JAMA Guide to Statistics and Methods

Number Needed to Treat
Conveying the Likelihood of a Therapeutic Effect

Jeffrey L. Saver, MD; Roger J. Lewis, MD, PhD

Effectively communicating clinical trial results to patients and clinicians is a requirement for appropriate application in clinical practice. In a recent issue of *JAMA*, Zhao et al[1] reported the results from a randomized clinical trial comparing dual antiplatelet therapy with aspirin monotherapy for preserving saphenous vein graft patency in 500 patients undergoing coronary artery bypass grafting. Dual antiplatelet therapy was found to be superior to aspirin monotherapy. The authors[1] used the number needed to treat (NNT) to communicate effect size, reporting that for every 8 patients treated with dual agents rather than aspirin alone, 1 additional patient would achieve saphenous graft patency at 1 year. The NNT may be defined as the number of patients who need to be treated with one therapy vs another for 1 additional patient to have the desired outcome. Since its first description 30 years ago,[2] the NNT has become an important means to express the magnitude of benefit conferred by a therapy.[3]

Explanation of the Concept

What Is the NNT?

When a clinical trial is completed, the fraction or proportion of patients experiencing the desired outcome is reported for the active and control groups. The NNT is derived from these values and indicates the magnitude of the therapy's treatment effect on the disease observed in the clinical trial. The NNT is computed by dividing 100 by the difference between the percentage response of the treatment group from that of the control group. Alternatively, the NNT is calculated by taking the reciprocal of the absolute risk reduction between the groups. The NNT indicates how many patients must be managed on average with active rather than control therapy to achieve 1 additional good outcome.

The number needed concept may be applied to many types of outcomes from both therapeutic and diagnostic studies. When a therapy increases desirable outcomes, the resulting value is the number needed to benefit (more often denoted as just NNT). When a therapy increases adverse events, the resulting value is the number needed to harm. When applied to diagnostic strategies, the resulting values are the number needed to screen for tests in asymptomatic individuals, and the number needed to diagnose for tests in symptomatic individuals.

Why Is the NNT Important?

The NNT is intuitively understandable by patients and clinicians. It is also quantitative, facilitating decision making when selecting among available therapeutic strategies. By including a 95% confidence interval (CI) around the observed NNT, the uncertainty in the benefit also can be communicated effectively.

Other well-established indices of treatment effect are not well suited for this purpose. For example, a statistically significant *P* value conveys statistical rather than clinical significance. The *P* value suggests there will be a difference in outcomes associated with choice of therapy, but not how large that difference will be.

Risk ratios and odds ratios convey the relative rather than the absolute differences in outcomes with different treatments.[4] They are interpretable only if the event rate in the control comparator group is also stated, and then require mental calculation not readily performable by many decision makers. For example, a treatment that increases by 1.5-fold the frequency of a desirable outcome (risk ratio = 1.5) will help only 1 of every 100 patients if the base rate of the desirable outcome in the control group is 2% (increased in the active group to 3%), but will help 20 of every 100 patients if the base rate of the desirable outcome in the control group is 40% (increased in the active group to 60%). In contrast, the NNT conveys the absolute size of differences in outcome proportions with different treatments in a readily interpretable manner.

Limitations and Alternatives to the NNT

Despite its several advantages, the NNT metric does have important limitations, and alternative indices of treatment effect magnitude are available that provide helpful complementary information. First, the NNT combines 2 proportions (the fraction of treatment success in each treatment group) into a single number, which sacrifices information. For example, the same NNT may represent increases in treatment success (eg, from 5% to 15% or from 85% to 95%) that may be viewed differently by patients and clinicians.

A second limitation is that it can be challenging to compare and integrate different NNTs because their values are expressed as fractions with different denominators. In contrast, the natural frequency metric (most often stated as benefit per hundred and harm per hundred) more readily facilitates comparisons because it expresses the treatment effect magnitude using a uniform (100) and familiar (from percentages) denominator.[5,6]

For example, consider the following statements describing the same treatment effect. The NNT to prevent 1 myocardial infarction is 25 patients, to prevent 1 ischemic stroke is 50, and to cause 1 major bleeding event is 33. For every 100 patients treated, 4 fewer will have a myocardial infarction, 2 fewer an ischemic stroke, and 3 more a major bleeding event. The different framing of the clinical trial result provided by the NNT and benefit per hundred can influence decision making despite the fact that they are numerically equivalent.

The NNT aligns more closely with the patient perspective because the patient will often be making a particular treatment decision only once ("my chance of benefit is 1 in X"). The benefit per hundred aligns more closely with the perspective of the clinician, who will often be making the same treatment decision tens of times during a career ("out of 100 patients I treat, I will help X").[7]

A limitation of the NNT shared by the natural frequency is that randomized clinical trial results fully specify NNT values only for binary outcomes (such as the occurrence of an infection, a rash, or death), but not for ordinal or continuous outcomes (such as reduced pain or degree of disability). This drawback has been partially mitigated by the development of methods that provide estimated NNT values for ordinal or continuous outcomes using automated or content expert–informed derivation techniques.[8] However, these methods require additional, often untestable, assumptions to estimate the distribution of an observed treatment group benefit among individual patients because the same clinical trial result can arise when many patients experience a small individual benefit or when fewer patients experience a large individual benefit.

Another limitation is that the NNT reflects the number not the importance of events. Different types of events are each given their own separate NNT values and the resulting quantitative statements may encourage overweighting of less important outcomes. For example, a therapy is clearly of substantial net benefit even if it has a nominally lower NNT to harm of 3 for a minor adverse effect (such as transient mild headache) accompanying a nominally higher NNT to benefit of 5 for a major beneficial effect (such as fatal cardiac failure). An alternative approach is to integrate multiple outcomes into a single measure of treatment effect using health-related utility values for each of the outcomes.[9,10] Once event values are converted to this single consistent measure, a NNT to achieve any given magnitude of benefit on the utility scale can be derived.[6] For example, the "number needed to save one life" was recently used to express the number of patients with acute ischemic stroke treated with thrombectomy required to achieve the same total benefit as saving the life of 1 patient who would have died and achieving a normal neurological outcome.[6]

Further limitations include that the NNT does not convey the financial costs and benefits of treatments and only expresses the magnitude of effect expected for a prototypical patient, reflecting the aggregate characteristics of the population enrolled in a particular clinical trial. In contrast, each individual patient has distinctive features modifying the baseline risk and treatment response. In addition, when patient outcomes vary over time, the reported NNT reflects the benefit at a particular time point and several different NNT values might be needed to capture varying benefits (eg, at early, middle, or late stages of the treatment course).

How Was the Concept of NNT Applied in This Particular Study?

In the Results section of the study by Zhao et al,[1] the primary efficacy end point findings were reported including each group's individual outcome proportions and 95% CIs, the relative treatment effect magnitude (relative risk, 0.48 [95% CI, 0.31-0.74]), the absolute treatment effect magnitude (risk difference, 12.2% [95% CI, 5.2%-19.2%]), and the statistical significance ($P < .001$). The authors restated this result as an NNT of 8 (the approximate reciprocal of 12.2%). Reporting the findings this way conveyed the probability of benefit in a clinically useful manner. The authors did not provide the 95% CI around the NNT value. The absence of a 95% CI around the NNT improves readability, but somewhat obscures the degree of uncertainty around the estimated value. An NNT to harm value was not provided for the increase in minor bleeding events that also was observed with double antiplatelet therapy. However, it is prudent in primary trial reports to state NNT values only for the lead efficacy and safety end points that were the prespecified focus of hypothesis testing.

How Should the NNT Be Interpreted in Zhao et al?

The absolute risk difference of 12.2% with the 95% CI of 5.2% to 19.2% reported by Zhao et al[1] indicates that approximately 8 patients (given by the reciprocal of 0.122) need to be treated with dual antiplatelet therapy as opposed to aspirin monotherapy to avoid 1 case of saphenous vein graft occlusion. However, the data are also consistent with this number being as low as 5 or as high as 19 (given by the reciprocals of 0.192 and 0.052, respectively). These values capture both the probability of benefit for the individual patient (approximately 1 in 8) and the uncertainty in that probability.

ARTICLE INFORMATION

Author Affiliations: Department of Neurology, Ronald Reagan–UCLA Medical Center and David Geffen School of Medicine, University of California, Los Angeles (Saver); Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Lewis); Department of Emergency Medicine, David Geffen School of Medicine, University of California, Los Angeles (Lewis); Berry Consultants LLC, Austin, Texas (Lewis).

Corresponding Author: Jeffrey L. Saver, MD, Reed Neurologic Research Center, 710 Westwood Plaza, Los Angeles, CA 90095 (jsaver@mednet.ucla.edu).

Section Editor: Edward H. Livingston, MD, Deputy Editor, *JAMA*.

Published Online: February 7, 2019. doi:10.1001/jama.2018.21971

Conflict of Interest Disclosures: None reported.

Disclaimer: Dr Saver is an associate editor of *JAMA*, but he was not involved in any of the decisions regarding review of the manuscript or its acceptance.

REFERENCES

1. Zhao Q, Zhu Y, Xu Z, et al. Effect of ticagrelor plus aspirin, ticagrelor alone, or aspirin alone on saphenous vein graft patency 1 year after coronary artery bypass grafting: a randomized clinical trial. *JAMA*. 2018;319(16):1677-1686.

2. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728-1733.

3. Mendes D, Alves C, Batel-Marques F. Number needed to treat (NNT) in clinical literature. *BMC Med*. 2017;15(1):112.

4. Norton EC, Dowd BE, Maciejewski ML. Odds ratios—current best practice and use. *JAMA*. 2018;320(1):84-85.

5. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine: communicating statistical information. *Science*. 2000;290(5500):2261-2262.

6. Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2018;378(1):11-21.

7. Peng J, He F, Zhang Y, et al. Differences in simulated doctor and patient medical decision making. *PLoS One*. 2013;8(11):e79181.

8. Saver JL. Optimal end points for acute stroke therapy trials. *Stroke*. 2011;42(8):2356-2362.

9. Irony TZ. The "utility" in composite outcome measures. *JAMA*. 2017;318(18):1820-1821.

10. Hong KS, Ali LK, Selco SL, et al. Weighting components of composite end points in clinical trials. *Stroke*. 2011;42(6):1722-1729.