Dissecting clinical trials with 'number needed to treat'

Calculation suggests a study's value to your patients

Leslie Citrome, MD, MPH
Professor of psychiatry
New York University School of Medicine
New York, NY

Director, Clinical Research and Evaluation Facility
Nathan S. Kline Institute for Psychiatric Research
Orangeburg, NY

C linical trials produce a mountain of data that can be difficult to interpret and apply to clinical practice. When reading about studies such as the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) for schizophrenia, you may wonder:

- How large is the effect being measured?
- Is it clinically important?
- Are we dealing with a result that may be statistically significant but irrelevant for day-to-day patient care?

Number needed to treat (NNT) and number needed to harm (NNH)—two tools of evidence-based medicine (EBM, *Box 1*[1,2])—can help answer these questions. This article shows how to calculate NNT and NNH, then applies these tools to published results from CATIE phases 1 and 2.

> **Box 1**
> ## What does 'evidence-based' mean?
>
> **Evidence-based medicine (EBM)** is a process by which a clinician extracts information from the medical literature and applies it in day-to-day patient treatment. Gray and Pinson[1] summarize EBM's 5 steps as:
> - formulate the question
> - search for answers
> - appraise the evidence
> - apply the results
> - assess the outcome.
>
> This is not a trivial task. To help clinicians, EBM pioneers such as Gordon Guyatt, MD, MSc, and Drummond Rennie, MD, have published useful, readable, short reviews of EBM methods in the "Users' Guides to the Medical Literature" in the *Journal of the American Medical Association*.[2]
>
> **Internet resources** also are available, including:
> - Centre for Evidence-Based Medicine, University of Toronto. www.cebm.utoronto.ca
>
> - Eskind Biomedical Library, Vanderbilt University. Evidence-based knowledge portal. www.mc.vanderbilt.edu/biolib/ebmportal/login.html
>
> - Hayward Medical Communications. Evidence-Based Medicine: What is...? series. www.evidence-based-medicine.co.uk/What_is_series.html.

WHAT IS NNT?

NNT helps us gauge effect size—or *clinical* significance. It is different from knowing if a clinical trial result is *statistically* significant.

NNT allows us to place a number on how often we can expect to see a difference between two interventions. If we see a therapeutic difference once every 100 patients (an NNT of 100), the difference between two treatments is not of great concern under most circumstances. But if a difference in

Use confidence intervals to determine if NNT is statistically significant

Calculating number needed to treat (NNT) or number needed to harm (NNH) does not tell you whether the result is statistically significant. You can find out by examining a range of values called the confidence interval (CI).

An NNT with a 95% CI means that the truth probably lies between the lower and upper bounds of the interval with a probability of 95%. A 95% CI with an NNT of 5 to 15 means we have an NNT that with 95% certainty falls between 5 and 15.

Formulas can be used to calculate CIs.[3] One useful online calculator is available at: www.cebm.utoronto.ca/practise/ca/statscal.

Sometimes the lower bound of a CI is a negative number and the upper bound is a positive number (such as –10 to +10). This occurs when the result is not statistically significant. Having a negative number and a positive number in the CI means when comparing intervention A to intervention B, intervention A might be better than B, or B might be better than A. We could not conclude that a difference exists between the two interventions.

CALCULATING NNT AND NNH

NNT and NNH are easy to calculate:
- First determine the difference between the frequencies of the outcome of interest for two interventions.
- Then calculate the reciprocal of this difference.

For example, let's say drugs A and B are used to treat depression, and they result in 6-week response rates of 55% and 75%, respectively. The NNT to see a difference between drug B versus drug A in terms of responders at 6 weeks can be calculated as follows:
- Difference in response rates = 0.75 – 0.55 = 0.20
- NNT = 1 / 0.20 = 5.

In this example, you would need to treat 5 patients with drug B instead of drug A to see 1 extra responder. If the NNT had been 5.5, you would round up to the next whole number (6) because you can't treat a fraction of a person.

Interpreting the importance of NNT values is easy, too. The smaller the NNT, the larger the clinical difference between interventions; the larger the NNT, the smaller the difference.
- An NNT of 100 or more usually means little difference exists between interventions for the outcome of interest.
- An NNT of 2 would be hugely important and is rarely encountered.

Keep in mind, however, that some NNTs may be clinically important even though they are relatively large. An NNT of 500, for example, could be important if the outcome measured is death. Similarly, relatively small NNTs may be clinically irrelevant, such as an NNT of 5 when the outcome is a mild dry mouth.

Example. We can calculate the NNT (actually, NNH) for risk of new-onset diabetes mellitus attributable to second-generation antipsychotics (SGAs), using data from a study that compared diabetes rates in patients given SGAs versus conventional antipsychotics.[4] Differences in new-onset

outcome is seen once in every 5 patients being treated with one intervention versus another (an NNT of 5), the result will likely influence day-to-day practice. Together with calculating a confidence interval (*Box 2*),[3] the NNT can help you judge the *clinical* significance of a *statistically* significant result.

NNT is useful when examining differences in binary outcomes such as treatment response (yes/no), remission (yes/no), or avoidance of hospitalization (yes/no). NNT also is useful when we compare two medications' side effects. Under these circumstances, we call NNT the "number needed to harm" (NNH).

diabetes rates across ≤25 months were 2.03%, 0.80%, 0.63%, and 0.05% for clozapine, quetiapine, olanzapine, and risperidone, respectively, versus first-generation antipsychotics (FGAs).

The NNH for clozapine compared with FGAs is 1 / 0.0203 = 49. This means you would need to treat 49 patients with clozapine instead of an FGA for up to 25 months to encounter 1 extra case of new-onset diabetes mellitus. NNH calculations for quetiapine, olanzapine, and risperidone compared with FGAs would be 125, 159, and 2,000, respectively.

APPLYING NNT AND NNH TO CATIE

An ongoing controversy in schizophrenia treatment is the relative merit of using the more-expensive SGAs versus FGAs. The National Institute of Mental Health-funded CATIE study addressed this issue.[5-7]

In CATIE phase 1, which was double-blinded, 1,493 patients with schizophrenia were randomly assigned to 1 of 5 antipsychotics—perphenazine, olanzapine, quetiapine, risperidone, or ziprasidone—for up to 18 months. Patients who discontinued phase 1 before 18 months could participate in phase 2, where 543 patients were randomly assigned to 1 of 5 SGAs that they did not receive in phase 1. Those who prematurely discontinued phase 2 were offered open-label treatment with one or two antipsychotics. When they enrolled, patients were told these switches were possible.

Nearly one-half of all patients who enrolled finished 18 months of follow-up. What resulted, however, was a morass of percentages and p values that were misinterpreted by various parties—including *The New York Times*, which published an article headlined, "Little difference found in schizophrenia drugs."[8] We can apply NNT and NNH to the CATIE study results, however, and discover that:

- important differences do exist between the drugs tested

• these differences are clinically and statistically significant.[3]

Overall effectiveness in the CATIE trial was measured by determining how long patients remained on the medications to which they were randomly assigned. All-cause discontinuation—the primary outcome measure—included discontinuation because of:

• lack of efficacy
• poor tolerability
• patient decision.

Thus, both clinician and patient input—and both efficacy and tolerability—affected all-cause discontinuation. In CATIE phase 1, 74% of patients stopped participating (all-cause discontinuation) before 18 months. The percentage of patients who ended phase 1 early ranged from 64% for olanzapine to 82% for quetiapine. Thus, calculating NNT comparing olanzapine and quetiapine on this measure yields:

• NNT = 1 / (difference in discontinuation rates) = 1 / (0.82 - 0.64) = 1 / 0.18 = 5.6.

By convention, we round up to the next whole number, in this case 6. This means that for every 6 patients randomized to olanzapine treatment, 1 extra patient completed phase 1 on his or her initially initial medication, compared with patients randomized to quetiapine treatment.

Similarly, we can calculate the NNT for all-cause discontinuation for olanzapine compared with ziprasidone, perphenazine, and risperidone, and find NNT of 7, 9, and 11, respectively. In general, a single-digit NNT is sufficiently small for the result to be clinically relevant in day-to-day patient treatment.

In measuring the number of hospitalizations for exacerbation of schizophrenia symptoms per total person-year of exposure, NNT ranged from 3 to 7 in favor of olanzapine compared with the other antipsychotics. This means that for every 3 to 7 patients treated with olanzapine versus another antipsychotic, 1 hospitalization was avoided.

Tolerability. Calculating NNH can show how often you could expect specific tolerability outcomes when comparing medications. In CATIE, differences in tolerability emerged among the medications, and each antipsychotic had a unique profile of relative strengths and weaknesses that can be expressed in NNT and NNH. For example, in CATIE phase 1:

• For every 5 to 8 patients treated with olanzapine compared to other antipsychotics, 1 additional patient gained >7% in body weight (NNH is 5 to 8; not corrected for duration of exposure to the medication)
• For every 13 to 18 patients treated with olanzapine versus another antipsychotic, 1 additional patient discontinued because of weight gain or metabolic effects.

Data from phase 2 were largely consistent with those from phase 1, with important advantages noted for clozapine. NNT in favor of cloza-

> NNT and NNH are best calculated from the results of well-controlled clinical trials

Bottom Line

Calculating number needed to treat (NNT) and number needed to harm (NNH) can help you judge the clinical usefulness of a study's findings. A small NNT shows that the benefit is large, and a large NNH shows that the risk of harm is small. Also make sure the study included patients similar to those in your practice.

pine for all-cause discontinuation was 3, 4, and 7 compared with quetiapine, risperidone, and olanzapine, respectively. In phases 1 and 2, ziprasidone presented with the most favorable metabolic profile, whereas risperidone appeared to have the best overall tolerability.

POTENTIAL PITFALLS

Different studies can provide different estimates of outcomes such as response, remission, hospitalization, or adverse events. Two studies of the risk of new-onset diabetes with antipsychotics demonstrate that these differences can be difficult to interpret, particularly when populations and study designs differ.

- A Department of Veterans Affairs study of data on 56,849 patients[4] produced an NNH of 159 when olanzapine was compared with conventional antipsychotics, meaning 1 extra case of new-onset diabetes was encountered for every 159 patients treated with olanzapine compared to conventional antipsychotics.
- In the CATIE study,[5] examining new prescriptions of antidiabetic agents yields an NNH of 61 when olanzapine is compared with perphenazine, meaning that 1 extra case of a new prescription of an antidiabetic agent was encountered for every 61 patients treated with olanzapine versus perphenazine.

A statistically significant NNT or NNH should carry more weight than a result that is not statistically significant. Even so, make sure the study included patients similar to individuals in your practice before applying the results.

NNT and NNH are best calculated from well-controlled clinical trials. However, the underlying study design and potential biases may affect how NNT and NNH apply to clinical practice. A more complete discussion of the CATIE NNT and NNH secondary analysis can be found

elsewhere,[3] but issues to consider include the impact of differential switching[9] and the possible effects of dosages.[10]

References

1. Gray GE, Pinson LA. Evidence-based medicine and psychiatric practice. *Psychiatr Q* 2003;74(4):387-99.

2. Guyatt GH, Rennie D. Users' guides to the medical literature [editorial]. *JAMA* 1993;270(17):2096-7.

3. Citrome L, Stroup TS. Schizophrenia, Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) and number needed to treat: how can CATIE inform clinicians? *Int J Clin Pract* 2006;60(8):933-40.

4. Leslie DL, Rosenheck RA. Incidence of newly diagnosed diabetes attributable to atypical antipsychotic medications. *Am J Psychiatry* 2004;161(9):1709-11.

5. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005;353(12):1209-23.

6. McEvoy JP, Lieberman JA, Stroup TS, et al. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. *Am J Psychiatry* 2006;163(4):600-10.

7. Stroup TS, Lieberman JA, McEvoy JP, et al. Effectiveness of olanzapine, quetiapine, risperidone, and ziprasidone in patients with chronic schizophrenia following discontinuation of a previous atypical antipsychotic. *Am J Psychiatry* 2006;163(4):611-22.

8. Carey B. Little difference found in schizophrenia drugs. *The New York Times.* September 20, 2005.

9. Essock SM, Covell NH, Davis SM, et al. Effectiveness of switching antipsychotic medications. *Am J Psychiatry* 2006;163(12):2090-5.

10. Citrome L, Volavka J. Optimal dosing of atypical antipsychotics in adults: a review of the current evidence. *Harv Rev Psychiatry* 2002;10(5):280-91.