NUMBER NEEDED TO TREAT:

TIME-RELATED ESTIMATION AND INTERPRETATION

Daniel Suissa,[1] MD, MSc, Brielan Smiechowski,[2,3] BSc, Samy Suissa, PhD.[2,3]

[1] Hôpital Notre-Dame, Centre hospitalier de l'Université de Montréal;

[2] Centre for Clinical Epidemiology, Lady Davis Institute - Jewish General Hospital;

[3] Department of Epidemiology and Biostatistics, McGill University;

Montreal, Quebec, Canada,

Funding: This research was funded by grants from the Canadian Institutes of Health Research (CIHR).

Please address correspondence and request for reprints to:

Dr. Samy Suissa

Centre for Clinical Epidemiology, Jewish General Hospital

3755 Cote Ste-Catherine, H-461

Montreal, Québec, Canada H3T 1E2

Tel: 514-340-7593 Fax: 514-340-7564

E-mail: samy.suissa@mcgill.ca

Words in text: 2,920 **Words in abstract: 200** **Tables: 1**

March 6, 2010

Abstract

The number needed to treat (NNT) is a simple measure of a treatment's impact, increasingly reported in randomized trials and observational studies. Its calculation in studies involving varying follow-up times or recurrent outcomes has been criticized. We discuss the NNT in these contexts, illustrating using several published studies.

The computation of the NNT is founded on the cumulative incidence of the outcome. Instead, several published studies use simple proportions that do not account for varying follow-up times, or use incidence rates per person-time. We show that these approaches can lead to erroneous values of the NNT and misleading interpretations. For example, after converting the incidence rate to a cumulative incidence, we show that a trial reporting a NNT of 4 "to prevent one exacerbation in 1 year" should have reported a NNT of 9. A survey of papers reporting NNT published in four major medical journals in 2009, found that 6 out of all 10 papers involving varying follow-up times did not correctly estimate the NNT.

As the "number needed to treat" becomes increasingly used in complex studies and in the comparative effectiveness of therapies, its accurate estimation and interpretation become crucial to avoid erroneous clinical and public health decisions.

Key words: Biostatistics; Epidemiology; Impact measure; Methods; Observational studies; Randomized controlled trials;

Introduction

The number needed to treat (NNT) is now used extensively in randomized trials and observational studies to provide an additional and user-friendly measure of the impact of a drug or treatment on a given disease outcome.[1,2] Its interpretation is appealing: it represents the number of patients with the disease under study that need to be treated with the drug or intervention to prevent the disease outcome in one patient. Its computation and interpretation have traditionally been based on a single occurrence of a dichotomous outcome assessed using the ideal randomized trial, namely a trial with equal follow-up for all patients. However, such ideal trials are rare. In practice, studies will involve unequal follow-up times, while certain trials will also study recurrent outcomes with multiple events, such as infections and exacerbations. The NNT has nonetheless been applied to these situations. Recently, its application to outcomes with repeated events has been criticized and its potential misuse with varying follow-up times highlighted, raising some questions about its validity and interpretation.[3,4]

To clarify these points, we expand on some issues raised in a recent letter,[4] by examining the use of the NNT measure in a number of trials published in four major journals, including the *New England Journal of Medicine,* the *Journal of the American Medical Association, The Lancet* and the *British Medical Journal*. In the context of these trials, all with varying follow-up times or outcomes with multiple events, we present the proper approaches to estimate, analyze and interpret the corresponding NNT.

The NNT measure

To better understand the origin of the NNT, recall that the frequency of disease outcome is typically measured as a cumulative incidence of the outcome per number of patients followed over a given time period. This will result in a proportion, whose value will be written for example as 0.5/100, 2/100 or 5/100, where fixing the denominator at 100 makes the magnitudes of outcome frequencies easily comparable. Alternatively, the motivation behind the NNT measure is to first make the cumulative incidences easily comparable by fixing the numerator at 1. In our example these become 1/200, 1/50 and 1/20. When these rates are inverted to 200:1, 50:1 and 20:1, they then represent the number of patients that need to be observed to find one patient with the outcome.

The use of the NNT relates more specifically to the effectiveness of a treatment that can be measured by the difference in the cumulative incidence of the outcome (CI) over a fixed follow-up time period, between two groups of patients, namely CI_0-CI_1, where 0 represents the reference treatment or placebo and 1 the treatment under study. This difference will represent the proportion of patients for whom the outcome was prevented. Inverting this difference will produce the number of patients that need to be treated to prevent one patient with the outcome. This patient-based interpretation corresponds to the traditional NNT.[1]

For example, in a trial with equal one-year follow-up for all patients, a difference in one-year cumulative incidence of the outcome of 9/100 – 5/100 = 4/100, represents a prevented cumulative incidence of 4/100 (preventing 4 patients with the outcome for every 100 patients treated) over one year. The NNT is then simply $1/(CI_0\text{-}CI_1) = 100/4 = 25$.

This inversion suggests that rather than using the 4/100 measure of prevented cumulative incidence, the NNT value of 25 will provide the more graphic interpretation of one patient avoiding the outcome over one year for every 25 patients treated.

Many studies, however, involve more complex study designs with varying follow-up times or outcomes with multiple events. In these situation, the calculation of the NNT requires some more careful consideration.

Patient-based NNT: varying follow-up

In practice, trials generally result in varying follow-up times between patients in which case the cumulative incidence of an outcome cannot be calculated simply as a proportion of subjects. It must instead use the Kaplan-Meier approach that accounts for varying follow-up times and provides a curve for the cumulative incidence over time.[5] The NNT can then be directly computed by inverting the difference in the cumulative incidence of the outcome between the two groups at the desired time of follow-up.[6] This NNT will represent the number of patients that need to be treated to prevent one patient with the outcome over the given effect time period, namely the traditional NNT interpretation. This approach has been properly used in several recent trials.[7-12]

Other trials, however, while at times also presenting Kaplan-Meier curves, incorrectly computed the NNT using the simple proportion of patients with the outcome. In some cases, this incorrect computation does not matter much because of the short and mostly complete follow-up.[13,14] However, when the follow-up times are more variable, computing the cumulative incidence without accounting for this may have resulted in a

distorted value of NNT.[15-18] For example, in the recent trial of the effect of adding zoledronic acid to a combination therapy in premenopausal women with endocrine-responsive early breast cancer, 1803 patients were followed for varying times (median 47.8 months).[18] Despite the facts that the Kaplan-Meier curves were estimated and that follow-up varied extensively between patients, the authors used the simple proportion of patients whose disease progressed to compute the NNT as $1/((83/904) - (54/899)) = 31$. Another example is a trial of nucleoside analogues against herpes simplex virus HSV type 2 (HSV-2), 1484 subjects were randomly assigned valacyclovir or placebo for 240 days.[15] The authors claimed that "one would expect to treat 38 persons with recurrent genital herpes for a year to prevent one case of HSV-2 infection". Here again, the authors used the simple proportion of patients with HSV-2 infection over the 240-day follow-up, and also extrapolated it to 365 days. Indeed, the value of 38 for the NNT comes from $(240/365) \times 1 / ((27/741) - (14/743))$. This is inaccurate not only because the authors used proportions that do not account for varying times, but also because this value was extrapolated from a trial period of 240 days to an unstudied one-year period.

Patient-time-based NNT

Rather than using the absolute number of patients with the outcome in computing the NNT in studies with varying follow-up times, as done above, trials may use, as a measure of outcome frequency, the incidence rate (IR) of the outcome. It is computed as the number of patients with the outcome divided by the total amount of person-time or patient-time, to account for varying follow-up times. In this case, the effect of a treatment

is measured by the difference in the incidence rate of the outcome between two groups, namely IR_0-IR_1, where 0 represents the reference treatment or placebo and 1 the treatment under study. This difference will represent the incidence rate of prevented events per person-moment. Some authors have computed the corresponding NNT as $1/(IR_0-IR_1)$ in several recent studies, with the interpretation that it represents the number of patients that need to be treated to prevent one outcome over a given time period.[19-25]

An example is from the trial of 1,801 frail elderly adults randomized to a hip protector or to a control group to assess the effect on the risk of hip fracture.[19] The subjects were followed for a mean of 1.1 years with follow-up times varying up to 2 years. The incidence rate of hip fracture in the hip-protector group was 21.3 per 1000 person-years compared with 46.0 per 1000 patient-years for the control group. The NNT for 1-year treatment was computed as 1/(0.0460-0.0213) = 40.5 which was interpreted as "the number needed to treat for one year to prevent one hip fracture was 41 persons". Moreover, the incidence rates were converted to five-year rates, namely to 106.5 versus 230.0 per 1000, and the corresponding NNT computed as 1/(0.2300-0.1065) = 8.1, which was interpreted as "the number needed to treat for five years to prevent one hip fracture was 8 persons".[19]

This interpretation is not accurate. Since this NNT is computed from the incidence rate of events, its inverse does not represent persons but rather person-time. This person-time-based NTT should thus instead be expressed as the number of person-moments needed to be treated. Indeed, 8 patients being treated for 5 years are not necessarily the same as 40 patient-years. This distinction is important because the former insinuates 8

distinct patients treated for 5 years, while the latter can as well equally imply 40 patients treated for 1 year or 20 patients treated for 2 years. Indeed, the incidence rates inherently assume that person-moments are intrinsically interchangeable so that one year of follow-up in one patient is equivalent to the first month of follow-up of 12 patients. The study in fact provided the Kaplan-Meier curves for the cumulative incidence of hip fracture, which indicate that the one-year cumulative incidence of hip fracture is around 5.0% for the hip-protector group and 2.1% for the control group. These correspond to a now accurate NNT of 35 patients needing to be treated for 1 year to prevent 1 hip fracture, rather than the reported 41. Similarly, in the Collaborative Atorvastatin Diabetes Study (CARDS), the authors used incidence rates to report that "27 patients would need to be treated for 4 years to prevent one (major cardiovascular) event", but the Kaplan-Meier curves for the cumulative incidence of a major cardiovascular event result in a NNT value closer to 20 patients at four years.[23]

Event-based NNT: multiple events

In some studies, the outcome involves a recurrent event that can occur more than once during the patient's follow-up.[26,27] In such studies, the frequency of disease outcome is also measured as an incidence rate computed as the total number of outcome events divided by the total amount of person-time. Here again, the corresponding NNT has been computed as $1/(IR_0\text{-}IR_1)$ with the interpretation that it represents the number of patients that need to be treated for a given time period to prevent one "event" in that period. An example is the TORCH trial of over 6,000 COPD patients randomized to one of

four treatment groups including fluticasone, salmeterol, their combination, or placebo, followed for three years.[26] With varying follow-up times, the incidence rate of exacerbation, which could occur more than once during a patient's follow-up, was 1.13 per patient-year in the placebo group and 0.85 in the combination group. The authors thus computed the NNT as 1/(1.13-0.85) = 3.6 rounded to 4, which was interpreted as "number needed to treat of four to prevent one exacerbation in 1 year".

This type of interpretation has been criticized on the basis that this event-based NNT deviates from its original meaning of the number needed to prevent one patient with an exacerbation but instead targets the number needed to prevent one exacerbation.[3] This particular criticism is not entirely justified because this event-based NNT is computed from the incidence rate of events so that it represents events and not patients with events. Instead, the criticism of the interpretation of this event-based version of the NNT should be directed to its claim about the time period: This NTT does not represent patients treated for a time period, but rather person-moments of treatment.

For example, the TORCH study's "number needed to treat of four to prevent one exacerbation in 1 year" would be more accurately expressed as "the number needed to treat of four patient-years to prevent one exacerbation".[26] This distinction is important because the former insinuates that 4 patients need to be treated for 1 year, while the latter can equally be expressed as an NNT of 48 patient-months to prevent one exacerbation or 208 patient-weeks to prevent one exacerbation, because of the nature of the underlying incidence rate. Indeed, with this measure, person-moments are intrinsically assumed to be interchangeable so that one year of follow-up in one patient is equivalent to the first month

of follow-up of 12 different patients. Thus, the interpretation must avoid the insinuation of a 12-month treatment period by clearly specifying 12 patient-months of treatment to reflect this interchangeability. The next section provides formulae that permit the patient-based interpretation.

From event-based to patient-based NNT

While the event-based NNT is valid and useful, albeit with a proper interpretation, one may anyhow also wish to quantify the corresponding patient-based NNT. To do so one may of course directly compute the cumulative incidence of the first event. Alternatively, one may use the relation between the Poisson and exponential distributions to relate the incidence rate and the cumulative incidence in the following way

$$CI = 1 - e^{-IR \times t}$$

where CI is the cumulative incidence of the outcome event up to time t and IR is the incidence rate of outcome events measured in the same time units as t. For simplicity, the time axis can be considered as the time to the first occurrence of the event during follow-up, although it is also the time between events. However, this formula depends on strong assumptions for the time to event distribution, namely a constant hazard and independence between successive inter-event times, which may not often hold. In this case, the formula cannot replace the direct computation of the cumulative incidence of the first event to occur during follow-up.

In the example of the TORCH trial, the paper did not provide Kaplan-Meir curves for the time to the first exacerbation, so that the cumulative incidence cannot be computed

directly.[26] Using the formula, however, the exacerbation rate of 1.13 per patient-year in the placebo group translates to a one-year cumulative incidence of an exacerbation of 0.68 while the incidence rate of 0.85 in the combination group translates to 0.57. Thus, the corresponding NNT from these one-year cumulative incidences is 1/(0.68-0.57) = 9.1, suggesting that 9 patients need to be treated for a year to prevent one patient with an exacerbation. This value is fairly different from the "number needed to treat of four to prevent one exacerbation in 1 year" reported in the paper, thus the importance of distinguishing the interpretation of the event-based NNT computed directly using incidence rates from that using cumulative incidence. This approach can easily be used to assess different time periods. For example, in the TORCH trial, the two-year cumulative incidence of an exacerbation can in this way be estimated to be 0.90 for placebo and 0.82 for combination, resulting in a NNT of 13 patients that need to be treated for two years to prevent one patient with an exacerbation. The Table provides such examples for the two recently published studies, using various time periods, highlighting the differences in results from the two approaches.[26,27]

Recent review: 2009

We reviewed all papers published in 2009 in four major medical journals, namely the *New England Journal of Medicine,* the *Journal of the American Medical Association, The Lancet* and the *British Medical Journal*, that reported the NNT. We used MEDLINE to search the papers (term: "number needed to treat") and excluded meta-analyses. We identified 19 such papers. We found that most papers (13/19) used this measure properly. However, 9

of the 19 studies involved simple designs with fixed follow-up times, which are not subject to miscalculation, since the cumulative incidence is a proportion at the fixed study end time. However, among the 10 other papers involving varying follow-up times, only 4 used the NNT properly. The other 6 studies incorrectly used incidence rates rather than cumulative incidence, or simply proportions of patients when the follow-up time was varying.

Conclusion

The "number needed to treat" is a simple and intuitive measure of the impact of a drug or treatment that is increasingly added to the reporting of study results. However, a clear understanding of what is being counted is crucial in its interpretation. The original NNT version is patient-based and counts the number of patients that need to be treated to prevent the outcome in one patient over a given time period. As most studies result in varying follow-up times, this measure inherently requires that the cumulative incidence of the outcome be used in its computation. We found, however, that several studies continue to use simple proportions rather than Kaplan-Meier curves to estimate the cumulative incidence, thus not properly accounting for varying follow-up times.

We also noted that studies that use the NNT formula not with cumulative incidences but with incidence rates to account for varying follow-up times are confusing. By incorrectly using a patient-based interpretation, the estimation can lead to some inconsistent values of the NNT and misleading interpretations. This error is also inherent in studies that involve recurrent outcomes with multiple events. We also observed that

some studies will extrapolate the NNT beyond the study time period, for example a two-year study used to compute a 5-year NNT.

This misuse of the NNT is still very current and ongoing. Our survey of papers published in 2009 in four major medical journals found that there was no problem with the calculation of the NNT when the studies involved simple designs with fixed follow-up times. In this case, the proportion of patients with the outcome is valid for the cumulative incidence. However, in the studies involving designs with varying follow-up times, 60% did not compute the NNT correctly. Thus, in studies using more complex designs that involve varying follow-up times or outcomes with multiple events, the NNT continue to date to be miscalculated and misinterpreted.

Such inaccuracies are not trivial: an incorrectly calculated NNT of 2 patients instead of 5 implies the treatment of 20,000 instead of 50,000 patients in any impact or economic calculation. Thus, as the "number needed to treat" becomes increasingly used to compare drugs and diseases, accurate estimates become crucial to avoid misleading clinical and public health decisions.

Table

Comparison between event-based NNT computed from incidence rates and patient-based NNT computed from corresponding cumulative incidences for two recent studies.[26,27]

			Event-based approach			Patient-based approach*		
Study	Outcome event	Time span for NNT	IR_1	IR_0	NNT	CI_1	CI_0	NNT
TORCH[26]	Exacerbations	1 year	0.85	1.13	4	0.57	0.68	9
		2 years	1.70	2.26	2	0.82	0.90	13
		3 years	2.55	3.39	1	0.92	0.97	23
Lacroix[27]	Transfusions	14 days	0.47	0.86	3	0.37	0.58	5
		28 days	0.94	1.72	2	0.61	0.82	5

* Patient-based NNT based on cumulative incidence approximated by converting the incidence rate using the exponential distribution.

References

[1] 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.

[2] McAlister FA. The "number needed to treat" turns 20--and continues to be used and misused. CMAJ 2008; **179**(6)**:** 549-553.

[3] Aaron SD, Fergusson DA. Exaggeration of treatment benefits using the "event-based" number needed to treat. CMAJ 2008; **179**(7)**:** 669-671.

[4] Suissa S. Calculation of number needed to treat. N Engl J Med 2009; **361**(4)**:** 424-425.

[5] Collett D. Modelling Survival Data in Medical Research. London: Chapman & Hall, 1994.

[6] Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. Br Med J 1999; **319**(7223)**:** 1492-1495.

[7] Wiebe S, Blume WT, Girvin JP, Eliasziw M. A randomized, controlled trial of surgery for temporal-lobe epilepsy. N Engl J Med 2001; **345**(5)**:** 311-318.

[8] Kalliomaki M, Salminen S, Arvilommi H, Kero P, Koskinen P, Isolauri E. Probiotics in primary prevention of atopic disease: a randomised placebo-controlled trial. Lancet 2001; **357**(9262)**:** 1076-1079.

[9] Meunier PJ, Roux C, Seeman E, Ortolani S, Badurski JE, Spector TD *et al.* The effects of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. N Engl J Med 2004; **350**(5)**:** 459-468.

[10] Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ *et al.* Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 2009; **360**(10)**:** 961-972.

[11] Glynn RJ, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ *et al.* A randomized trial of rosuvastatin in the prevention of venous thromboembolism. N Engl J Med 2009; **360**(18)**:** 1851-1861.

[12] Zinman B, Harris SB, Neuman J, Gerstein HC, Retnakaran RR, Raboud J *et al.* Low-dose combination therapy with rosiglitazone and metformin to prevent type 2 diabetes mellitus (CANOE trial): a double-blind randomised controlled study. Lancet 2010; **376**(9735)**:** 103-111.

[13] Bucaneve G, Micozzi A, Menichetti F, Martino P, Dionisi MS, Martinelli G *et al.* Levofloxacin to prevent bacterial infection in patients with cancer and neutropenia. N Engl J Med 2005; **353**(10)**:** 977-987.

[14] Hill N, Moor G, Cameron MM, Butlin A, Preston S, Williamson MS *et al.* Single blind, randomised, comparative study of the Bug Buster kit and over the counter pediculicide treatments against head lice in the United Kingdom. Br Med J 2005; **331**(7513)**:** 384-387.

[15] Corey L, Wald A, Patel R, Sacks SL, Tyring SK, Warren T *et al.* Once-daily valacyclovir to reduce the risk of transmission of genital herpes. N Engl J Med 2004; **350**(1)**:** 11-20.

[16] Tu JV, Bowen J, Chiu M, Ko DT, Austin PC, He Y *et al.* Effectiveness and safety of drug-eluting stents in Ontario. N Engl J Med 2007; **357**(14)**:** 1393-1402.

[17] Comi G, Martinelli V, Rodegher M, Moiola L, Bajenaru O, Carra A *et al.* Effect of glatiramer acetate on conversion to clinically definite multiple sclerosis in patients with clinically isolated syndrome (PreCISe study): a randomised, double-blind, placebo-controlled trial. Lancet 2009; **374**(9700)**:** 1503-1511.

[18] Gnant M, Mlineritsch B, Schippinger W, Luschin-Ebengreuth G, Postlberger S, Menzel C *et al.* Endocrine therapy plus zoledronic acid in premenopausal breast cancer. N Engl J Med 2009; **360**(7)**:** 679-691.

[19] Kannus P, Parkkari J, Niemi S, Pasanen M, Palvanen M, Jarvinen M *et al.* Prevention of hip fracture in elderly people with use of a hip protector. N Engl J Med 2000; **343**(21)**:** 1506-1513.

[20] Hurlen M, Abdelnoor M, Smith P, Erikssen J, Arnesen H. Warfarin, aspirin, or both after myocardial infarction. N Engl J Med 2002; **347**(13)**:** 969-974.

[21] Hulley S, Furberg C, Barrett-Connor E, Cauley J, Grady D, Haskell W *et al.* Noncardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II). JAMA 2002; **288**(1)**:** 58-66.

[22] Graham DJ, Staffa JA, Shatin D, Andrade SE, Schech SD, La GL *et al.* Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. JAMA 2004; **292**(21)**:** 2585-2590.

[23] Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HA, Livingstone SJ *et al.* Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial.

Lancet 2004; **364**(9435)**:** 685-696.

[24] Beckett NS, Peters R, Fletcher AE, Staessen JA, Liu L, Dumitrascu D *et al.* Treatment of hypertension in patients 80 years of age or older. N Engl J Med 2008; **358**(18)**:** 1887-1898.

[25] Forman JP, Stampfer MJ, Curhan GC. Diet and lifestyle risk factors associated with incident hypertension in women. JAMA 2009; **302**(4)**:** 401-411.

[26] Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW *et al.* Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med 2007; **356**(8)**:** 775-789.

[27] Lacroix J, Hebert PC, Hutchison JS, Hume HA, Tucci M, Ducruet T *et al.* Transfusion strategies for patients in pediatric intensive care units. N Engl J Med 2007; **356**(16)**:** 1609-1619.