes to spend with a patient, should that time be devoted to a blood pressure check, a counseling session about dietary fat, an inquiry about possible symptoms of transient cerebral ischemia, or a demonstration of how to use nicotine chewing gum?

Plenty of experts are quick to tell us how we should spend this valuable time; there are now even task forces and professional review bodies who, using rigorous rules for the admissibility of evidence, can give us some valuable advice.^{1,2} Thoughtful economic analyses are sometimes provided as well.³ However, all this advice tends to be focused on individual disorders and to provide yes or no answers. The practicing clinician still needs some yardstick with which to measure and compare the benefits and risks of various preventive, diagnostic, therapeutic, or rehabilitative approaches. After the pros and cons have been discussed with the patient, the treatment with the biggest payoff can be given top priority.

Such a yardstick would ideally have four properties. First, it would compare the consequences of doing nothing (the patient's risk for an adverse event if no intervention is provided) with the potential benefits of doing something (the extent to which this risk would be reduced by the use of a specific clinical treatment). Second, it would summarize the harm that would ac-

company the treatment in the form of side effects and toxicity to the patient. Third, it would identify patients who are both at high risk for an event and responsive to therapy. Finally, the yardstick would incorporate a measure that would permit a comparison of the consequences of applying one approach to the prevention, diagnosis, and treatment of one condition with the consequences of applying other approaches to other conditions, making it possible for individual clinicians and their patients to decide where best to focus their efforts.

In this essay we will describe several of the available yardsticks, illustrate their advantages and deficiencies, and suggest that one of them — the reciprocal of the absolute risk reduction (i.e., the number of patients with a given disorder that a physician must treat in order to protect one of them from the disorder's potential consequences) — may be a highly useful measure for clinicians.

MEASURES OF CLINICAL BENEFIT

Because randomized, placebo-controlled trials provide the most valid data on which to base all measures of the benefits and risks of particular therapies, we have used data from such trials here. In order to provide a clear picture of the differences between the various measures of benefit, we will use the same example for all of them: the Veterans Administration cooperative study on hypertension.⁴ This three-year study compared the efficacy of antihypertensive therapy (a combination of hydrochlorothiazide, reserpine, and hydralazine) with that of placebo. A subset of the data was collected in 201 men who were under the age of 50 and whose diastolic blood pressures were 90 to 114 mm Hg at the time of entry into the trial. Among patients who had target-organ damage at the time of entry, the rate of adverse events (sudden death, stroke, myocardial infarction, congestive heart failure, accelerated hypertension, and dissecting aneurysm)

From the Departments of Medicine, and Epidemiology and Biostatistics, University of Western Ontario, London, Ont. (A.L.), and the Departments of Medicine (D.L.S.) and Clinical Epidemiology and Biostatistics (A.L., D.L.S., R.S.R.), McMaster University, Hamilton, Ont., Canada. Address reprint requests to Dr. Laupacis at University Hospital, Rm. 5 OP3, P.O. Box 5339, Terminal A, London, ON N6A 5A5, Canada.

Dr. Laupacis is a Career Scientist of the Ontario Ministry of Health. Part of this work was done while he was a Fellow of the Medical Research Council of Canada.

was 22.2 percent in the control group and 8.5 percent in the treatment group (Table 1). Among patients without target-organ damage, the rate of adverse events was 9.8 percent in the control group, as compared with 4.0 percent in the treatment group.

Relative Risk Reduction

Relative risk reduction is the reduction of adverse events achieved by a treatment, expressed as a proportion of the control rate. In other words, it is the difference in event rates between the control and treatment groups, divided by the event rate in the control group.

In the Veterans Administration study, the relative risk reduction in patients with target-organ damage at the time of entry into the trial was $(0.222 - 0.085) \div 0.222 = 0.617$, or 62 percent. In patients without target-organ damage, the relative risk reduction was an almost identical 0.592, or 59 percent $[(0.098 - 0.040) \div 0.098]$. Thus, although the risk of adverse events among both treated and untreated patients was more than twice as high if they had target-organ damage, the relative risk reductions obtained by treating the two groups were almost equivalent. This illustrates the main disadvantage of using the relative risk reduction in clinical decision making; it does not reflect the magnitude of the risk without therapy, and it will therefore overestimate or underestimate the absolute impact of therapy when adverse events in untreated patients are very rare or very common, respectively.

Odds Ratio

The odds ratio is the traditional epidemiologic expression of the relative likelihood of an outcome. A probability (P) of 0.25 (one chance in four) represents odds of 1:3, or $\frac{1}{4}$. The probability of an event's occurrence generally corresponds to odds of $P \div (1 - P)$. Whereas the relative risk is the ratio of the probabilities of adverse outcomes in two treatments being compared, the odds ratio is the ratio of the odds of these adverse outcomes. In the Veterans Administration study (Table 1), the odds ratio for the group with target-organ damage was $(0.085 \div [1 - 0.085]) \div (0.222 \div [1 - 0.222]) = 0.325$; for the group without target-organ damage, it was $(0.040 \div [1 - 0.040]) \div (0.098 \div [1 - 0.098]) = 0.384$. As with the relative risk reduction, even though the two groups of patients had markedly different rates of adverse outcomes, the treatment effect expressed by the odds ratio was similar.

The odds ratio is often used as an approximation of relative risk in case-control studies, but it is also a valid measure of treatment effect in randomized trials. The odds ratio has distinct statistical advantages over the relative risk in terms of its sampling distribution

Table 1. Measures of Efficacy in the Veterans Administration Trial.⁴

PATIENTS' CONDITION AT ENTRY	RATES OF ADVERSE EVENTS		RELATIVE RISK REDUCTION (%) [*]	ABSOLUTE RISK REDUCTION [†]	NO. NEEDED TO BE TREATED [‡]
	PLACEBO GROUP	TREATMENT GROUP			
Target-organ damage	0.222	0.085	62	0.137	7
No target-organ damage	0.098	0.040	59	0.058	17

^{*}Calculated by dividing the difference between the event rates in the placebo and treatment groups by the event rate in the placebo group.

[†]The difference between the event rates in the placebo and treatment groups.

[‡]The number of patients who had to be treated in order to prevent one event. Calculated as the reciprocal of the absolute risk reduction.

and suitability for modeling⁵ and has become the preferred statistic for pooling data across trials in the form of meta-analyses⁶ — partly because of the strength and simplicity of the Mantel-Haenszel technique,⁷ but also because the odds ratio may be inherently more stable than other measures of treatment effect when applied across studies with various incidences of adverse outcomes. Although the odds ratio has certain statistical advantages over the relative risk reduction, it shares the latter's insensitivity to differences in the magnitude of risk without therapy, and its clinical usefulness suffers accordingly.

Absolute Risk Reduction

The absolute risk reduction (sometimes called the attributable risk reduction) is the difference in event rates between the control and treatment groups. Among patients with target-organ damage in the Veterans Administration trial, the absolute risk reduction was $0.222 - 0.085 = 0.137$ (Table 1). In patients without target-organ damage, the absolute risk reduction was 0.058, less than half that obtained in patients with target-organ damage. The advantage of the absolute risk reduction over the relative risk reduction and the odds ratio is that it is an expression of the consequences of giving no treatment and therefore provides an additional measure of clinical effect. In the example given, it also gains face validity by reflecting the common practice of treating hypertension more aggressively when target-organ damage is present than when it is absent. However, because the expression of the absolute risk reduction as a decimal fraction may not seem sensible to practicing physicians, this measure may be difficult for them to remember and incorporate into clinical practice.

Number Needed to Be Treated

The "number needed to be treated" is the number of patients who must be treated in order to prevent one adverse event. For example, in the Veterans Administration trial, if 100 control patients without target-organ damage had been followed for three years (risk of adverse event, 0.098), 10 events would have been expected. If, however, 100 such patients had been treated with antihypertensive agents and followed for three years (risk of adverse event, 0.040),

only four events would have been expected. Thus, on average, treating 100 such patients for three years would have prevented six ($10 - 4$) adverse events, meaning that 17 patients ($100 \div 6$) would have had to be treated in order to prevent one event. However, similar calculations reveal that among patients with initial target-organ damage, only seven would have had to be treated for three years in order to prevent one event.

Mathematically, the number needed to be treated is equivalent to the reciprocal of the absolute risk reduction. The number needed to be treated has the same advantage over the relative risk reduction and odds ratio as the absolute risk reduction in that it expresses efficacy in a manner that incorporates both the base-line risk without therapy and the risk reduction with therapy. Moreover, the resulting number needed to be treated is more useful than the absolute risk reduction because it tells clinicians and patients in more concrete terms how much effort they must expend to prevent one event, thus allowing comparisons with the amounts of effort that must be expended to prevent the same or other events in patients with other disorders.

The effect of the base-line risk and risk reduction on the number needed to be treated is illustrated in Table 2. If the event rate in the control group is high, even a small relative risk reduction will produce a low number needed to be treated (e.g., in a disease with a base-line risk of 90 percent, a relative risk reduction of only 10 percent will yield a number needed to be treated of 11). Conversely, if the event rate in the control group is low, the risk reduction must be large in order to produce a low number needed to be treated (e.g., if the base-line risk is 30 percent, a relative risk reduction of 30 percent is needed to yield a number needed to be treated of 11).

The general principle of the number needed to be treated can be used to assess clinical approaches other than long-term drug therapy. One can calculate the number of patients needed to be operated on to prevent one adverse event by carrying out a similar analysis of randomized trials of surgical procedures. For

example, in the European Coronary Study Group trial, only six patients with stable angina and stenosis of the left main coronary artery had to undergo coronary-artery bypass surgery for one life to be saved after five years of follow-up.⁸ Similarly, one can calculate the number of persons needed to be immunized by combining the incidence of infection with the results from vaccine trials. Over 200 persons need to be immunized with hepatitis B vaccine in order to prevent one case of hepatitis in the low-risk general population of the United States, as compared with only 8 persons in a high-risk population of homosexual men.⁹⁻¹¹ The concept can also be adapted to the realm of diagnosis and screening by calculating the number of patients needed to be examined to prevent one death from breast cancer. A comparison of screened and unscreened women in the Swedish National Board of Health and Welfare study showed that approximately 1592 women between 50 and 74 years of age had to be screened with mammography in order to prevent one death from breast cancer seven years after the screening was instituted.¹² Finally, studies of risk factors can yield a number needed to be exposed. A large cohort study estimated that 1306 persons must be exposed to the passive smoke of 20 or more cigarettes per day for 14 years before one such person dies of lung cancer.¹³

Approximate and exact methods for calculating a 95 percent confidence interval (which provides an estimate of the range within which the true value will lie with 95 percent confidence) are available for the number needed to be treated as well as for all the other measures of benefit discussed here.¹⁴

*Shortcomings of the Foregoing Measures of Clinical Benefit

Although these measures of benefit (especially, perhaps, the number needed to be treated) have some clinically useful properties, their shortcomings must be acknowledged. These shortcomings result from the properties of the measures themselves as well as from the data used. Ideally, the data are obtained in randomized trials, but although such trials are superior to other sources of information on the benefits and risks of therapy (e.g., comparisons of treated patients with nonrandomized contemporaneous or historical controls, case series), they share other deficiencies with them.

First, there is a disadvantage to combining the base-line risk and risk reduction into a single number (as is done to determine the absolute risk reduction and number needed to be treated). The fact that a physician must treat 11 patients in order to prevent one adverse event tells us nothing about the fate of the other 10 patients. On average, when the base-line risk of an event is 0.9 and the relative risk reduction is 10 percent, 9 of the remaining 10 treated patients will have an event. However, when the base-line risk is 0.3 and the relative risk reduction is 30 percent, an average of only 2 of the remaining 10 treated patients will

Table 2. The Effect of the Base-Line Risk and Relative Risk Reduction on the Number Needed to Be Treated.

BASE-LINE RISK*	RELATIVE RISK REDUCTION BY A NEW THERAPY (%)						
	50	40	30	25	20	15	10
	<i>number needed to be treated</i>						
0.9	2	3	4	4	6	7	11†
0.6	3	4	6	7	8	11	17
0.3	7	8	11†	13	17	22	33
0.2	10	13	17	20	25	33	50
0.1	20	25	33	40	50	67	100
0.05	40	50	67	80	100	133	200
0.01	200	250	333	400	500	667	1,000
0.005	400	500	667	800	1000	1333	2,000
0.001	2000	2500	3333	4000	5000	6667	10,000

*Risk of an adverse event in control patients.

†Numbers used as examples in the text.

have an event. Yet in both situations, the number needed to be treated is 11. Thus, the number needed to be treated reflects the average number of patients that must be treated in order to prevent one event but does not indicate the fate of the other patients or the virulence of the disease process.

A number needed to be treated of 11 means that 10 of 11 patients either do not need therapy or will not respond to it. Unfortunately, at our present level of knowledge, it is impossible to predict which 1 of the 11 patients is both at risk for an event and responsive to therapy; in order to prevent one adverse event, we must treat all 11 patients. Similarly, we are unable to identify the patients in whom the therapy will have side effects. These drawbacks affect the clinical application of any group-based assessment of treatment; this is one reason for the current interest in "n of 1" trials, in which the principles of the randomized controlled trial are applied to the individual patient.¹⁵ This approach is limited, however, to short-term, reversible outcomes in stable disease processes.

Second, for reasons of precision and efficiency, the results of randomized clinical trials are often expressed in disease-specific terms. For example, death from coronary heart disease and nonfatal myocardial infarction were the primary outcomes in the Lipid Research Clinics Coronary Primary Prevention Trial of cholestyramine for hypercholesterolemia, and deaths that were unrelated to coronary heart disease were excluded from the efficacy analysis.¹⁶ Although this is an appropriate way to maximize the efficiency of a trial, all deaths are important from a patient's point of view, and the clinician who is considering applying a trial's results may want to include deaths from all causes in calculating a measure of the benefit of therapy.

Third, when estimating the likely benefit of a treatment to their own patients, clinicians must also consider the likelihood that their patients will comply with the intervention. For good reasons, trials may incorporate compliance-improving strategies not applied in routine clinical practice. For example, in a trial evaluating the efficacy of a low-protein diet in retarding the progression of chronic renal failure, all patients had a consultation with a dietitian every two months.¹⁷ The results of that study may have overestimated the benefit of a low-protein diet that is prescribed without such intensive dietary guidance and emotional support.

Fourth, any measure of the benefit of treatment may vary considerably in different trials of the same or similar therapy because of different patient populations, trial designs (e.g., whether the therapy is evaluated in a setting designed to maximize compliance or as part of routine patient care),¹⁸ or chance. For example, some studies have shown a benefit of beta-blockers in the prevention of myocardial infarction, whereas others have not.⁶ Although a meta-analysis can generate an overall estimate of therapeutic efficacy, it may obscure differences between trials that

are important aids to a clinician deciding on therapy for a patient, especially when that patient resembles those in a particular trial. On the other hand, although it is tempting to use the estimate of therapeutic benefit provided by a particular trial, it must be remembered that apparent differences in treatment effect between trials conducted in relatively small numbers of patients may be due to chance alone.

Fifth, since clinical trials (like subexperimental studies) are of finite duration, they can provide only hints about the consequences of continuing therapy beyond the period of the trial. Moreover, the benefits of treatment may change with time. For example, the Beta-Blocker Heart Attack Trial Research Group reported the efficacy of propranolol in preventing death one, two, and three years after myocardial infarction.¹⁹ The relative risk reduction was higher during the first year of treatment than during the second year. The relative risk reduction over the first two years of the trial was 22 percent, but a similar proportion of patients in both the treatment and control groups died between years 1 and 2 (0.0347 in the propranolol group and 0.0316 in the placebo group). Although this may suggest that propranolol is no more effective than placebo after one year of therapy and that it should be discontinued at that time, the fact that more patients in the placebo group died during the first year than in the treatment group makes it likely that the propranolol-treated patients who were alive at the end of the first year were at higher risk than the corresponding placebo-treated patients. Therefore, discontinuation of propranolol at the end of the first year might have resulted in the death of a number of high-risk patients whose lives had been saved by propranolol until that time. This illustrates a common problem in interpreting any information on long-term therapy. The concern that an excess number of patients will have adverse events shortly after the cessation of therapy may explain why many clinicians elect to continue treatment beyond the duration of efficacy demonstrated by a trial, as long as the treatment has no unacceptable side effects in the patient.

Finally, some treatments are not effective until long after they have been started. The Lipid Research Clinics Coronary Primary Prevention Trial of cholestyramine demonstrated a relative risk reduction of 14 percent after 7.4 years of follow-up.¹⁶ However, there was no difference in primary outcome events between the two groups for at least three years, indicating that patients must continue to take cholestyramine (with its attendant side effects) for three years before any benefit can be expected. There is a similar situation in the assessment of surgical therapy (e.g., coronary-artery bypass surgery). The short-term mortality associated with surgery may be greater than that associated with medical therapy, but longer follow-up reveals an overall benefit of surgery. These issues are inherent aspects of clinical decisions; their more formal analysis by methods such as dis-

counting²⁰ and quantification of "risk aversion"²¹ will not be discussed here.

MEASURES OF HARM

In addition to conferring benefits, most effective therapies have clinically important side effects, and in some instances it is not clear whether the benefits outweigh the harm. How should the risks of harm be expressed and compared with measures of benefit? Although some clinical trials include detailed quality-of-life assessments that reveal both the benefits and harms of treatment in all patients, such assessments can be difficult and expensive and are therefore often not performed.

Alternatively, the absolute harm of therapy can be quantitated in the same way as the absolute benefit — by subtracting the harm in the control group from that in the treatment group. For example, in the Medical Research Council hypertension trial (in which patients with diastolic blood pressures of 90 to 109 mm Hg were randomly assigned to receive placebo, bendroflumethiazide, or propranolol for five years), the incidence of gout sufficiently severe to cause withdrawal from the trial was 0.0128 in men taking bendroflumethiazide, as compared with 0.0009 in those taking placebo.²² The absolute harm from gout caused by treatment with bendroflumethiazide is thus $0.0128 - 0.0009$, or 0.0119, and the relative risk of harm is $(0.0128 - 0.0009) \div 0.0009$, or 13.2. The advantages and disadvantages of using the absolute and relative harms as measures of the side effects of therapy are similar to those mentioned in the discussion of therapeutic benefits.

The number needed to be treated can incorporate the harm as well as the benefit of therapy. For example, in the Medical Research Council hypertension trial, the combined incidence of stroke and coronary events was decreased by both bendroflumethiazide and propranolol, and the number needed to be treated to achieve these outcomes in men taking bendroflumethiazide was 107. This trial also reported the reasons for withdrawal from therapy, which were impotence (4 percent), gout (5 percent), impaired glucose

tolerance (3 percent), and lethargy (1 percent). The reported side effects in men taking placebo and bendroflumethiazide can be incorporated into the calculation of the number needed to be treated, showing that to prevent one adverse event (by treating 107 men with bendroflumethiazide for five years), 4 men were forced to withdraw because of impotence, 5 because of gout, 2 because of impaired glucose tolerance, and 1 because of lethargy. The number of patients who had these symptoms but were not forced to stop therapy was even greater. This strategy permits clinicians to contrast the positive and negative consequences of therapy in a manner that is useful for both the physician and patient. Other negative aspects of therapy, such as monetary cost and inconvenience, can be added to these considerations if pertinent data are available, although the means for doing so (and, at times, the interpretation of such data) are beyond the scope of this paper and the resources of most clinicians.

COMPARING THE RELATIVE BENEFIT AND HARM OF TREATING DIFFERENT CONDITIONS

Although ideally all patients would be treated with all the effective preventive and therapeutic interventions, this is sometimes unfeasible because of limitations on the time of patients and physicians, patients' preferences, drug interactions, and treatment costs. If a clinician uses the same measure of benefit and harm for different conditions, it should be possible to compare these factors and determine which efforts on the parts of the clinician and patient will achieve the greatest net benefit for the patient. In order to make such comparisons of numbers needed to be treated, however, an additional problem must be solved. In Table 3 we have summarized the benefits reported in positive randomized trials of stepped-care therapy for diastolic blood pressure of 115 to 129 mm Hg,²³ coronary-artery bypass surgery for stenosis of the left main coronary artery of more than 50 percent,⁸ aspirin therapy for transient ischemic attacks,²⁴ cholestyramine therapy for asymptomatic men with hypercholesterolemia,¹⁶ isoniazid prophylaxis in patients with inactive tuberculosis,²⁵ and stepped-care therapy for diastolic

Table 3. The Benefits of Treating Six Different Conditions.

THERAPY	EVENTS	FOLLOW-UP (Yr)	BASE-LINE RISK	RELATIVE RISK REDUCTION (%) ^a	NO. NEEDED TO BE TREATED ^a	NO. NEEDED TO BE TREATED ^b
Stepped care for diastolic blood pressure of 115-129 mm Hg ²³	Death, stroke, myocardial infarction	1.5	0.13	89	9	3
Left main coronary-artery bypass surgery ⁸	Death	5	0.32	56	6	6
Aspirin for transient ischemic attacks ²⁴	Death, stroke	2.2	0.23	31	14	6
Cholestyramine for hypercholesterolemia ¹⁶	Death, myocardial infarction	7.4	0.12	14	60	89
Isoniazid for inactive tuberculosis ²⁵	Active tuberculosis	5	0.01	75	96	96
Stepped care for diastolic blood pressure of 90-109 mm Hg ²²	Death, stroke, myocardial infarction	5.5	0.05	14	128	141

^aCalculated for the actual length of follow-up in each trial.

^bAdjusted to the number needed to be treated after five-year follow-up, assuming constant benefit over time (see text for calculation).

blood pressure of 90 to 109 mm Hg.²² The problem arises from the different durations of these randomized trials; the average duration of follow-up ranged from 1.5 to 7.4 years, resulting in "raw" numbers needed to be treated ranging from 6 to 128. However, the number needed to be treated is time-dependent, which may affect the relative rankings of the numbers needed to be treated in trials of different lengths. If a treatment produces a constant relative risk reduction over time, the number needed to be treated decreases as the length of follow-up increases.

One can convert the number needed to be treated in a trial with a follow-up of T years to the approximate equivalent of a standard duration of S years by using the following formula:

$$\text{NNT:T} \times T \div S = \text{NNT:S},$$

where NNT:T and NNT:S are the numbers needed to be treated for T and S years, respectively. For example, the NNT:2.2 in the Canadian Cooperative Study Group trial of aspirin therapy for transient ischemic attacks²⁴ was 14. The NNT:5 for that study can be estimated to be $14 \times 2.2 \div 5 = 6$. This formula yields approximate figures and is less accurate for higher event rates (more than 10 percent per year).

Any method of comparing the results of trials of different lengths (regardless of whether one uses the relative risk reduction, absolute risk reduction, or number needed to be treated) must take into account any changes in benefit and harm that occur over time. Although it is convenient to assume that these factors remain constant over time, any measure of efficacy will be either underestimated or overestimated when they do not. For example, the previously described trial of beta-blockers in heart attack reported the efficacy of propranolol in preventing death one, two, and three years after myocardial infarction.¹⁹ The relative risk reduction was higher during the first year of treatment than during the subsequent two years. Consequently, conversion of the number needed to be treated to a common NNT:3 yields markedly different results, depending on the length of follow-up. For example, the calculated NNT:3 after one year of follow-up is 14, as compared with the actual NNT:3 of 28, obtained with use of the study data after three years of follow-up. Thus, extrapolating data from one interval to another is always dangerous and can only be done with confidence when the benefit and harm are known to be constant.

CONCLUSIONS

Although randomized controlled trials provide the most valid estimates of the benefit and harm of a treatment, the application of these results to individual patients may be difficult. In reviewing the different ways that benefit and harm can be expressed, we conclude that the absolute risk reduction is superior to the relative risk reduction because it incorporates both the base-line risk and the magnitude of the risk reduction. Its reciprocal, the number needed to be treated, expresses the absolute risk reduction in a manner that is

easily understood by clinicians, and can be used to describe the harm as well as the benefit of therapy and other clinical maneuvers.

We are indebted to Janice Coffey for assistance in the preparation of the manuscript, and to numerous colleagues for their helpful comments.

REFERENCES

1. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979; 121:1193-254.
2. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986; 89:2S-3S.
3. Levine MN, Drummond MF, Labelle RJ. Cost-effectiveness in the diagnosis and treatment of carcinoma of unknown primary origin. *Can Med Assoc J* 1985; 133:977-87.
4. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension. III. Influence of age, diastolic pressure, and prior cardiovascular disease; further analysis of side effects. *Circulation* 1972; 45:991-1004.
5. Mosteller F. Association and estimation in contingency tables. *J Am Stat Assoc* 1968; 63:1-28.
6. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; 27:335-71.
7. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *JNCI* 1959; 22:719-48.
8. European Coronary Study Group. Long-term results of prospective randomised study of coronary artery bypass surgery in stable angina pectoris. *Lancet* 1982; 2:1173-80.
9. Szmuness W, Stevens CE, Harley EJ, et al. Hepatitis B vaccine: demonstration of efficacy in a controlled clinical trial in a high-risk population in the United States. *N Engl J Med* 1980; 303:833-41.
10. Health and Public Policy Committee, American College of Physicians. Hepatitis B vaccine. *Ann Intern Med* 1984; 100:149-50.
11. Mulley AG, Silverstein MD, Dienstag JL. Indications for use of hepatitis B vaccine, based on cost-effectiveness analysis. *N Engl J Med* 1982; 307:644-52.
12. Tabár L, Fagerberg CJG, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Workshop Group of the Swedish National Board of Health and Welfare. *Lancet* 1985; 1:829-32.
13. Hirayama T. Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan. *Br Med J* 1981; 282:183-5.
14. Thomas DG, Gart JJ. A table of exact confidence limits for differences and ratios of two proportions and their odds ratios. *J Am Stat Assoc* 1977; 72: 73-6.
15. Guyatt G, Sackett DL, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy — randomized trials in individual patients. *N Engl J Med* 1986; 314:889-92.
16. Lipid Research Clinics Program. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. *JAMA* 1984; 251:351-64.
17. Rosman JB, ter Wee PM, Meijer S, Piers-Becht TPM, Sluiter WJ, Donker AJM. Prospective randomised trial of early dietary protein restriction in chronic renal failure. *Lancet* 1984; 2:1291-6.
18. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301:1410-2.
19. β -Blocker Heart Attack Trial Research Group. A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. *JAMA* 1982; 247:1707-14.
20. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. VII. To understand an economic evaluation (part B). *Can Med Assoc J* 1984; 130:1542-9.
21. O'Connor AM, Boyd NF, Warde P, Stolbach L, Till JE. Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome description, elicitation technique and treatment experience on preferences. *J Chronic Dis* 1987; 40:811-8.
22. Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *Br Med J* 1985; 291:97-104.
23. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension: results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. *JAMA* 1967; 202:1028-34.
24. The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfapyrazone in threatened stroke. *N Engl J Med* 1978; 299:53-9.
25. International Union Against Tuberculosis Committee on Prophylaxis. Efficacy of various durations of isoniazid preventive therapy for tuberculosis: five years of follow-up in the IUAT trial. *Bull WHO* 1982; 60:555-64.