

has been observed elsewhere in biology. Consider this excerpt from Barry Lopez's book *Arctic Dreams*<sup>4</sup>:

Many Western biologists . . . comprehend that, objectively, what they are watching is deceptively complex. . . . They know that while experiments can be designed to reveal aspects of the animal, the animal itself will always remain larger than any set of experiments. They know they can be very precise about what they do, but that does not guarantee they will be accurate. They know that the behavior of an individual animal may differ strikingly from generally recognized behavior of its species; and that the same species may behave quite differently from place to place, from year to year.

The same statement can be made even more strongly on research conducted with human animals.

The answer to the question posed in the opening paragraph, by the way, is option number two: Rogers and Rohrbaugh absolutely nail a narrow and ultimately not very interesting question with tangential relevance to the general use of genograms in daily practice. In doing so, however, the authors place us in their debt by clearly illustrating the complexity of studying questions lying outside traditional biomedical boundaries. The next researchers who wish to examine genograms should benefit from the authors' experience by seeking methods that have a better chance of producing results generalizable to daily practice.

The biopsychosocial model has already been discarded by some because it does not go far enough in eradicating linear causal thinking from research and clinical practice.<sup>5</sup> Philosophers of science have moved far ahead in developing a naturalistic perspective based on postpositivist theories that refute much of biomedicine's current scientific methodology.<sup>6</sup> Conceptions of what is and is not science are changing rapidly. Researchers in family medicine have much to gain by being first at the boundary.

Alfred O. Berg, M.D., M.P.H.  
Seattle, WA

## References

1. Rogers JC, Rohrbaugh M. The SAGE-PAGE trial. Do genograms make a difference? *J Am Board Fam Pract* 1991; 4:319-26.

2. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977; 196: 129-36.
3. Pellegrino ED. Patient care—mystical research or researchable mystique? *Clin Res* 1964; 12:421-5.
4. Lopez B. *Arctic dreams*. New York: Bantam Books, 1989:269.
5. Foss L, Rothenberg K. *The second medical revolution: from biomedicine to infomedicine*. Boston: New Science Library, 1987.
6. Lincoln YS, Guba EG. *Naturalistic inquiry*. Beverly Hills, CA: Sage Press, 1985.

## Measures Of Clinical Effectiveness: The Numbers Needed To Treat

---

What are the best measures of clinical effectiveness to use when presenting health care decisions to individual patients and when determining what health care needs take priority when resources are restricted? The measures are numerous: disease-specific and all-cause mortality rates, morbidity rates, years of life lost before a specified age, population-attributable risk, relative risk reduction, odds ratios, absolute or attributable risk reduction, to name a few. Each has advantages and disadvantages related to the question being asked and to the inherent statistical properties of the measure.

In this issue of the *Journal*, Grumbach<sup>1</sup> applies another measure, "the number needed to treat," to the question of how best to measure the consequences of pharmacologic management of hypercholesterolemia using outcome and side effect data from five major clinical trials. The statistic "the number needed to treat" (NNT) provides the number of persons needed to be treated in order to reach a given end point, for example, prevention of one myocardial infarction, prevention of one death, or causation of side effects in one patient. NNT is the inverse of the absolute risk reduction (ARR), which, in

---

Submitted 19 July 1991.

From the Department of Family Medicine, University of Washington, Seattle. Address reprint requests to William E. Neighbor, Jr., M.D., Department of Family Medicine, RF-30, University of Washington, Seattle, WA 98195.

turn, is the difference in rates of event between the treatment and control groups. ARR can be calculated indirectly by multiplying the relative risk reduction by the event rate. Laupacis, Sackett, and Roberts<sup>2</sup> in 1988 reviewed the properties of this statistic and applied it to the benefits of treating a number of conditions, including treating hypercholesterolemia with cholestyramine. Grumbach applies the statistic more broadly to the issue of treating hypercholesterolemia. I will not focus on the implications of the review for screening and management of hypercholesterolemia. Instead, because the author's statements are based on inferences from NNT statistics, I will examine the nature of the statistic itself and limitations on inference from the statistic.

In this respect, three questions need to be addressed: (1) what is the level of confidence in the statistic; (2) what are the assumptions on which the statistic is based, particularly with respect to adjusting for treatment effects over time; and (3) how does the NNT calculated for different therapies for hypercholesterolemia compare with NNT for therapies for other conditions?

### What Is the Level of Confidence in the Statistic?

Any statistic has a calculable measure of confidence, which is determined by the sample sizes of the treatment and control groups and, in the case of proportions, the magnitude of the proportions  $p_1$  and  $p_2$  (e.g., event rates in the control and treatment groups) and their complements  $1-p_1$  and  $1-p_2$ . The range within which the true value of the statistic will occur with 95 percent confidence is the 95 percent confidence interval (CI). Table 1 shows the approximate 95 percent

CIs of NNT for total death, cardiovascular morbidity, and death plus cardiovascular morbidity from the Helsinki Heart Study<sup>3</sup> and the niacin arm of the Coronary Drug Project.<sup>4</sup> These intervals were arrived at by calculating the 95 percent CIs for the ARR, that is,  $p_2 - p_1$ , using the method described by Fleiss.<sup>5</sup> Because  $NNT = 1/(p_2 - p_1)$ , the 95 percent CI for NNT was taken to be the inverse of the CI end points for the ARR. This method is similar to that referenced by Laupacis, et al.<sup>2</sup>

This analysis shows that in the niacin arm of the Coronary Drug Project, while NNT is 16, we are 95 percent confident that the true value could be as low as 10 or as high as 36 to prevent one death over 15 years. Similarly, in the Helsinki study we are 95 percent confident that true value for NNT to prevent one nonfatal myocardial infarction (cardiovascular morbidity) could be as low as 43 but also could be as high as 358 patients over 5 years. This information will influence interpretation of the NNT statistic in clinical practice, though it is not provided in either the paper by Grumbach or that by Laupacis, et al.

The statistics for death plus cardiovascular morbidity in the Helsinki study show two important properties of the NNT statistic. First, the CI for ARR includes 0, indicating it is not statistically significant, and, by inference, the NNT is not statistically significant (and the 95 percent CI does not even include 83, the NNT), even though the relative risk reduction statistic cited in Grumbach's review is significant. I believe Grumbach makes the error of assuming that if relative risk reduction is statistically significant, the NNT statistic is significant. This is not the case. NNT and ARR statistics are not as "robust" as the relative risk

**Table 1. Numbers Needed to Be Treated and 95% Confidence Intervals (CI).**

Study	Outcome	Incidence per		Incidence per		Absolute Risk Reduction (95% CI)	Numbers Needed to be Treated (95% CI)
		1000 in Treatment Group	Sample Size of Treatment Group	1000 in Control Group	Sample Size of Control Group		
Coronary drug project, niacin group	Total death	520	1119	582	2789	0.0625 (0.027, 0.097)	16 (10, 36)
Helsinki heart study	CV morbidity	22	2051	35	2030	0.013 (0.0028, 0.023)	77 (43, 358)
	Death plus CV morbidity	44	2051	56	2030	0.012 (0.0014, 0.025)	83 (-729, 39)*

\*See text for discussion.  
CV = cardiovascular.

reduction statistic. This is one reason why ARR has not been as widely employed as other measures of effect.

Second, because of the mathematical relation of ARR to NNT, the smaller the ARR, that is, the closer it comes to indicating no treatment advantage, the larger the NNT becomes. A disadvantage of NNT is that while other statistics allow an easy statement of the null hypothesis, for example, "ARR equals 0" or "relative risk equals 1," the null hypothesis for the NNT statistic is "NNT equals infinity." The difficulty is that the researcher is left to decide how large NNT needs to be before he or she says it approaches infinity close enough that the null hypothesis can be accepted and that, therefore, there is no statistically significant treatment effect.

### **What Are the Assumptions and Limitations of the Statistic?**

The benefits of the NNT statistic, such as that it takes into account the magnitude of risk in the untreated group, that it is intuitively meaningful, and that it can be applied to compare side effect experiences, are described by both Grumbach and Laupacis, et al. Additionally, they point out limitations common to most statistics (e.g., comparison of summary statistics from various studies must be made with knowledge of differences in methods and criteria for sample selection, differences in compliance rates, different measurements employed, and differences in duration of treatment and follow-up). Laupacis, et al. also point out that if NNT equals 10, one patient will benefit, but the statistic does not provide any information about the experience of the other 9 (except that they did not benefit from treatment).<sup>2</sup>

There are two other limitations in interpretation of the NNT statistic that might be emphasized. First, the term *number needed to treat* implies that this is the actual number of patients we as physicians will need to have under treatment to see an effect. This assumption is somewhat misleading because in most clinical trials of treatment of hypercholesterolemia, 100 percent compliance was not achieved, but nevertheless a positive treatment effect was found. Second, as with other statistics, even if treatment A has a smaller NNT than treatment B for the same time period, one needs to be cautious in inferring that resources are better allocated to

treatment A. The profiles distribution of benefits and risks over time might be significantly different. For example, all of the risks of treatment A may become manifest in the first few years following treatment, whereas those of treatment B can be spread more equally over time. Some patients, because of quality-of-life evaluations might choose treatment B over A even though one treatment has a smaller NNT than another for the same time period. An example of treatment A might be carotid endarterectomy, which probably provides greater benefit than medical treatment after 3 years of follow-up but at the price of a higher risk of death or disability accrued in the first year of treatment.

Third, the data from the Helsinki study, the Coronary Drug Project, and the Lipid Research Clinics study<sup>6</sup> indicate that, at least for cardiac end points, the benefits of treatment are not linearly related to time. For example, cholestyramine did not show any benefit in the first 3 years, but did by 7 years of follow-up. (Unfortunately, none of these studies to my knowledge provided graphs of treatment effects on non-cardiac end points over time.) This finding violates the assumption of equal distribution of benefits occurring over time, which is inherent in the equation used by Grumbach and Laupacis, et al. to adjust the NNT statistic for differences in duration of studies:

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

where NNT:T and NNT:S are the numbers needed to be treated for T and S years, respectively.<sup>2</sup> In the three studies cited above, the actual benefit is greater than the calculated future benefit. For example, using data from Grumbach's Table 3, NNT:6.2 years equals 100 for the Coronary Drug Project-niacin group, and adjusted to 15 years, NNT calculates to be 41. The actual NNT at 15 years was 16, however, a substantially smaller number needed to be treated than calculated.

### **How Does the NNT for Hypercholesterolemia Compare with NNT for Other Conditions?**

Laupacis, et al. offer a wider perspective on the numbers needed to be treated for various conditions than what is presented by Grumbach, and it is worth repeating here.<sup>2</sup> In Table 2, I have adjusted the NNT:5 years statistics

**Table 2. Numbers Needed to be Treated for 10 Years for Various Therapies and Conditions.\***

Therapy and Condition	Events	No. Needed to Treat: 10 Years
Stepped care for diastolic blood pressure of 115–129 mmHg	Death, stroke, myocardial infarction	1.5
Left main coronary-artery bypass surgery	Death	3
Aspirin for transient ischemic attacks	Death, stroke	3
Isoniazid for inactive tuberculosis	Active tuberculosis	48
Stepped care for diastolic blood pressure 90–109 mmHg	Death, stroke, myocardial infarction	72
Niacin for hypercholesterolemia	Death plus cardiovascular morbidity	14
Gemfibrozil for hypercholesterolemia	Death plus cardiovascular morbidity	42
Cholestyramine for hypercholesterolemia	Death plus cardiovascular morbidity	53

\*Modified from Laupacis, et al.<sup>2</sup>

calculated by Laupacis, et al. to NNT:10 years to facilitate comparison with the numbers provided by Grumbach, which are included in the table. What is interesting is that one needs to treat fewer patients of the kind included in the hypercholesterolemia clinical trials than patients having a diastolic blood pressure of 90 to 109 mmHg to avert one death or cardiovascular morbidity event. It also appears that on average fewer patients need to be treated for hypercholesterolemia to prevent death or cardiovascular morbidity than need to be treated for inactive tuberculosis to prevent one case of active tuberculosis. (Of course, cost, duration of treatment, and other considerations, such as communicability of tuberculosis, need to be factored if one needs to determine how to distribute health care resources.) It would be interesting to extend the comparison further to other therapies and conditions.

### Conclusions

The paper by Grumbach contributes an important perspective to the treatment of hypercholesterolemia. The perspective needs to be seen in light of the properties of the statistical method employed. In sum, the statistic *number needed to treat* has several advantages over other

statistics as a way of looking at consequences of treatment. These advantages include being intuitively meaningful and taking into account the magnitude of risk in the untreated group. Limitations of the statistic include the fact that it is less robust than other statistics and so may not be statistically significant when other statistics are and that it does not lend itself easily to testing of the null hypothesis (no treatment effect). Those drawing statistical inferences from NNT should take into account the level of confidence in the statistic, be guarded when projecting to time intervals beyond the duration of the study, and should take into account the compliance rates of the studies from which the statistic is derived. Finally, one should not only compare different treatments for the same condition but also compare treatments for different conditions when judging relative treatment effects.

What the concept *number needed to treat* makes most apparent is that, though we, as physicians, may do no harm in the classical sense, for many conditions we often benefit only a few of the patients to whom we offer treatment. As the review by Grumbach points out, one of the aims of medical research is to help identify those patients who will benefit.

William E. Neighbor, Jr., M.D.  
Seattle, WA

### References

1. Grumbach K. How effective is drug treatment of hypercholesterolemia? A guided tour of the major clinical trials for the primary care physician. *J Am Board Fam Pract* 1991; 4:437-45.
2. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318:1728-33.
3. 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. Safety of treatment changes in risk factors, and incidence of coronary heart disease. *N Engl J Med* 1987; 317:1237-45.
4. Canner PL, Berge KG, Wenger NK, Stamler J, Friedman L, Prineas RJ, et al. Fifteen year mortality in coronary drug project patients: long-term benefit with niacin. *J Am Coll Cardiol* 1986; 8:1245-55.
5. Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley & Sons, 1981:101-2.
6. Lipid Research Clinics Program. The lipid research clinics coronary primary prevention trial results. Reduction in incidence of coronary heart disease. *JAMA* 1984; 251:351-64.