The “number needed to treat” turns 20 — and continues to be used and misused

Finlay A. McAlister MD MSc

In the 20 years since the initial description of the number needed to treat, this method of expressing the efficacy of an intervention has become widely used. Indeed, the Consolidated Standards of Reporting Trials statement recommends that the number needed to treat be reported in randomized trial publications, and journals of secondary publication (e.g., American College of Physicians Journal Club) routinely calculate and report the number needed to treat for studies of therapy. As well, there have been increasing calls for health care policy makers to use numbers needed to treat to inform their recommendations, and league tables comparing numbers needed to treat have appeared in the literature and on the internet (See www.cebm.utoronto.ca/glossary/nnts.html#table and www.jr2.ox.ac.uk/bandolier/band50/b50-8.html for examples from different branches of medicine).

Having attended hundreds of journal clubs as well as departmental and divisional rounds over the past 2 decades, I am consistently impressed by the frequency with which audience members display skepticism about a therapy if its efficacy is presented only in relative terms such as odds ratios or relative risk reductions. Not infrequently, this skepticism is healthy — the dangers of misinterpreting the importance of a therapy when relying solely on relative effect estimates are well known. However, I have also been struck by the extent to which discussions of a therapy’s number needed to treat, and even comparisons between therapies on this basis, are accepted at face value. A review of the literature and their experiences in journal club and critical appraisal settings led Chong and colleagues to also express concern that many clinicians appear to hold “the impression that NNT [number needed to treat] values in and of themselves are broadly comparable” and display “an implicit belief that an unadjusted NNT value adequately captures the overall worth of a treatment.”

In this article, I explore the factors (beyond the efficacy of a therapy) that influence the number needed to treat and that must be taken into account when comparing these values between therapies.

What is the number needed to treat?

The number needed to treat is an aggregate measure of clinical benefit that represents the number of patients who would need to be treated to prevent 1 additional adverse event. It is calculated by taking the reciprocal of the absolute risk reduction between 2 treatment options. This number is a useful way to summarize the potential impact of a therapy when discussing treatment options with patients. A detailed discussion of how to personalize this number to each patient’s situation, including means to incorporate potential harms as well as patient values and preferences, has been published.

Given the many heuristics that guide medical decision-making, it is not surprising that the number needed to treat has also been embraced by those wishing to compare 2 or more therapies. Proponents use it as though it offers a single dimensionless metric. Although the number needed to treat may appear to be an absolute measure of clinical benefit, it is in fact specific to a single comparison in a single study because it is the reciprocal of the difference in event rates between 2 treatment options. Thus, this number should not be considered in isolation. It should be viewed as specific to a particular comparison, not to a particular therapy. In addition, there are 3 other factors that can influence the number needed to treat for any therapy, above and beyond the efficacy of the therapy and the comparator.

Nontherapy factors that affect the number needed to treat

Baseline risk

Although the relative efficacy of drug therapies is often similar across patient subgroups at different risk, the number needed to treat varies inversely with baseline risk. As a re-

Dr. McAlister is with the Division of General Internal Medicine, University of Alberta, Edmonton, Alta.
sult, the number needed to treat for any therapy rarely appears favourable if it is evaluated in low-risk populations. For example, although the mortality relative risk reduction with antihypertensive therapy is similar across risk strata (about 9%–12%), the number needed to treat for antihypertensive therapy to prevent 1 death over 5 years ranges from 1157 in healthy young women to 17 in older men with other cardiovascular risk factors.12

By the same token, the number needed to treat will be larger if cointerventions reduce the frequency of the outcome. For example, consider the case of a new therapy being tested for a condition for which several efficacious therapies have already become standard therapy or for which secular changes have improved the baseline risk of adverse outcomes. Because contemporary trial participants will have a lower baseline risk than those enrolled in the earlier trials, the novel therapy will always exhibit a larger number needed to treat than the earlier therapy. In early trials involving patients with myocardial infarctions, the number needed to treat for Aspirin was 42 to prevent 1 in-hospital death.13 However, if Aspirin were being tested today as a novel therapy in such patients, β-blockers, angiotensin-converting enzyme inhibitors and statins would be mandated for virtually all trial patients receiving Aspirin or placebo and the baseline mortality rate would be about 3.5%.14 As a result, the number needed to treat would at best be 124 (assuming the relative risk reduction was the same 23% as found in the earlier trials). Put another way, consider the potential impact of statins for primary prevention given the secular trends in coronary mortality seen over the past 2 decades: if statins were available in 1975, the number needed to treat to prevent 1 coronary death would have been 83 for men and 286 for women. However, in 1995, this would have been 154 for men and 1075 for women.15

**Time frame**
The number needed to treat is inherently a time-dependent measure — it depends on when the outcomes are counted. Even if the relative risk reduction from a long-term therapy is constant over time, the number needed to treat will decrease with increasing follow-up as events accrue and the absolute event rate increases. However, it is unlikely to be linear as time passes, given the increasing contribution from competing risks (i.e., other adverse outcomes not usually affected by the therapies being compared) and the use of concurrent medications that may also impact the events of interest. Of course, the relative risk reduction of a therapy is not always constant over time. In particular, surgical therapies usually involve a trade-off of early excess risk for long-term benefits.16 However, even some drug therapies demonstrate differential effects over time. In the Anglo-Scandinavian Cardiac Outcomes Trial–Lipid-Lowering Arm (ASCOT-LLA),17 statin therapy was associated with relative reductions in coronary events (reported as hazard ratios) of 0.33 (95% confidence interval [CI] 0.14–0.78) at 90 days but only 0.64 (95% CI 0.50–0.83) at the end of the study (3.3 years). As a result, the number needed to treat to prevent a coronary event ranged from 364 patients (95% CI 210–1362) treated for 90 days to 93 patients (95% CI 59–208) treated for 3.3 years (Table 1).

Clearly, one cannot assume that a number needed to treat of 30 people over 5 years can be converted to a number needed to treat of 150 over 1 year or to a number needed to treat of 15 over 10 years. However, people often make the mistake of trying to intrapolate or extrapolate from one time period to another to standardize comparisons between interventions. Various methods to calculate the number needed to treat for different time frames within the same study have been proposed. They include using the survival curves to estimate annual event rates and multiplying hazard ratios with the survival rates in the control group at the times of interest.18 Although these methods work well if trial data with long-term follow-up exists, they do not help in the situation of chronic preventive therapies. In such situations, we wish to project the benefits demonstrated in randomized trials lasting 1–3 years out to treatment horizons lasting for several decades. However, this form of extrapolation is fraught with potential error.

In fact, the number needed to treat concept is best applied to acute conditions with outcomes that cluster closely in time without long-term sequelae (e.g., the treatment of gastrointestinal bleeds or ventricular arrhythmias). It applies less well to chronic conditions (e.g., hypertension, osteoporosis or atherosclerosis) for which adverse outcomes are not permanently avoided but are merely postponed. Indeed, an outcome for a patient with a chronic condition is only truly avoided if it is postponed longer than the patient’s remaining lifespan.19,20 Thus, it has been suggested that it may be more accurate to describe the potential impact of chronic preventive therapies

---

**Table 1: Example calculations for a trial evaluating the impact on fatal and nonfatal myocardial infarctions of statins for patients with hypertension and other cardiovascular risk factors.**17

<table>
<thead>
<tr>
<th>Time</th>
<th>Event rate in placebo arm</th>
<th>Event rate in intervention arm</th>
<th>Relative risk reduction* (95% CI)</th>
<th>Absolute risk reduction† (95% CI)</th>
<th>Number needed to treat‡ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 d</td>
<td>21/5121</td>
<td>7/5184</td>
<td>0.67 (0.23–0.86)</td>
<td>0.28 (0.07–0.48)</td>
<td>364 (210–1362)</td>
</tr>
<tr>
<td>12 mo</td>
<td>61/5121</td>
<td>34/5184</td>
<td>0.45 (0.16–0.64)</td>
<td>0.54 (0.17–0.92)</td>
<td>186 (109–601)</td>
</tr>
<tr>
<td>3.3 yr</td>
<td>(study end)</td>
<td>154/5121</td>
<td>0.36 (0.18–0.50)</td>
<td>1.08 (0.48–1.69)</td>
<td>93 (59–208)</td>
</tr>
</tbody>
</table>

Note: CI = confidence interval.

*Relative risk reduction = 1 – (event rate in intervention arm/event rate in placebo arm). For example, at 90 days, relative risk reduction = 1 – (7/5184)/(21/5121).
†Absolute risk reduction = event rate in placebo arm – event rate in intervention arm. For example, at 90 days, absolute risk reduction = 21/5121 – 7/5184.
‡Number needed to treat = 100/absolute risk reduction. It can also be calculated as number needed to treat = (1/absolute risk reduction) ×100. For example, at 90 days, number needed to treat = 1090.3.
in terms of average durations of life gained rather than focusing on differential survival at one point in the survival curve.\textsuperscript{19,20} For instance, rather than describing the benefits of angiotensin-converting enzyme inhibition among patients with heart failure enrolled in the Studies of Left Ventricular Dysfunction (SOLVD) trial as a number needed to treat of 22 to prevent 1 death over 3.5 years, it would perhaps be more informative to describe a potential gain of 1.9 months in life expectancy with 3.5 years of angiotensin-converting enzyme inhibitor treatment.\textsuperscript{11,21} However, a randomized trial comparing these 2 presentation formats proved that lay people had an easier time understanding the number needed to treat format. They were more likely to accept therapy when the benefits were expressed as the number needed to treat compared with the average duration of life gained.\textsuperscript{21}

Outcomes

Most therapies impact more than 1 outcome. This means that more than 1 number needed to treat that needs to be incorporated into treatment decision-making. In doing so, it takes substantial clinical experience to move beyond simple expressions of frequency. Instead, one should weigh the severity and importance of these different numbers needed to treat and integrate patient preferences. Methods to integrate efficacy and safety data from trials with each patient’s risks and values have been discussed elsewhere.\textsuperscript{9}

How important are these 3 factors?

This question is perhaps best answered by considering the magnitude of changes in the number needed to treat that arise if these numbers are standardized to a model population at a predefined risk with a common outcome and study duration. For example, Caro and colleagues\textsuperscript{24} standardized the numbers needed to treat from 18 cardiovascular trials by inputting the relative efficacy data for each therapy into simulated population models in which baseline risk, treatment duration, outcomes and comparators were standardized. They found marked changes in the numbers needed to treat after standardization ranging from a 91% decrease to a 223% increase compared to the crude numbers needed to treat reported in each trial.\textsuperscript{22} Importantly, the authors reported that there were no factors that could predict which crude numbers needed to treat were most likely to change markedly after standardization. This emphasizes the importance of only comparing numbers needed to treat for therapies if they have been derived against similar comparators, for the same outcome, in populations at the same stage of disease, and followed for the same duration. To the extent that any of these conditions are not met, a face-value comparison between numbers needed to treat can be misleading.

Are there other limitations?

The number needed to treat can only be expressed for binary outcomes (such as death v. survival, hospitalization v. not). It cannot be calculated for continuous outcomes, which are still relevant to patients and are frequently the intermediate targets of our therapeutic endeavours, such as changes in blood pressure, bone density or serum lipid levels. Thus, the number needed to treat may not be the best metric with which to compare chronic preventive therapies. For example, although the number needed to treat to prevent a recurrent vertebral fracture with bisphosphonate therapy was 15 in the Fracture Intervention Trial (FIT), this does not mean that 14 of every 15 patients didn’t benefit; 89% of women given bisphosphonate demonstrated at least some improvement in the bone mineral density in their lumbar spine over the first 12 months of treatment.\textsuperscript{25,26}

Although confidence intervals can be generated around the number needed to treat (by calculating the reciprocals of the confidence interval for the absolute risk reduction), it is common to see it reported as a single number, especially if the result is not statistically significant. This situation has arisen because of the difficulty in describing the confidence interval around a nonsignificant number needed to treat because it stretches between 2 values via infinity (since the reciprocal of a statistically nonsignificant absolute risk reduction, which incorporates 0, must encompass infinity).\textsuperscript{27} For example, a therapy exhibiting an absolute risk reduction of 10% with a 95% confidence interval of –5% to 25% would be expressed as an absolute risk reduction of 5% to 25% extending from a number needed to treat of 10 with a 95% confidence interval around the absolute risk reduction or the relative risk reduction are much easier to express and understand.

The number needed to treat is an expression of the frequency of an outcome event, not its utility. Patients and physicians vary their treatment decisions depending on cost, side-effect profile, ease of application, the severity of the outcome it is supposed to prevent, and personal values and preferences. Thus, a number needed to treat alone is insufficient to declare a therapy worthy of use.\textsuperscript{28–30} For example, a number needed to treat of 100 may be acceptable for a drug that is cheap, easy to take and has few side effects; however, a number needed to treat of 5 may be too high for an expensive drug that carries substantial potential toxicities.

What other factors should be considered when interpreting trial-based estimates?

Although the relative effects of many therapies are often the same in routine clinical practice as in trials (if given to comparable patients),\textsuperscript{11,31,32} the number needed to treat is rarely so. Generalizing a number needed to treat from a particular trial to routine care in a different setting may lead to erroneous conclusions. For example, because trial participants tend to be younger, healthier and have better prognoses (i.e., lower absolute risk) than nontrial participants,\textsuperscript{33} a trial-based number needed to treat may overestimate the number needed to treat for that therapy when used in clinical practice where baseline risks are higher.

On the other hand, trial-based numbers needed to treat may sometimes underestimate the numbers needed to treat in clinical practice. For example, patients in routine practice are likely to have more comorbidities than trial participants, such
that their risk of competing causes of death increases, thereby minimizing the potential benefits of a therapy targeting a specific mode of death (e.g., implantable cardioverter defibrillators for ventricular arrhythmias). Moreover, as patients in routine clinical practice are unlikely to take a medication under the same conditions as participants in a clinical trial (i.e., the same dose, the same intensity of monitoring by clinicians as experienced as those who participated in the trials, the same high levels of adherence, and the same low use of interventions as mandated in trial protocols), trial-based numbers needed to treat are likely to underestimate the numbers needed to treat in clinical practice.

What is the role of the number needed to treat?

Some people argue that advances in pharmacogenomics and proteomics will render the number needed to treat obsolete when we are able to personalize treatment recommendations by taking into account each patient’s biology. However, this situation is likely far in the future (and given the well-recognized pitfalls of multiple subgroup analyses,9,31 the future is perhaps not quite as rosy as it may first appear). Thus, there is clearly still a need to express the potential impacts of a therapy when discussing options with patients. The presentation of evidence about any therapy should incorporate absolute risk as well as relative benefits (and harms), and the number needed to treat remains a useful means of doing so. Methods of incorporating patient values and preferences into these discussions are familiar to experienced clinicians and formal methods of calculating numerical values such as the likelihood of being helped or harmed have been described in full elsewhere.9,36

It is well recognized that how efficacy and safety data is presented to physicians, patients and health care policy makers influences their decisions. In fact, most studies suggest that all 3 groups make more conservative decisions about therapies when they are presented with numbers needed to treat than when they are presented with the same data presented as relative risk ratio.37–41 However, it is not entirely clear that a more conservative decision is necessarily the right one. For example, many British patients with atrial fibrillation who were likely to benefit from anticoagulant therapy because of their risk profiles and their similarity to the participants in randomized trials supporting the efficacy of warfarin declined warfarin therapy when presented with the data about their absolute risks and benefits.42

This raises questions about how easy the number needed to treat is to understand, particularly since this measure is not familiar to lay people. For instance, department stores advertise sale prices in terms of relative risk reductions (e.g., “save 20% off the regular price”) or absolute risk reduction (e.g., “save $5 off the regular price”) rather than numbers needed to treat (Have any readers seen signs trumpeting “the number of discounted items that would have to be purchased to get 1 item free is X”?). Indeed, surveys of patients and medical students demonstrated that few (7% and 25% respectively) correctly interpreted the number needed to treat and could identify which treatment was most likely to be beneficial if efficacy was expressed in this way.43,44 A survey of the general public in Norway reported that although 93% of respondents consented to therapy when presented with the number needed to treat, only 35% reported that the number needed to treat was “very easy” or “somewhat easy” to understand.23

The consistent finding in multiple studies that patients and physicians are insensitive to the magnitude of the number needed to treat when making treatment decisions also supports the contention that this format of expressing treatment efficacy is not as easily understood as formats such as relative or absolute risk reduction.26,36,45 It has been suggested that use of the “likelihood of being helped or harmed” modification of the number needed to treat (in which the ratio between the number needed to treat and the number needed to harm is adjusted for individual patient utilities and preferences) will promote understanding.36 However, this hypothesis that has not yet been adequately tested. Thus, for now, the most prudent course of action to maximize understanding is to continue to present information on the potential benefits from a therapy in terms of absolute risk reduction, relative risk reduction and number needed to treat. Of course, such a discussion should also include information on potential adverse effects as well as costs and inconvenience.

Final thoughts on the number needed to treat

The point of this article is not to dissuade readers from using the number needed to treat when discussing treatment options with their patients. This measure remains a valuable tool in our efforts to improve communication with patients at the bedside and in the office. However, I hope to remind readers that numbers needed to treat for different therapies can only be compared if the therapies were tested in similar populations with the same condition (and preferably at the same stage of disease), for their effects on the same outcomes, against the same comparator and over the same time frame. To the extent that these conditions are not met, comparisons based on numbers needed to treat may be misleading. The words of the editors of Bandolier (in a footnote which frequently accompanies number needed to treat league tables provided on their evidence-based medicine website) are worth remembering: numbers needed to treat “are reproduced here with a simple health warning — that readers should always go back to original papers to get all the nuances of the original studies.”36
REFERENCES

2. Moher D, SchulzKF, Altman D. The CONSORT statement: revised recommenda-
3. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommenda-
4. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
5. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
6. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
7. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
8. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
9. Kumana CR, Cheung BMY, Lauder IJ. Gauging the impact of statins using num-
10. Alter DA, Manuel DG, Gunraj N, et al. Age, risk-benefit trade-offs, and the pro-
11. McAlister FA. Relative treatment effects are consistent across a spectrum of under-
on cardiovascular outcomes in women and men. A meta-analysis of individual pa-
13. Antithplatelet Trials’ Collaboration. Collaborative overview of randomised trials of
antithplatelet therapy. Prevention of death, myocardial infarction, and stroke by pro-
14. Fox KA, Steg PG, Eagle KA, et al. Effect of ACE inhibitor trandolapril on life expectancy of patients with reduced left-
Group. Users’ guides to the medical literature: XX. Integrating research evidence
16. Halvorsen PA, Selmer R, Kristiansen IS. Different ways to describe the benefits of
enolapril on survival in patients with reduced left-
17. Christensen PM, Kristiansen IS. Number needed to treat (NNT) — Needs treat-
dering notions of number need to treat and other common risk reduction formats.
18. Otterstad JE, Sleight P. The HOPE study: comparison with other trials of second-
19. Tan LB, Murphy R. Shifts in mortality curves: Saving or extending lives?
Arch Intern Med 2005;165:848-56.
21. SOLVD Investigators. Effect of enalapril on survival in patients with reduced left-
Effect of ACE inhibitor trandolapril on life expectancy of patients with reduced left-
23. Halvorsen PA, Selmer R, Kristiansen IS. Different ways to describe the benefits of
enolapril on survival in patients with reduced left-
an NNT model.
25. Christensen PM, Kristiansen IS. Number needed to treat (NNT) — Needs treat-
dering notions of number need to treat and other common risk reduction formats.
26. Halvorsen PA, Selmer R, Kristiansen IS. Different ways to describe the benefits of
enolapril on survival in patients with reduced left-
27. McAlister FA, O'Connor AM, Wells G, et al. Number Needed to treat (NNT) —
implication in rheumatology clinical practice.
implication in rheumatology clinical practice.
implication in rheumatology clinical practice.
treat? The different perspectives of Canadian patients and family physicians.
31. Vist GE, Hagen KB, Devereaux PJ, et al. Subgroup analysis and other (mis)uses of
32. Straus SE. Individualizing treatment decisions. The likelihood of being helped or
33. McAlister FA, O’Connor AM, Wells G, et al. When should hypertension be
34. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of
35. McAlister FA, O’Connor AM, Wells G, et al. When should hypertension be
36. Straus SE. Individualizing treatment decisions. The likelihood of being helped or
37. Naylor CD, Chen E, Strauss B. Measured enthusiasm: Does the method of reporting
38. Fenow L, Taylor WC, Arnold RM. Absolutely relative: how research results are
39. Hus JE, Naylor CD. Communicating the benefits of chronic preventative therapy;
Does the format of efficacy data determine patients’ acceptance of treatment? Med
40. Crammery W, Walley T. Same information, different decisions: the influence of ev-
idence on the management of hypertension in the elderly. Br J Gen Pract 1996;
46:661-3.
41. Bacher HC, Weinbacher M, Gry K. Influence of method of reporting study results
on decision of physicians to prescribe drugs to lower cholesterol concentration.
42. Protheroe J, Fahey T, Montgomery AA, et al. The impact of patients’ preferences
in the treatment of atrial fibrillation: observational study of patient-based decision
43. Sheridan SL, Pignone M. Numeracy and the medical student’s ability to interpret
44. Sheridan SL, Pignone MP, Lewis CL. A randomized comparison of patients’ un-
derstanding of number needed to treat and other common risk reduction formats.
45. Kristiansen IS, Gyrd-Hansen D, Nexoe J, et al. Number needed to treat: Easily un-
derstood and intuitively meaningful? Theoretical considerations and a randomized

Correspondence to: Dr. Finlay McAlister, 2E3.24 WMC, University of Alberta Hospital, 8440 112 St., Edmonton AB T6G 2R7; fax: 780 407-2680; Finlay.McAlister@ualberta.ca