

The Numbers Needed to Treat and Harm (NNT, NNH) Statistics: What They Tell Us and What They Do Not

Chittaranjan Andrade, MD



Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

Research papers and research summaries frequently present information in the form of derived statistics such as the number needed to treat (NNT) and the number needed to harm (NNH). These statistics are not always correctly understood by the reader. This article explains what NNT and NNH mean; presents a simple, nontechnical explanation for the calculation of the NNT; addresses the interpretation of the NNT; considers applications of the NNT; and discusses the limitations of this statistic. The NNH is also briefly considered.

J Clin Psychiatry 2015;76(3):e330–e333
(doi:10.4088/JCP.15f09870).

© Copyright 2015 Physicians Postgraduate Press, Inc.

Clinical Question

A meta-analysis of randomized controlled trials (RCTs) found that antidepressants were effective in pediatric depression; the number needed to treat (NNT) was 9.¹ What information does NNT = 9 provide?

Introduction

Research results are presented in the form of summary statistics such as the mean improvement or the mean response rate in different treatment groups. These summary statistics can be directly compared, such as to determine whether the mean improvement or the mean response rate is significantly greater in one group versus the other.

Readers may wish to know whether an identified advantage is small or large. For example, in an RCT, an antidepressant may outperform placebo by 3 points on the Hamilton Depression Rating Scale; is this 3-point advantage for the antidepressant drug meaningful? Measures of effect size are available to answer such a question. These measures include statistics such as the standardized mean difference, relative risk, odds ratio, number needed to treat, and number needed to harm (NNH).^{2,3}

Understanding NNT

The NNT is a derived statistic. It is calculated from the observed response rates. It tells us how many patients need to be treated with a particular intervention for 1 extra patient to experience a favorable outcome such as treatment response. What does this mean?

We know that there are many reasons why patients respond to treatment even if they actually receive placebo⁴; therefore, with reference to the Clinical Question outlined at the start of this article, we know that if depressed children and adolescents were treated with an antidepressant, some of those who improved would probably have improved even if they hadn't received the antidepressant. The NNT provides us with an idea of the *unique* contribution of the antidepressant toward improving outcomes. Therefore, if the NNT for antidepressants is 9 in pediatric depression, it means that we need to treat 9 depressed pediatric subjects with an antidepressant for 1 *extra* patient to respond. Expressed otherwise, had these 9 subjects not received the antidepressant, then the number of responders would have been fewer by 1.

To explain the situation more fully, if 9 depressed pediatric subjects are treated with an antidepressant drug, 1 will respond specifically because of the antidepressant; others will respond because of placebo-related mechanisms, and the rest will not respond.

Note that NNT = 9 does *not* mean that antidepressant treatment will result in a response rate of 1 in 9. Antidepressant treatment, in fact, will result in several out of 9 patients responding, except that some of these responders would have responded anyway, through a placebo mechanism.

- The number needed to treat (NNT) is a derived statistic that tells us how many patients must receive a particular treatment for 1 additional patient to experience a favorable outcome such as treatment response.
- The number needed to harm (NNH) is a derived statistic that tells us how many patients must receive a particular treatment for 1 additional patient to experience a particular adverse outcome.
- Lower NNT and higher NNH values are associated with a more favorable treatment profile.
- The NNT and NNH statistics have limitations; therefore, clinicians should consult the actual response and adverse events rates to be better informed about likely treatment outcomes.

Understanding NNT: Breaking Up the Numbers

In the pediatric depression meta-analysis,¹ the response rate to antidepressants versus placebo was 60% vs 49%, respectively. So, if 100 depressed children and adolescents are treated with an antidepressant drug, 49% would respond because of placebo-related mechanisms, an additional 11% would respond because of the unique contribution of the antidepressant treatment (making the total response rate 60%), and the remaining 40% would not improve.

With regard to the $NNT=9$ estimate, if 9 depressed pediatric subjects are treated with an antidepressant drug, 49% of these (4½ patients) would anyway have responded because of placebo mechanisms, 11% (1 patient) would respond because of the additional benefit associated with antidepressant treatment, and 40% (3½ patients) would not respond. The NNT, of course, does not tell us all this; it merely tells us that we need to treat 9 patients with an antidepressant for 1 additional patient to respond. Readers may pardon the reference to “half patients”; mathematics does not always work in whole numbers!

Understanding NNH

The NNH, like the NNT, is also a derived statistic. It is calculated from the observed adverse effect rates, and it tells us how many patients need to be treated with a particular intervention for 1 extra patient to experience a specified adverse outcome. What does this mean?

March et al⁵ reported that the NNH for suicidality was 64 when sertraline was used to treat pediatric depression. This means that 64 depressed children and adolescents need to receive sertraline for 1 extra patient to experience suicidality as an adverse outcome. Or, if 64 such patients receive sertraline, some would experience suicidality as part of the illness, some would not experience suicidality, and 1 would become suicidal as a unique contribution of the drug.

NNTs and NNHs are calculated and interpreted in the same manner, and so only NNTs will be discussed for the most part in the remainder of this article.

Calculating the NNT

NNT is very simply calculated. Consider a hypothetical RCT in which 120 depressed patients were randomized to an experimental antidepressant drug ($n=58$) or placebo ($n=62$). After 8 weeks, 32 of the 58 antidepressant-treated patients met response criteria; the antidepressant response rate was 32/58, or 55%. At this same treatment endpoint, 26 of the 62 placebo-treated patients responded, yielding a placebo response rate of 26/62, or 42% (for simplicity, figures are rounded to the nearest integer).

In this example, antidepressant treatment raised the basal (placebo-related) response rate from 42% to 55%; that is, by 13%. This is like saying that if 55% of the patients responded to the antidepressant, 42% would have responded to placebo anyway, and so the unique contribution of the antidepressant was only 13%.

From this example, we conclude that antidepressant treatment of 100 depressed patients results in 13 extra responders. So, how many patients will need to be treated for there to be 1 extra responder? By a simple mental calculation, we determine that 7.7 patients (ie, 100/13) will need to be treated with that antidepressant for 1 extra patient to respond. Here, “extra” refers to treatment response over and above the placebo-related basal response rate of 42%. The NNT is usually rounded up to the nearest integer, and so the NNT in this worked example is 8.

In the Tsapakis et al pediatric depression meta-analysis,¹ the response rates for antidepressant vs placebo were 60% vs 49%, respectively. Antidepressant treatment resulted in 11 extra responders for every 100 subjects treated. The NNT is 100/11, or approximately 9.

Notes Related to the Application of the NNT

NNTs are calculated only when response rates are available. For example, it is not usual to report NNTs for Alzheimer’s disease treatments, because one does not expect the condition to respond to treatment in the usual sense of the term.

The NNT is not limited to response rates; it can be estimated for remission rates, as well. For example, in a meta-analysis of RCTs of aripiprazole versus placebo for acute mania, the NNT was 6 for response and 14 for remission.⁶ Likewise, NNTs can be estimated for relapse prevention during maintenance therapy, or even prevention of an adverse outcome such as death.

The NNT is not limited to comparisons with placebo; it can be estimated for comparisons with an active control, as well. For example, in a meta-analysis of RCTs comparing escitalopram and citalopram, Montgomery et al⁷ found that escitalopram was associated with better response ($NNT=12$) and remission ($NNT=6$) rates than citalopram.

The NNT can be accompanied by 95% confidence intervals (CIs). For example, Udina et al⁸ found that the NNT was 12 (95% CI, 7–38) for the efficacy of selective serotonin reuptake inhibitors in the prevention of occurrence of a major depressive episode during antiviral treatment of chronic hepatitis C infection. The interpretation of 95% CIs was discussed in an earlier article in this column.⁹

NNT and the Importance of Its Value

What is the importance of the value of the NNT? Obviously, the smaller the NNT, the greater the unique contribution of the drug toward the outcome. So, if the NNT for a drug is 4, it means that just 4 patients need to be treated with that drug for 1 additional patient to respond; in contrast, if the NNT is 18, it means that as many as 18 patients need to receive the drug for 1 additional patient to respond.

Note that the NNT cannot lie between 0 and 1; it is impossible, for example, for half a patient to be treated for 1 additional patient to respond. The lowest value for NNT (NNT = 1) is obtained in the impossibly ideal situation in which the response rates to drug and placebo are 100% and 0%, respectively. The highest possible value for NNT is infinity; this is when the response rate is the same in treatment and control groups.

Should we prefer drugs that have a lower NNT over those that have a higher NNT? Or should we reject a drug if it is associated with a high NNT? Not necessarily. Consider the following situations:

1. Antipsychotic A is associated with an NNT of 5 and Antipsychotic B with an NNT of 10. On the surface, this would seem to suggest that Antipsychotic A is better, and it may indeed be so. However, it is also possible that the 2 drugs were studied in different kinds of patients; for example, the patients in the RCTs of Antipsychotic B may have had greater comorbidity and may have been more refractory to treatment at baseline. This makes it inappropriate to compare NNTs. If comparisons between drugs are to be drawn, these should be done in head-to-head studies.
2. A treatment is associated with an NNT of 50–200 across a treatment period of 5 years. On the surface, it would seem ridiculous to treat as many as, say, 200 patients for 5 years for just 1 extra patient to obtain a treatment-related benefit. However, what if that benefit happened to be the prevention of a major vascular event? Where prevention of mortality or major morbidity is concerned, even large NNTs may be acceptable.^{10,11}

In passing, it may be noted that it is easier to demonstrate statistical significance when comparing active drug with placebo than when comparing active drug with an active control. This is why the NNT is higher in the latter situation. The larger the margin of separation between 2 treatments, the smaller the value of the NNT.

Comparing NNT and NNH Values

Comparisons are sometimes drawn between the NNT and the NNH for a specified drug. For example, in 8–10 week trials of levomilnacipran for major depressive disorder, the NNTs for response and remission were 9 and 14, respectively, and the NNHs for different adverse effects ranged from 10 to

31; the NNH for dropout due to adverse events was 19.¹² If the NNT is smaller than the NNH, does it mean that the risk-benefit ratio is favorable? Not necessarily. Clinicians need to make a subjective judgment about the value of the benefit and the seriousness of the risk. For example, a drug may have an NNT of 3 for pain relief and an NNH of 50 for a blood dyscrasia. Even though the blood dyscrasia is far less likely to occur than pain relief, the seriousness of the dyscrasia would certainly discourage the clinician from prescribing the drug.

Limitations of the NNT

The NNT is an academically useful statistic, but it has limited value for the practicing clinician. This section explains why.

If the NNT is, say, 9, we understand that 9 patients will need to receive the treatment for 1 extra patient to respond. We do not know how many patients will respond anyway because of placebo-related mechanisms, nor do we know how many patients will not respond at all. To obtain this information, we need to return to the data from which the NNT was calculated. In the earlier-discussed Tsapakis et al meta-analysis¹ in which the NNT was 9, 49% of patients showed a placebo response, an additional 11% responded to antidepressant medication, and 40% did not respond to treatment. Or, as already explained in the section Understanding NNT: Breaking Up the Numbers, 4½ out of 9 patients will show a placebo response, 1 additional patient will respond to the antidepressant drug, and 3½ patients will not respond to treatment.

Now, here is something interesting. Consider a situation in which drug versus placebo response rates are 12% versus 1%, respectively; the advantage for the drug is 11%, and the NNT is 9. Consider another situation in which the drug versus placebo response rates are 99% versus 88%, respectively; the NNT is again 9. These 2 situations are strikingly different. In the first situation, there is almost no placebo response, and medication is associated with a relatively large treatment gain. In the second situation, there is a large placebo response, and medication is associated with a relatively small treatment gain. Yet, the NNT is the same in the 2 situations. So, it is really important for clinicians to know not only what the unique contribution of the drug is (NNT) but also what the placebo response and nonresponse rates are.

Readers may also note that the NNT is a crude measure. This is because it is based on the response rate, which is also a crude measure. When response is defined, for example, as 50% attenuation of scores on a rating scale, then patients are classified as responders whether they improve by 50% or 100%, and they are classified as nonresponders whether they improve by 49% or 0%. Thus, a lot of information is lost when outcomes are dichotomized into response and nonresponse categories.^{13,14} It is far better to directly examine by what margin drug outperforms placebo on a rating scale than to see by what margin drug outperforms placebo on an arbitrary cutoff value that defines response on that rating scale.

With regard to adverse events, these may happen or not; in such situations, adverse event rates in drug and placebo groups and the NNH value are appropriate estimates. The occurrence of nausea as an adverse effect of serotonin reuptake inhibitors is a case in point. However, if outcomes can be quantified, then additional information can be clinically useful. For example, it could be helpful to know the NNH for the occurrence of akathisia with aripiprazole, but it could also be helpful to know by what margin akathisia is rated as more severe with aripiprazole as compared with placebo.

As a final limitation: the NNT and its CI do not convey any indication of statistical significance, and so there is little point in presenting an NNT and its 95% CI if a treatment effect is not statistically significant. Limitations of the NNT have also been discussed by Hutton.¹⁵

Parting Notes

1. Very surprisingly, almost nobody who cites an NNT value mentions the time frame for that value. Consider the statin NNT that was referred to earlier: 50 to 200 low-risk subjects must take a statin for 5 years for 1 additional subject to benefit.^{10,11} What if the time frame had been just 3 months, or if it had been 20 years? Obviously, a longer time frame is more discouraging.

2. When NNTs are cited for psychotropic agents, these apply to the duration of the RCTs that generated the data. Whereas RCTs for acute mania are usually around 3 weeks in duration, those for anxiety and depression last 6–8 weeks, and those for schizophrenia last 2 months or longer. Naturally, maintenance therapy RCTs last still longer. The reader needs to actually check the trial duration to conclude, for example, that 10 depressed patients need to take an antidepressant drug for 1 extra patient to benefit *across a 2-month treatment period*.

3. NNTs should be interpreted in the context of the definition of response (or remission). This definition may

not be the same in different studies even though the drug and disorder are the same.

4. Finally, it is theoretically possible for the NNT to be negative; that is, with a value of –1 and below. This happens if the response rate is lower with drug than with placebo.

REFERENCES

1. Tsapakis EM, Soldani F, Tondo L, et al. Efficacy of antidepressants in juvenile depression: meta-analysis. *Br J Psychiatry*. 2008;193(1):10–17.
2. Citrome L. Relative vs absolute measures of benefit and risk: what's the difference? *Acta Psychiatr Scand*. 2010;121(2):94–102.
3. Streiner DL, Norman GR. Mine is bigger than yours: measures of effect size in research. *Chest*. 2012;141(3):595–598.
4. Andrade C. There's more to placebo-related improvement than the placebo effect alone. *J Clin Psychiatry*. 2012;73(10):1322–1325.
5. March JS, Klee BJ, Kremer CM. Treatment benefit and the risk of suicidality in multicenter, randomized, controlled trials of sertraline in children and adolescents. *J Child Adolesc Psychopharmacol*. 2006;16(1–2):91–102.
6. Fountoulakis KN, Vieta E, Schmidt F. Aripiprazole monotherapy in the treatment of bipolar disorder: a meta-analysis. *J Affect Disord*. 2011;133(3):361–370.
7. Montgomery S, Hansen T, Kasper S. Efficacy of escitalopram compared to citalopram: a meta-analysis. *Int J Neuropsychopharmacol*. 2011;14(2):261–268.
8. Udina M, Hidalgo D, Navinés R, et al. Prophylactic antidepressant treatment of interferon-induced depression in chronic hepatitis C: a systematic review and meta-analysis. *J Clin Psychiatry*. 2014;75(10):e1113–e1121.
9. Andrade C. A primer on confidence intervals in psychopharmacology. *J Clin Psychiatry*. 2015;76(2): e228–e231.
10. Mihaylova B, Emberson J, Blackwell L, et al; Cholesterol Treatment Trialists' (CTT) Collaborators. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet*. 2012;380(9841):581–590.
11. Andrade C. Primary prevention of cardiovascular events in patients with major mental illness: a possible role for statins. *Bipolar Disord*. 2013;15(8):813–823.
12. Citrome L. Levomilnacipran for major depressive disorder: a systematic review of the efficacy and safety profile for this newly approved antidepressant—what is the number needed to treat, number needed to harm and likelihood to be helped or harmed? *Int J Clin Pract*. 2013;67(11):1089–1104.
13. Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry*. 2002;47(3):262–266.
14. Andrade C. Categorizing continuous variables. *Can J Psychiatry*. 2002;47(9):886.
15. Hutton JL. Misleading statistics. *Pharmaceutical Med*. 2010;24(3):145–149.