STATISTICS FOR RESEARCHERS

Number Needed to Treat

NJ Gogtay, UM Thatte

Introduction

When any patient is treated with a drug, device, vaccine, or undergoes surgery or even an investigation with a diagnostic test in a clinical research setting, two broad outcomes are possible -the patient gets better[benefit] or the patient gets worse [harmed]. An important aspect of research is to meaningfully summate data obtained in terms of measures of effect that can readily be understood and used by physicians and researchers alike. A measure that is *a single number* that can be used to compare both benefits[1] and harms of two or more preventive, diagnostic, therapeutic, or rehabilitative approaches would be extremely useful. One such single number is the "*Number Needed to Treat*" (NNT) and is defined as the number of patients that need to be treated with an intervention [relative to another intervention] in order for it to have an impact [benefit or harm] on ***one*** patient.

Since the NNT applies in equal measure to both benefit (B) and harm (H), the terms NNT-B and NNT-H[2] are used to indicate them. We have used this convention throughout this article. However,in literature, the reader may find that the term "NNT" is used synonymously for benefit (NNT –B) while the "*Number Needed to Harm"*(NNH)is used to indicate the NNT - H. We also discuss in this paper, a lesser-known metric, the *Number Needed to Screen* [NNS], one that is useful for policy makers for the use of screening tests in populations.

Table 1: The association between number of patients with fracture and treatment with either placebo or Zolendronic acid

	Number of men with fractures	Number of men without fractures	Total
Placebo	28	546	574
Zolendronic acid	09	544	553

History and Origins of the Number Needed to Treat [NNT]

The concept of NNT was first introduced in 1988 by Laupacis A *et al*,[3] who defined it as "the number of patients a clinician should treat in order to prevent one adverse outcome". Its original intended use was for benefit. The NNT concept is essentially one based on noting the frequency of occurrence of an outcome [benefit or harm] measured as a cumulative incidence of that outcome per number of patients followed up over a given time period of time.[4] This will result in a proportion of patients with the outcome over time [out of the total number followed up], which we then write as a percentage.

Understanding NNT- B, how it is Calculated and its Clinical Applications

Simply put, NNT equals the reciprocal of the Absolute Risk Reduction [ARR]. Let us understand the calculation and application of NNT using the example of a paper by Boonen and colleagues.[5] The authors evaluated the efficacy of intravenous Zolendronic acid *vs.* placebo in preventing fractures in a double-blind, randomized controlled trial. They randomized n = 1199 men with primary or hypogonadism-associated osteoporosis who were between 50 and 85 years of age to receive either 5 mg of intravenous Zolendronic acid or placebo at the beginning of the study and at 12 months. The endpoint of interest was the proportion of patients with one or more new morphometric fractures over a period of 24 months.

The results of the study were as follows – 28/574 [4.87%] men who received placebo developed fractures over 24 months compared to 09/553 [1.68%] men who received Zolendronic acid, with the difference being statistically significant [p = 0.002].

What the authors have done here is really evaluate the *association* between the use [or lack thereof] of Zolendronic acid and the reduction in number of fractures. From a previous article on Measures of Association,[6] you would remember the 2 x 2 table that we use to present this type of binary [fracture/no fracture] data. We will use the same table now to calculate the NNT- B [Table 1] in three steps.

Step 1

We first calculate the Absolute Risk [AR] of getting fracturesin both groups. We use AR1 to indicate the risk of fractures with placebo and AR2 for the risk with Zolendronic acid. This measure

Department of Clinical Pharmacology, Seth GS Medical College and KEM Hospital, Mumbai, Maharashtra
Received: 03.07.2017; Accepted: 07.07.2017

would be calculated as a proportion (or a percentage) of the number of patients who developed fractures with either treatment.

Thus, Absolute Risk [AR1] for getting fractures with placebo would be

28/574 = 0.0487 or 4.87%

And the Absolute Risk [AR2] for getting fractures with zolendronic acid would be

9/553 = 0.0168 or 1.68%

For the sake of simplicity, we round off these values to 5% and 2% respectively in the subsequent calculations.

Step 2

It is obvious from the data that the Zolendronic acid group has a significant protective effect as it has fewer fractures relative to placebo. Thus, there is a *risk difference* [RD] between the two groups and we calculate this next. The risk difference is simply the difference between the two absolute risks or AR1- AR2, and is also called Absolute Risk Reduction [ARR]

Thus, the ARR would be 5% - 2% or 3% (i.e. 3/100 = 0.03)

Step 3

The final step is the actual calculation of NNT-B, which is given by the formula

NNT-B = 1/ARR

NNT-B = 1/0.03 =33 patients.

How did we arrive at this formula and thus the number?

If Zolendronic acid were to be completely ineffective [which would be our null hypothesis], the fracture risk with both Zolendronic acid and placebo would be identical at 5%. The ARR would then be zero [5%-5%]. Zolendronic acid is however effective and only 2% patients treated with it get fractures relative to the 5% treated with placebo. If we were to see the impact on 100 patients, 2/100 would get fractures with Zolendronic acid and 5/100 with placebo.Thus, every time that n = 100 patients are treated with Zolendronic acid [rather than placebo], 3 patients [5 minus 2] would be spared the adverse outcome [fractures in this case]. Thus, if 1 patient were to be spared the adverse outcome, how many patients would be needed to be treated with Zolendronic acid?

This would be

$$\frac{100 \times 1}{3}$$

Or 33 patients.

You will realize that we have already arrived at this number using the formula 1/ARR

Thus, NNT- B is that *single number* which tells the practicing clinician about the number of patients he would need to treat with one intervention rather than another, to prevent one adverse outcome [for a defined period under defined conditions]. It can also be defined as the number of patients that would need to treated with one intervention rather than another, to prevent one *additional* adverse outcome.

In our example, we can interpret the NNT- B as follows – A total of 33 patients need to be treated with Zolendronic acid [rather than placebo] to prevent one patient from getting afracture [or additional fracture]over a 24-month period.

A perfect NNT-B would really be 1! This means that every time one patient is treated, one patient is prevented from getting an adverse outcome. It is intuitive that as the NNT-B increases, fewer and fewer patients would be helped. As a general rule of thumb, lower the NNT-B, the better is the treatment. For example, Quetiapine monotherapy has an NNT - B of 6 and the combination of olanzapine and fluoxetine an NNT - B of 4, both single digit NNTs relative to their placebo comparators[7] for the management of acute bipolar depression [approved by the US FDA].

On the other hand, Zolendronic acid [relative to placebo; the Pivotal Fracture Trial,],[8] was approved for its anti-fracture use [also by the US FDA]for the management of post-menopausal osteoporosis with a NNT of 14 for new morphometric vertebral fractures and a NNT of 91 for hip fractures. A very low NNT thus may not always be possible or necessary to allow for marketing approval and would depend upon the disease, outcome and intervention being evaluated.

Number Needed to Treat-harm [NNT-H]

The number needed to treat that we have discussed above is for *benefit*. A similar metric is the number needed to treat to harm or NNT-H or NNH as it is frequently referred to in literature is defined as the number of patients who need to be treated with one intervention [rather than another] for one patient to be harmed or for one patient to have an adverse outcome. Let us understand this as well with a published example.

Montelescot G and colleagues[9] evaluated n=4033 patients with non-ST segment elevation [NSTEMI] acute coronary syndromes to assess the effect of the timing of administration of Prasugrel (a P2Y12 antagonist) *vis à vis* the angiography, on major ischaemic events within 30 days (Table 2). The patients were divided in a 1:1 ratio into two groups- one that received 30mg of Prasugrel pre-angiography followed by 30 mg Prasugrel post angiography in the event that percutaneous intervention [PCI] was undertaken and the second group that received placebo initially followed by 60mg of Prasugrel in the event that percutaneous intervention [PCI] was needed. Safety was assessed according to the Thrombolysis in Myocardial Infarction [TIMI] criteria of major and minor bleeding episodes regardless of whether or not they were related to the PCI. The safety data of the two groups is described in Table 2. The group that received Placebo pre-treatment

Table 2: The association major and minor bleeding episodes in patient with NSTE with differential timings of Prasugrel

	Number of patients with major or minor bleeding episodes	Number of patients without major or minor bleeding episodes	Total
Prasugrel 30 mg pre angiography followed by 30 mg in the event of PCI [n = 2037]	52	1985	2037
Placebo pre-angiography followed by 60 mg Prasugrel in the event of PCI [n= 1996]	27	1969	1996

followed by 60 mg of Prasugrel had fewer bleeding episodes and this difference was statistically significant [p = 0.003].

Similar to the NNT -B, we calculate the NNT –H or NNH in 3 steps

Step 1

The absolute risk of bleeding with Prasugrel (30mg) pre-angiography [AR1] followed by 30 mg post intervention is 52/2037 or 2.6%. The absolute risk of bleeding with Placebo pre-angiography followed by 60 mg Prasugrel post intervention is 27/1996 or 1.4%.

Step 2

We next calculate the risk difference [RD] or Absolute risk reduction [ARR] as

AR1-AR2 or 2.6 – 1.4 = 1.2% (i.e. 1.2/100 = 0.012)

Step 3

The NNT –H or NNH is calculated as 1/ARR or 1/0.012 or 83 patients.

This is interpreted to mean that everytime 83 patients are treated with Prasugrel [30 mg] pre angiography followed by 30 mg post intervention rather than placebo followed by Prasugrel [60 mg], one *additional* patient will experience a major or minor bleeding episode [harm]. The authors of the paper concluded that pre – treatment with Prasugrel increased the rate of bleeding complications. The study was also stopped by the Data Safety Monitoring Committee for safety concerns.

Unlike the NNT-B, the NNT-H or NNH should be high as this indicates lesser likelihood of harm relative to the comparator. For example, the NNT- H of Valbenazine, a newly approved drug for the management of Tardive Dyskinesia is 76 [for discontinuation due to an adverse event] compared to a NNT - B of 4 [both NNT-B and NNT-H being relative to placebo comparator] over a six-week period.[10]

Likelihood to be Helped or Harmed [LHH] – the Ratio of NNH to NNT

Since interventions can produce both benefits and harm, any comparison of two interventions will produce two NNTs – one for benefit [NNT-B] and one for harm [NNT-H]. A lesser-used metric called the *"Likelihood to be helped / harmed"* [LHH] is calculated as the ratio of NNT-H to NNT-B since treatment decisions are almost always a trade-off between harm and benefit. Intuitively, the value of LHH should be greater than 1 and the further away from 1 that the value is, greater is the likelihood that the intervention produces more benefit than harm. Let us understand this with an example.

Srivastava and Ketter[7] in their eloquent narrative review evaluated RCTs that studied quetiapine, olanzapine and lamotrigine among other drugs for the management of acute bipolar depression [all studies with placebo comparators]; a difficult to treat disorder. Quetiapine [a second-generation anti-psychotic] had a NNT-B of 6 and a NNT-H of 6 for sedation giving a LHH for efficacy: sedation of 1. Olanzapine [also a second-generation anti-psychotic] had a NNT-B of 12, and a NNT-H for sedation of 7 giving a LHH value of 0.58. Lamotrigine, a mood stabilizer had a NNT-B similar to olanzapine of 12, but a NNT-H for sedation of 42, giving an efficacy: sedation LHH value of 3.5. The LHH values thus indicate superiority of Lamotrigine over quetiapine and quetiapine over olanzapine [in that order] in terms of risk *vs.* benefits, thus enabling the clinician to make an informed choice.

The Number Needed to Screen [NNS]

National strategies for disease screening to identify patients at risk of developing disease or with yet undetected disease [for example, use of the Prostate Specific Antigen (PSA) for the diagnosis of prostate cancer or mammography for the detection of breast cancer] require evidence that measures the value addition that any screening test provides. The Number Needed to Screen [NNS] is one such metric and was first developed by Rembold in 1998[11] as a metric to define the number of people that are needed to be screened to prevent one death or one adverse event or one life-year gained.While it is conceptually similar to the metrics of NNT- B and NNT- H, its calculation differs slightly.

We require the knowledge of two elements before beginning the calculation of NNS

- The background risk or prevalenceof the disease in the population
- Knowledge of mortality or an adverse outcome in screened and unscreened cases

Let us now understand the steps in the calculations of NNS with an example.

Step 1- Calculate the cumulative rate of deathsin the two groups

Unscreened group- The Cancer Intervention and Surveillance ModelingNetwork (CISNET) USA,

estimates that the mortality from breast cancer in the absence of any screening mammography would be 3% for a woman aged 40 years or older.[12] Thus, the death rate without screening would be 3% or 30 per 1000 women screened.

Screening group- Let us assume that a screening technique X is developed that reduces mortality by 90%. Now the deaths will be 3 per 1000 women screened.

Step 2 –Calculate the number of deaths prevented [lives saved] due to screening[13]

Because intervention X is 90% effective, 27 lives per 1000 women screened are saved [or deaths prevented]

Step 3- Calculate the number needed to screen as the reciprocal of the absolute difference in cumulative mortality[13]

Since 27 deaths were prevented for 1000 women screened, for one death to be prevented, 1000/27 or 37 women would need to be screened [NNS = 37 for an intervention that reduces mortality by 90% from 30/1000 to 3/1000].[13]

If intervention Y were to produce only a 10% reduction in morality, 0.1 x 30 = 3 lives per 1000 women screened would be saved

Thus, to save 1 life, 1000/3 or 333 women need to be screened [NNS = 333 for an intervention that produces a 10% reduction in mortality]

Logically therefore, the lower the NNS, the more useful is the screening test.

Challenges Associated with the Use of NNTs

Interpret with caution and with an understanding of the baseline risk

A risk is essentially the probability that something will happen. If we toss an unbiased coin once for instance, the "risk" of heads would be 50%. Similarly, we can define risks in medicine as well. Let us take a hypothetical example of a drug A reducing the risk of dying of myocardial infarction from 3% to 2%. The NNT would be 100. Drug B reduces the risk of dying from rabies from 100% to 99%. The NNT is again 100. These NNTs simply cannot be compared! The reason is that rabies is a disease with 100% mortality [baseline risk] and even a 1% reduction [giving a NNT of 100] will make a huge impact to the disease. In the MI example, while the NNT of 100 is the same, given that the baseline risk of death itself is low, the 1% reduction may or may not really be meaningful. Thus, any comparison of NNTs [benefits or harm] must be made with a clear understanding of the baseline risks that are involved.[14]

When NNTs are needed to be calculated for "time to an event"

When we carry out survival analysis [time to an event analysis],[15] the calculation of the number needed to treat can become difficult as patients will have varying follow up times and some of them may be censored as well. The calculation is thus *significantly dependent upon time*. In survival or time to an event study, we use the term NNT[t] and calculate more than one NNT[t] at several time points [which can be fixed in advance]. The calculation at each time point is based *on the survival probability at that time point* which is estimated either by the Kaplan-Meier method or Cox regression method. A time specific NNT(t) is defined as the average number of patients needed to be treated to observe one event-free patient *more* in the intervention group relative to the control group at a given time point *t*.[16]

Presenting NNTs in Research Papers – Key Points to be Remembered

Though NNTs are now widely used, their reporting in literature is less than optimal.[17] The following need to be remembered while presenting NNTs in publications.

Confidence Intervals for the Number Needed

Similar to other measures we estimate in statistics [for example risk ratio, odds ratio] where we give confidence intervals to help the reader gauge the margin of error or uncertainty that was seen with the study, the number needed to treat similarly needs to be accompanied by a confidence interval. Several methods exist for the calculation of CIs and the Wald method is a commonly used one.[18]

Stating the direction of effect

Although the NNT was originally devised as a measure of benefit, as interventions can produce both harm and benefit, simply stating the NNT without giving the direction [benefit or harm] can become difficult for the reader. Thus, the terms NNT-B to indicate benefit and NNT-H are recommended for use while presenting this metric.[2]

The importance of stating the comparator

Since the NNT makes use of Absolute risks in both groups, it is logical that both groups are alluded to when presenting the NNT. However, this does not happen routinely in literature. For instance, simply stating that Drug X has a NNT of 25 makes no sense, unless the comparator Drug Y is clearly stated.

Stating the time frame of the study

Randomized controlled trials [RCTS] are often conducted over a long period of time. Hence stating the time frame along with the presentation of the NNT becomes very important. Let us understand this with an example. Study 1 compares two treatments X and Y over a two-year period and yields a NNT-B of 25. Study 2 compares two treatments A and B over a 10-year period and gives a NNT-B of 25. Though the NNT – B for both studies are identical, it is a very different matter to produce benefit in 1 patient for every 25 patients treated over 2 years versus over 10

years! Thus, mentioning the time frame for the study is a crucial component for presenting the NNT. In the example of association of fractures with Zolendronic acid, the NNT of 33 is over a 24-month period.

Criticisms of the NNT

Katz N and colleagues[19] in their eloquent narrative review have put together evidence on and summarized the challenges associated with the use of the NNT. These range from the NNT having an infinite value when the ARR is zero or close to zero [when the interventions tested have very similar effects], to the NNT being dependent upon the choice of the binary outcome or not translating into the same NNT value when the intervention is actually used in the real-world setting.

NNT and the Cost and Reimbursements of Treatments

Graziano and colleagues[20] have suggested that the NNT can be used by policy makers for pricing negotiations with the pharmaceutical industry and have expounded on this using the example of NNT for regorafenib [salvage therapy for metastatic colorectal cancer, the CORRECT trial].[21]

Conclusions

RCTs report results in a wide variety of ways that include relative risk, odds ratio, hazard ratio and the p value when two interventions are being compared. Many clinicians have difficulty in translating these findings into actual patient care as the answer to the question "which therapy between the two should I use in my patient?" is often not clear to them.[22] The NNT offers clinicians a "yardstick" for measurements as it helps compare benefits and harms of treatment by converting them into a single number. It helps practicing clinicians make an informed choice when more than one intervention is available. The explanation and elaboration document of the CONSORT guidelines suggest that the NNT-B and NNT-H can be presented as metrics for binary data from Randomized Controlled Trials.[23] Clinicians should learn how to derive and use NNT from results of RCTs as the reciprocal of the Absolute risk difference. However, since benefits and risks are two sides of the same coin, each intervention in a RCT would have both a NNT-B and a NNT-H and both need to be considered in tandem so as to make careful, individualized, patient -centric as also policy decisions.

Acknowledgment

The authors are grateful to Dr. Robin Ferner from the West Midlands Centre for Adverse Reactions Monitoring, Birmingham, for his constructive inputs that helped refine the manuscript.

References

1. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995; 310:452–454.
2. Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998; 317:1309e12.
3. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318:1728e33.
4. Suissa S. The Number Needed to Treat: 25 Years of Trials and Tribulations in Clinical Research. *Rambam Maimonides Medical Journal* 2015; 6:e0033.
5. Boonen S, Reginster JY, Kaufman JM, et al. Fracture risk and zoledronic acid therapy in menwith osteoporosis. *N Engl J Med* 2012; 1;367:1714-23.
6. Gogtay NJ, Deshpande S, Thatte UM. Measures of Association. *J Assoc Phy Ind* 2016; 64:70-3.
7. Srivastava S, Ketter AT. Clinical relevance of treatments for acute bipolar disorder: balancing therapeutic and adverse effects. *Clinical Therapeutics* 2011; 33:B40-B48.
8. Black ND, Delmas DP, Eastell R, et al. Once-yearly zolendronic acid for the treatment of post menopausal osteoporosis. *N Engl J Med* 2007; 356:1809-22.
9. Montalescot G, Bolognese L, Dudek D, et al. ACCOAST Investigators. Pretreatment with prasugrel in non-ST-segment elevation acute coronary syndromes. *N Engl J Med* 2013; 369:999-1010.
10. Citrome L. Valbenazine for tardive dyskinesia: A systematic review of the efficacy and safety profile for this newly approved novel medication-What is the number needed to treat, number needed to harm and likelihood to be helped or harmed? *Int J Clin Pract* 2017; May 12.
11. Rembold CM. Number needed to screen: Development of a statistic for disease screening. *BMJ* 1998; 1:317:307.
12. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under differentscreening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009; 151:738–747.
13. Hendrick ER, Helvie MA. Mammography screening: a new estimate of number needed to screen to prevent one breast cancer death. *AJR* 2012; 198:723-28.
14. McCormack PJ. Number needed to treat and baseline risks. *CMAJ* 2008; 179:1174.
15. Gogtay NJ, Thatte UM. Survival analysis. *J Assoc Phy Ind* 2017; 65:80-84.
16. Altman DG, Andersen PK. Calculating the number needed to treat where the outcome is time to an event. *BMJ* 1999; 319:1492-1495.
17. Hildebrandt M, Vervölgyi E, Bender R. Calculation of NNTs in RCTs with time-to-event outcomes: A literature review. *BMC Medical Research Methodology* 2009; 9:21.
18. Bender R. Calculating confidence intervals for the number needed to treat. *Controlled Clinical Trials* 2001; 22:101-110.
19. Katz N, Paillard CF, Inwegen VR. A review of the use of the Number needed to Treat to Evaluate the efficacy of analgesics. *The Journal of Pain* 2015; 16:116-123.
20. Graziano F, Rulli E, Biagioli E, et al. Number needed to treat for pricing costly anti-cancer drugs. The example of regorafenib in metastatc colorectal cancer. *Annals of Oncology* 2016; 27:247-48.
21. Grothey A, Van Cutsem E, Sobrero A, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multi-centre, randomised, placebo-controlled, phase 3 trial. *Lancet* 2013; 381:303–312.
22. Cordell HW. Number needed to treat [NNT]. *Annals of Emergency Medicine* 1999; 33:433-36.
23. Moher D, Hopewell S, Schulz FK, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340:c869.