Lingua Medica

Talking about Treatment: The Language of Populations and the Language of Individuals

In recent years, clinical decisions about treatment have been increasingly guided by the findings of randomized clinical trials. These trials use the strongest type of study design for treatment assessment. They offer the best approach to reducing bias and balancing the distribution of known or unknown patient attributes that could threaten the internal validity of the study (1). The quantitative measures of risk reduction used to report randomized clinical trials describe the collective experience of the participants in the trial. To apply the results of such trials in clinical practice, physicians must translate the concepts and measures used to describe groups of patients into a language that can inform the decisions of an individual patient. A clinical vignette illustrates the tensions inherent in this process of translation.

You are caring for a 74-year-old man whose blood pressure averages approximately 175/85 mm Hg. He has type 2 diabetes mellitus that is controlled by oral medications. On the basis of the findings of a randomized clinical trial, the Systolic Hypertension in the Elderly Program (SHEP) (2), you believe that treatment with antihypertensive medication will reduce his risk for a cardiovascular event. The patient asks you, "If I take this medication, how will it help me?" (The **Table** provides data from the SHEP trial to inform your answer.)

Three problems impede an accurate response to your patient's seemingly simple question. First, reports of randomized clinical trials do not usually account for individual characteristics that modify the effect of treatment (3). Second, clinicians and patients often have difficulty interpreting quantitative data about risk and risk reduction (4, 5). Finally, the results of treatment studies can be expressed, or "framed," in different ways; this can influence how both physicians and patients perceive the efficacy of treatment (6).

You suspect that diabetes increases your patient's risk for cardiovascular disease. He may therefore receive more benefit from treatment of hypertension. In the SHEP trial, as in most randomized clinical trials, some participants had risk factors that increased their likelihood of a cardiovascular event; 13% of enrolled patients were current smokers, 10% had a personal history of diabetes, and 5% had previous myocardial infarction (2). However, this particular trial provides only the average estimated benefit for all participants rather than an estimate that takes your patient's diabetes into account.

Interpreting the Data

The measures of risk reduction in a randomized clinical trial are calculated from two proportions: the proportion of patients in the control group with an adverse outcome (P_C) and the proportion of patients in the treatment group with the same adverse outcome (P_T). These proportions can be used to calculate two measures of treatment effect: relative risk reduction and absolute risk reduction. Formulas for both calculations are provided in the **Table**.

The relative risk reduction is the traditional method of reporting the findings of a randomized clinical trial. To describe to your patient the reduction in all cardiovascular events observed in the SHEP trial by using the relative risk reduction, you could say, "Taking antihypertensive medications for 5 years will reduce your risk for stroke or cardiac event by 30%." Because the event rate appears in both the numerator and the denominator of the relative risk reduction, it cancels out of their ratio (3). As a result, the same relative risk reduction can be observed for adverse outcomes that occur at much different rates. For example, the relative risk reduction for all adverse cardiovascular events in the SHEP trial (30%) was approximately the same as the relative risk reduction for a fatal stroke (29%), even though 12.2% of treated patients experienced some type of cardiovascular event and only 0.4% had a fatal stroke (**Table**).

See editorial comment on pp 604-606.

This paper is also available at http://www.acponline.org.

Table. Measures of Treatment Efficacy from the Systolic Hypertension in the Elderly Program (SHEP) Trial*

Outcome	Event Rate in Controls (P_C)	Event Rate in Treated Patients (P_T)	Relative Risk†	Relative Risk Reduction‡	Absolute Risk Reduction§	Number Needed To Treat‖
	%			%		
Total deaths	10.2	9.0	0.88	12	1.2	83
Cardiovascular death	4.7	3.8	0.81	19	0.9	111
Death from stroke	0.59	0.42	0.71	29	0.17	588
Fatal or nonfatal stroke	6.9	4.5	0.65	35	2.4	42
All cardiovascular events	17.5	12.2	0.70	30	5.3	19

* Calculated with data from reference 2.
† P_T/P_C
‡ $(P_C - P_T)/P_C$.
§ $P_C - P_T$.
‖ 1/absolute risk reduction = $1/(P_C - P_T)$.

Laupacis and colleagues (7) proposed that clinical decisions about treatment should incorporate the reciprocal of the absolute risk reduction, calculated as $1/(P_C - P_T)$, which they termed the *number needed to treat*. The number needed to treat to prevent any cardiovascular event in the SHEP trial was 19, roughly 30-fold less than the number needed to treat to prevent a fatal stroke (**Table**). You could describe the number needed to treat to your patient by saying, "About 19 persons need to be treated for hypertension for 5 years to prevent one stroke or cardiac event." The number needed to treat has become the standard for presenting the results of randomized clinical trials in *ACP Journal Club* (8).

Framing the Effectiveness of Treatment

The relative risk reduction and number needed to treat frame therapeutic decisions differently. The number needed to treat unambiguously assumes the population perspective because it explicitly states the number of people who must be treated and for what length of time to protect one person. Although the relative risk reduction is also a population statistic, it seems to promise a benefit that is individual, large, and immediate. In studies of hypothetical treatment decisions, physicians (9–12) and patients (13, 14) viewed treatments more favorably when benefits of these treatments were framed in terms of relative risk reduction. In a questionnaire study of 235 physicians (9), 49% of respondents reported that they would be more likely to prescribe a lipid-lowering drug with a relative risk reduction of 24% than a drug with an absolute risk reduction of 0.4%, even though both statistics were calculated from the Lipid Research Clinics Coronary Primary Prevention Trial (15). In another questionnaire study that used data from the Helsinki Heart Study (16), 88% of patients indicated that they would be willing to take a medication that caused a 34% relative risk reduction in heart attacks. However, only 31% of patients said that they would take this medication when the same information was expressed as a number needed to treat (71 persons needed to be treated for 5 years to prevent one heart attack) (14). Thus, in hypothetical cases, physicians and patients seem more likely to promote or accept treatment when it is framed in individual terms rather than in population terms.

Little research has assessed the content of actual clinical conversations about risk. In one study, discussions about risk occurred in 26% of audiotaped primary care office visits (17). Only 3.4% of risk discussions were quantitative; relative risk reduction was used in those conversations. The physician usually told the patient that an adverse outcome was certain unless the patient adopted the recommended treatment; in other words, that $P_C = 1$ (17). These limited clinical data support studies of hypothetical treatment decisions (9–14) by suggesting that physicians may deliberately or unconsciously choose to discuss the benefits of a treatment by using a framing strategy that does not provide neutral information but is intended to guide the treatment choices of their patients.

The Language of Populations: From Groups to Subgroups

You could estimate the benefits of hypertension treatment for your patient more precisely if information were available about the subgroup of diabetic patients enrolled in the SHEP trial. A valid subgroup analysis of a randomized clinical trial identifies characteristics of prognostic importance before data analysis, then compares outcomes between treated patients and controls within the subgroup of patients who have these characteristics (18). Subgroup analyses can provide useful information to guide decision making when the subgroups are large, the subgroup analyses are specified in advance, and the statistical analysis is appropriately conservative (19, 20). Although the initial SHEP

study (2) did not evaluate the effect of treatment of hypertension in persons with diabetes, a subsequent SHEP report found that the cumulative incidence of adverse cardiovascular events at 5 years among the subgroup of diabetic persons who received antihypertensive drugs was 21.4% compared with 31.5% among diabetic persons receiving placebo (21). When the formulas in the **Table** are used, these findings represent a relative risk reduction of 32% and a number needed to treat of 10. Thus, as we would expect to find in clinical practice, diabetic patients in the SHEP trial were at greater risk for adverse events and received more benefit from treatment of hypertension. With this additional information, you could say to your patient, "On average, 19 patients need to be treated for hypertension for 5 years to prevent a stroke or cardiac event. For persons with diabetes, only 9 to 10 patients need to be treated for the same period to achieve the same benefit."

In a similar way, you could use other single risk factors—or your overall assessment of your patient's risk for a cardiovascular event if he does not receive antihypertensive medication—to modify the average estimate of treatment benefit from the SHEP trial. A full discussion of this process, which is called *Bayesian reasoning*, is beyond the scope of this paper, but the literature contains excellent introductions (22, 23). In a busy clinical practice, neither the findings of the SHEP trial itself nor the additional data necessary to further particularize the benefits of treatment are readily available. Computerized decision support systems to provide such information in real time are being developed (24).

Problems of statistical power limit the applicability of subgroup analyses in randomized clinical trials to individual decisions. The sample size for a randomized clinical trial is calculated to provide a precise estimate of the average treatment effect. Nevertheless, estimates of the relative risk reduction or the number needed to treat from any study retain statistical uncertainty, which can be made explicit by reporting CIs around the point estimate of the treatment effect (25). A subgroup analysis compounds that uncertainty because the boundaries of the CI widen as the number of persons in the subgroup decreases. For a small subgroup, the CI may be so wide that predictions of benefit become too imprecise for clinical use. For example, in addition to diabetes, the most relevant subgroup analysis for your patient would include such variables as sex, age, baseline systolic blood pressure, and presence of baseline abnormalities on electrocardiography. However, the subgroup of patients with all of these characteristics in the SHEP trial was probably too small to provide meaningful estimates of treatment benefit.

If the number needed to treat were 1, we could guarantee benefit to an individual patient and would not need to choose between measures of treatment effectiveness. However, the formula used to calculate the number needed to treat is $1/(P_C - P_T)$. In this formula, the number needed to treat would equal 1 only when all untreated persons developed the adverse outcome ($P_C = 1$) and all treated persons avoided it ($P_T = 0$). In other words, the treatment would need to be invariably effective for a condition with uniformly bad outcomes. Few diseases are that bad, and no treatments are that good. In a recent compilation, the smallest number needed to treat was 1.1 (26). Thus, the number needed to treat never provides assurance of individual benefit.

In a few cases, the effect of treatment can be fully individualized by an *n*-of-1 randomized clinical trial (27–29). In this design, patients are assigned to an active treatment or placebo and then are crossed over at random during a series of treatment intervals. The patient and clinician remain blinded to treatment assignment during these intervals, and the patient maintains a quantitative record of the outcome of interest during each period. After patient and clinician are unblinded, they can compare treatment effects during active periods with those observed during control periods. Investigators have used *n*-of-1 trials to assess reversible outcomes of stable diseases, such as treatments for pain or dyspnea (27). They cannot be conducted for a condition such as hypertension, in which adverse outcomes are rare, catastrophic, and long delayed.

We can rarely translate with certainty the average benefit reported in randomized clinical trials to a precise assessment of treatment benefit for an individual patient. Subgroup analyses from randomized clinical trials can refine average effects of treatment into subgroup-specific effects, but analysis of a subgroup small enough to include all of the relevant risk factors of an individual patient may lack the precision necessary to be clinically useful. The number needed to treat equals 1 only under the constrained conditions of an *n*-of-1 randomized clinical trial. The language of populations can bring us closer to informing our patient about the consequences of treatment but cannot convey all that must be said.

The Language of Individuals: From Subgroups to the Clinical Encounter

Although you prescribe treatment to your patient with hypertension to help him individually, the number needed to treat makes explicit the fact that the benefits of treatment are realized by a population.

You cannot know in advance whether your patient's decision to accept treatment will prevent him from having a stroke. If, years later, he asks whether he can stop taking his antihypertensive medication, you cannot know whether he has avoided an adverse event that would otherwise have occurred. His scrupulous adherence is unlikely to be rewarded, and his failure to comply will rarely be penalized. All you can say to him with assurance is that on the basis of the best available evidence, everything possible is being done to prevent an adverse outcome.

The choice between the relative risk reduction and the number needed to treat poses a dilemma for the physician. Because both statistics are calculated from P_C and P_T, both are intrinsically accurate portrayals of the results of a randomized clinical trial. We should tell our patients the truth, but which truth should we tell them? If we are to inform them fully about the benefits of treatment, we should use the language of populations. But if a treatment is to benefit a population, our patients must make daily decisions to adhere to it. How many patients are sufficiently committed to the health of the population that they will take medications for years, knowing that only some will benefit if all comply? No research exists to help us decide which strategies of risk communication promote adherence to therapy in clinical practice.

As physicians, we need to become bilingual—that is, we must speak the language of populations as well as the language of individual patients (30). If we are fluent in both languages, we will constantly remind ourselves that we offer our patients choices, not treatments; patients are the final arbiters who must decide on a daily basis whether to take their medication. At the same time, we will remind our patients of the equally important truth that whatever their treatment choices, the outcomes of health care are uncertain. Members of the medical profession have never been comfortable recognizing or sharing uncertainty (31–33). But in the long run, the attempt to protect patients from uncertainty may backfire by jeopardizing trust and impeding full discussion of treatment options, including the option to do nothing (34). It may even be advisable to risk sharing the uncertainties of therapeutic decisions with our patients; they will rarely be surprised to learn that our treatments do not always succeed or that we cannot guarantee good outcomes (33). As a substitute for false certainties, we can offer our commitment to care for our patients as individual persons whatever their choices and outcomes (35).

If you learn to speak the language of populations as well as the language of individual persons, you may be able to say to your patient, "If 10 patients similar to you are treated for 5 years, 1 is likely to avoid a stroke or cardiac event. Although I can't be sure that you will be the one to benefit, I would recommend it because your risk is high and because then we'll both know that we are doing our best to prevent bad outcomes. And whatever you decide, I will care for you." By learning to speak both of these languages, physicians can reconcile the number needed to treat with another statistic—the number needed to care, which always equals one.

John F. Steiner, MD, MPH
University of Colorado Health Sciences Center
Denver, CO 80262

Acknowledgments: The author thanks Ralph Gonzales, MD MSPH; Thomas MacKenzie, MD, MSPH; and Robert Anderson, MD, for helpful suggestions.

Requests for Reprints: John F. Steiner, MD, MPH, University of Colorado Health Sciences Center, Box B-180, 4200 East Ninth Avenue, Denver, CO 80262.

Ann Intern Med. 1999;130:618-622.

References

1. **Sackett DL.** Readers' guide for therapy: was the assignment of patients to treatments randomized? [Editorial] ACP J Club. 1991 May-June;114:A12.
2. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). SHEP Cooperative Research Group. JAMA. 1991;265: 3255-64.
3. **Feinstein AR.** Invidious comparisons and unmet clinical challenges [Editorial]. Am J Med. 1992;92:117-20.
4. **Berwick DM, Fineberg HV, Weinstein MC.** When doctors meet numbers. Am J Med. 1981;71:991-8.
5. **Schwartz LM, Woloshin S, Black WC, Welch HG.** The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966-72.
6. **Tversky A, Kahneman D.** The framing of decisions and the psychology of choice. Science. 1981;211:453-8.
7. **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.
8. Glossary. ACP J Club. 1997 Jan-Feb;126:28.
9. **Forrow L, Taylor WC, Arnold RM.** Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med. 1992;92:121-4.
10. **Naylor CD, Chen E, Strauss B.** Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med. 1992;117:916-21.
11. **Bobbio M, Demichelis B, Giustetto G.** Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet. 1994;343:1209-11.
12. **Bucher HC, Weinbacher M, Gyr K.** Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. BMJ. 1994;309:761-4.
13. **Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM.** The framing effect of relative and absolute risk. J Gen Intern Med. 1993;8:543-8.
14. **Hux JE, Naylor CD.** Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? Med Decis Making. 1995;15:152-7.
15. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. JAMA. 1984;251:351-64.
16. **Frick MH, Elo O, Haapa K, Heinonen OP, Heinsalmi P, Helo P, et al.** Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. N Engl J Med. 1987;317:1237-45.
17. **Kalet A, Roberts JC, Fletcher R.** How do physicians talk with their patients about risks? J Gen Intern Med. 1994;9:402-4.
18. **Detsky AS, Naglie IG.** Subgroup analyses: primary and secondary [Editorial]. ACP J Club. 1995 May-June;122:A12-4.
19. **Yusuf S, Wittes J, Probstfield J, Tyroler HA.** Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA. 1991;266:93-8.
20. **Oxman AD, Guyatt GH.** A consumer's guide to subgroup analyses. Ann Intern Med. 1992;116:78-84.
21. **Curb JD, Pressel SL, Cutler JA, Savage PJ, Applegate WB, Black H, et al.** Effect of diuretic-based antihypertensive treatment on cardiovascular disease risk in older diabetic patients with isolated systolic hypertension. Systolic Hypertension in the Elderly Program Cooperative Research Group. JAMA. 1996;276:1886-92.
22. **Browner WS, Newman TB.** Are all significant *P* values created equal? The analogy between diagnostic tests and clinical research. JAMA. 1987;257: 2459-63.

23. **Atkins CD.** A clinician's view of statistics. J Gen Intern Med. 1997;12:500-4.
24. **Elson RB, Connelly DP.** Computerized decision support systems in primary care. Prim Care. 1995;22:365-84.
25. **Braitman LE, Davidoff F.** Predicting clinical states in individual patients. Ann Intern Med. 1996;125:406-12.
26. **McQuay HJ, Moore RA.** Using numerical results from systematic reviews in clinical practice. Ann Intern Med. 1997;126:712-20.
27. **Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, et al.** A clinician's guide for conducting randomized trials in individual patients. CMAJ. 1988;139:497-503.
28. **Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT.** The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. Ann Intern Med. 1990;112:293-9.
29. **Larson EB, Ellsworth AJ, Oas J.** Randomized clinical trials in single patients during a 2-year period. JAMA. 1993;270:2708-12.
30. **Greenlick MR.** Educating physicians for population-based clinical practice. JAMA. 1992;267:1645-8.
31. **Katz J.** The Silent World of Doctor and Patient. New York: Free Pr; 1984.
32. **Naylor CD.** Grey zones of clinical practice: some limits to evidence-based medicine. Lancet. 1995;345:840-2.
33. **Logan RL, Scott PJ.** Uncertainty in clinical practice: implications for quality and costs of health care. Lancet. 1996;347:595-8.
34. **Reiser SJ.** Words as scalpels: transmitting evidence in the clinical dialogue. Ann Intern Med. 1980;92:837-42.
35. **Cassell EJ.** Doctoring: The Nature of Primary Care Medicine. New York: Oxford Univ Pr; 1997.

I visited Johns Hopkins when the hospital was beginning to perform autologous bone marrow transplants on patients in first remissions. Until that point, due to the riskiness of the procedure, transplants were performed only as a last resort on patients who had already relapsed and gained a second, or even third, remission. I had come looking for information to help me decide if I should be one of the first to jump right into the transplant arena or take my chances with the 80 percent probability of a recurrence. At that time we met a doctor who stunned us by graciously inviting us into his office at precisely the agreed-upon hour of our appointment. After all my experiences, after all the debasing episodes at the outpatient clinic, this was the first time that I had ever been seen by a doctor without enduring an excruciatingly long wait.

I sat down across a cluttered wooden desk and looked up to notice a small framed drawing hanging over the light switch on the wall. The drawing was a simple rendering of a hammer labeled with the letters "BMT," and below it, at the bottom of the picture, was the printed message "When all you have is a hammer, everything looks like a nail."

I repeated the saying in my head a few times while I stared at the picture. Slowly, through the confusion, I recognized that the letters "BMT" stood for bone marrow transplant.

"Just a little reminder," the doctor said. I hadn't known that he'd noticed me studying the drawing. "It helps keep us honest."

I was astonished. I was an in-the-flesh, first-person to a group of doctors, or, who cares? Just one doctor, who framed and hung a cartoon for the purpose of reminding himself to question his own judgment! I wanted to run around the desk and hug this man.

Evan Handler
Time on Fire
New York: Little, Brown; 1996

Submitted by:
Edmund O'Rourke
Philadelphia, PA 19106

Submissions from readers are welcomed. If the quotation is published, the sender's name will be acknowledged. Please include a complete citation (along with the page number on which the quotation was found), as done for any reference.—*The Editor*