

Number Needed to Treat: What It Is and What It Isn't, and Why Every Clinician Should Know How to Calculate It

Leslie Citrome, MD, MPH

Number needed to treat (NNT) is a number that helps the clinician assess the clinical relevance of a statistically significant result. Case in point: you are being told that treatment A is better than treatment B because treatment A results in 25% more “responders” than treatment B. Moreover the *P* value is less than .0001. Is this compelling evidence that should impact medical decision making?

The astute clinician will already be asking about the definition of response (Is a 20% decrease in symptoms on a rating scale clinically important? Is 30%? Is 50%?). The next question is to clarify what the rates of response are for treatment A vs treatment B. The relative difference of a “25% better” claim for A would be of little importance if the actual rates of response were abysmally low. Finally, it needs to be recognized that the *P* value doesn't tell us anything about the clinical relevance or the effect size of the treatment. It merely tells us how likely the result can or cannot be due to chance. A very low probability that the result is due to chance gives us confidence that we are dealing with a real difference, but it doesn't inform us about the magnitude of this difference. Moreover, very large sample sizes can make even the smallest difference highly statistically significant, yet upon closer inspection that difference can be clinically irrelevant.¹

Of all the measures of effect size, NNT is perhaps the easiest to calculate and is likely the most clinically intuitive. It answers the question: How many patients do I need to treat with treatment A vs treatment B before I would expect to encounter 1 additional outcome of interest, such as response? In our example in which A is “25% better” than B in responder rates, let's say that for all patients receiving either A or B, response for A was 12.5% and that for B, 10%. A rate of 12.5% is indeed “25% better” than 10%, but is this clinically relevant when treating patients during the normal course of the day? We can calculate the NNT; it turns out to be 40. We would need to treat 40 patients with treatment A rather than treatment B before expecting 1 additional responder. In the course of a day, the clinician would be hard-pressed to detect a difference in responder rates between A and B. In fact, there may be other characteristics of the patient and the treatment that would quite likely be far more important, such as prior history of response to A or B (or drugs similar to A or B), an individual patient's sensitivity toward certain side effects that may be more frequently encountered with A vs B, or the cost and availability of the medication in question.

How NNT Is Calculated

The beauty of NNT is in its simplicity, in both concept and calculation. NNT is easily determined by subtracting the rates in question (ie, response rates, remission rates), and calculating the reciprocal of this difference. Mathematically, it can be expressed as

$$\text{NNT} = 1/(\text{Rate}_1 - \text{Rate}_2).$$

For our example above, $\text{NNT} = 1/(0.125 - 0.10) = 1/0.025 = 40$.

Calculating the confidence interval (CI) for the NNT is more complex. The confidence interval will inform us about the precision of the NNT. Usually a 95% CI is calculated, providing a lower bound and an upper bound, whereby with 95% certainty, the true NNT is said to be between. With a non-statistically significant result, the NNT is difficult to interpret, and the 95% CI is bracketed by a negative and a positive number, with infinity included as a possible value for the true NNT (an infinite NNT would mean that there is

no difference between the interventions being compared on the outcome of interest). Formulas for the CI, as well as for other measures of effect size can be found elsewhere.²

How NNT Is Interpreted

The smaller the NNT, the larger the effect size difference between the 2 interventions being compared. For example, a NNT of 2 would mean you would expect to encounter an additional outcome of interest for every 2 patients treated with 1 treatment vs the other. It would be a large effect size difference. A NNT of 100 would mean you would need to treat 100 patients before expecting to encounter an additional outcome of interest—unlikely to be noticed in routine clinical practice.

Because smaller means more important, the convention is to always round up to the next highest whole number when calculating a NNT. For example, a NNT of 2.1 or 2.7 would both be rounded up to 3 because we do not want to exaggerate a potential difference between 2 treatments and thus err on the side of caution when describing these differences.

The smallest NNT in the real world is 2 because no treatment is 100% efficacious, and after rounding up, even a NNT of 1.1 becomes a NNT of 2. By “rule of thumb,” a NNT less than 10 is ordinarily considered clinically meaningful because a treatment difference would be routinely encountered in day-to-day clinical practice.

NNT's Evil Twin: NNH

The concept of NNT can be used to compare treatments in terms of potential adverse events. By convention, the term is *number needed to harm* (NNH). NNH is calculated the same way as NNT but is used to describe the number of patients we would need to treat with treatment A vs treatment B before we would expect to encounter 1 additional adverse outcome of interest. Possible outcomes subject to this calculation include the occurrence of weight gain in excess of a certain threshold, such as 7%; the occurrence of akathisia; or a complaint of sedation. The higher the NNH, the less important the difference between the 2 interventions with respect to the adverse outcome of interest. It is possible to calculate a NNH of drug vs placebo for all of the adverse events reported in a medication's product labeling, which can be used to make indirect comparisons between medications regarding adverse events such as weight gain, sedation, or akathisia.^{2,3}

Some advocate for using the terms *number needed to treat for an additional beneficial outcome* (NNTB) and *number needed to treat for an additional harmful outcome* (NNTH), instead of NNT and NNH, respectively. Both sets of terms are commonly found in the medical literature.

Can We Combine NNT and NNH?

For a treatment choice to be compelling, we would like the NNT to be low and the NNH to be high. This means that we would encounter differences in benefits more often than differences in harms. The ratio of NNH to NNT can illustrate the trade-offs between a specific benefit and a specific harm. This ratio has been termed the *likelihood of being helped vs harmed* (LHH). Great care is required when selecting the NNT and NNH to make this calculation, because each part of the ratio must be clinically relevant and of similar importance to the patient, but it does answer

the question of how much more likely it is for a medication to be associated with a benefit than a harm.^{3,4}

From the perspective of a clinician, one way of appraising a clinical trial comparing a medication to placebo is to calculate the NNT for discontinuation because of lack of efficacy and the NNH for discontinuation because of an adverse event. These rates for discontinuation are commonly provided in clinical trial reports no matter the disease state being treated. A question that can be asked, and quantified, is, how much more often is the benefit encountered (avoidance of discontinuation because of lack of efficacy) compared to a harm (discontinuation because of an adverse event)? A LHH of 5 (ie, NNH/NNT = 5) would mean that the test medication is 5 times more likely to lead to an avoidance of discontinuation because of lack of efficacy than discontinuation because of an adverse event. When available, the NNT for response or remission can also be used to contrast benefit vs harm in the calculation of the LHH ratio. A caveat is that the definition of response or remission has to be clinically meaningful, something for which, at this time, there is no universal agreement for some disorders such as schizophrenia.

What's Wrong With NNT and NNH?

NNT and NNH are only calculable for binary or dichotomous outcomes at a specific point in time. NNT does not capture information about trajectory of improvement. When the data are available, it can be interesting to examine how NNT changes over time as a patient continues to receive a therapeutic intervention.⁵

The assessment of effect sizes for continuous measures, such as point change on a rating scale or kilograms gained over time, requires other techniques.^{6,7} These are substantially more complex to calculate but are necessary when designing clinical trials (to determine sample size requirements) and are useful when attempting to fully understand clinical trial outcomes.

It should be kept in mind that while definitions of response and remission used in determining NNT are relatively standard for some illnesses, such as major depressive disorder or bipolar disorder, the definition of "harm" is widely variable, and can include specific harms such as akathisia, sedation, or body weight gain. Severity of harm can vary as well as the time course of the harm. Thus, it might be difficult to justify comparing the NNT for remission for a severe disorder with the NNH for transient nausea. Alternatively, tardive dyskinesia and agranulocytosis are uncommon harms with a high NNH but for which consequences are severe.

A final caveat: the NNT or NNH is only as good as the data behind it. If the clinical trial that is used to calculate NNT or NNH is flawed, then the NNT or NNH can be useless. If the patients enrolled in the clinical trial are too dissimilar to the patient you are treating, then the data may not be generalizable, and the NNT or NNH may be quite misleading.

Summary

NNT and NNH provide a means for every clinician to rapidly calculate the relevance of a clinical trial result, provided these results are presented as a binary outcome (ie, absence or presence of response, remission, relapse, clinically relevant weight gain, sedation, akathisia, and others). The *P* value does not answer the question of how clinically important the result is. Calculating the NNT or NNH of a statistically significant result can help put that result into clinical perspective.

Author affiliation: Department of Psychiatry, New York University School of Medicine, New York. **Potential conflicts of interest:** Dr Citrome is a consultant for, has received honoraria from, or has conducted clinical research supported by the following: Abbott Laboratories, AstraZeneca Pharmaceuticals, Avanir Pharmaceuticals, Azur Pharma Inc, Barr Laboratories, Bristol-Myers Squibb, Eli Lilly and Company, Forest Research Institute, GlaxoSmithKline, Janssen Pharmaceuticals, Jazz Pharmaceuticals, Merck, Novartis, Pfizer Inc, Sunovion, Valeant Pharmaceuticals, and Vanda Pharmaceuticals. **Funding/support:** None reported. **Disclaimer:** No writing assistance was utilized in the production of this article. **Corresponding author:** Leslie Citrome, MD, MPH, 11 Medical Park Drive, Suite 106, Pomona, NY 10970 (nntman@gmail.com). *J Clin Psychiatry* 2011;72(3):412-413 (doi:10.4088/JCP.11ac06874)
© Copyright 2011 Physicians Postgraduate Press, Inc.

REFERENCES

1. Citrome L. Compelling or irrelevant? using number needed to treat can help decide. *Acta Psychiatr Scand*. 2008;117(6):412-419.
2. Citrome L. Quantifying risk: the role of absolute and relative measures in interpreting risk of adverse reactions from product labels of antipsychotic medications. *Curr Drug Saf*. 2009;4(3):229-237.
3. Citrome L. Adjunctive aripiprazole, olanzapine, or quetiapine for major depressive disorder: an analysis of number needed to treat, number needed to harm, and likelihood to be helped or harmed. *Postgrad Med*. 2010;122(4):39-48.
4. Citrome L, Kantrowitz J. Antipsychotics for the treatment of schizophrenia: likelihood to be helped or harmed, understanding proximal and distal benefits and risks. *Expert Rev Neurother*. 2008;8(7):1079-1091.
5. Cookson J, Gilaberte I, Desai D, et al. Treatment benefits of duloxetine in major depressive disorder as assessed by number needed to treat. *Int Clin Psychopharmacol*. 2006;21(5):267-273.
6. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990-996.
7. Citrome L. Relative vs absolute measures of benefit and risk: what's the difference? *Acta Psychiatr Scand*. 2010;121(2):94-102.

ASCP Corner offerings are not peer reviewed by the *Journal* but are peer reviewed by ASCP. The information contained herein represents the opinion of the author.

Visit the Society Web site at www.ascpp.org