Introduction

The need to express estimates of risk in an understandable manner is a challenge faced regularly by those who work with the results of epidemiological research and who need to convey their meaning to clinicians. Traditional statistical measures from systematic reviews cannot be immediately applied to clinical practice. Clinicians also need a readily understandable tool for weighing the risks of various treatments. Ideally this should be feasible without recourse to complicated statistical concepts. The concept of rate difference is converted into a number of individuals, a more intuitively understandable quantity. This quantity was named “number needed to treat” (NNT). NNT is becoming a popular measure of treatment effect. This number can be calculated easily from raw data or from statistical estimates and can be applied to different end points (treatment efficacy, harm, and other outcomes).
The aims of this review are to summarize the value and limitations of "NNT" as a measure of treatment effect in randomized controlled trials.

Traditional measures of treatment effect

The traditional measures of treatment effect are relative risk, odds ratio (OR) and absolute risk reduction (ARR). Table 1 summarizes the results of a study of a new treatment for prevention of postoperative nausea and vomiting (PONV) against a control (existing treatment).

Relative risk (RR)

Relative risk (also known as the "risk ratio") is defined as the probability of an event in the study group divided by the probability of an event in the control group. Mathematically in the above study it is

\[
\frac{a}{a+c} = \frac{b}{b+d}.
\]

It expresses the relative probability that an event will occur when the two groups of patients are compared. The complement of relative risk (1-RR) is known as relative risk reduction (RRR). RRR is an estimate of the percentage of baseline risk that is removed by the treatment. A RRR of 0 indicates no benefit or harm from the treatment compared to the control group, whereas a RRR of 1 indicate a "100% benefit".

Odds ratio (OR)

"Odds" expresses the probability that a particular event will occur against the probability it will not occur. In the above example, the odds of PONV are \(a/c\) in the study group and \(b/d\) in the control group. The OR is defined as the odds of an event in the study group divided by the odds of an event in the control group. OR is the traditional epidemiological expression of the relative likelihood of an outcome. Because OR has statistical advantages over relative risk in terms of sampling distribution, it is preferred in meta-analyses. ORs are often used as approximations of RRs. This association is generally considered valid as long as the event is relatively rare (incidence less than 10%). As the incidence of the event increases, this relationship breaks down and OR tends to overestimate benefits and harms of treatment.

It is important to note that RR, RRR and OR represent measures of treatment effect relative to the control event rate and do not reflect changes in the baseline risk of an event. A RRR of 40% may be significant, but if the baseline risk of the event in the population is very small, potential toxicity and cost of the treatment may not justify its use.

Absolute risk reduction (ARR)

ARR is the absolute difference in risk of an event between the experimental and control groups. In the above example, ARR is mathematically equal to: \(b/(b+d) - a/(a+c)\). ARR represents the number of events (in this case PONV) prevented as a proportion of all patients treated. In contrast, the RRR expresses the number of events prevented as a proportion of the number of events expected. It does not reflect the risk of the event without treatment, and cannot discriminate large treatment effects from small ones. The ARR incorporates the baseline risk of the event without therapy and the risk reduction with therapy. It, like RR, RRR, and OR, is not an intuitively easy means by which clinicians can assess treatment effect in trial reports and apply them at the bedside.

Number needed to treat (NNT)

In the context of randomized trials on the desirable effects of treatment, Sackett proposed a method of converting rate differences into a more intuitive quantity called "NNT". NNT is the "number of patients one would need to treat with the experimental therapy in order to prevent one additional adverse event or attain one additional benefit". Mathematically it is calculated as 1/ARR. NNT is usually expressed as a whole number (as its point estimate) and can be obtained for any trial that has reported a binary outcome (i.e. event or no event). Alternatively, NNT may be calculated using the risk of the event in the control group and the RRR as follows:

\[
\text{NNT} = \frac{1}{(Pc \times RRR)}
\]

where Pc is the risk of the event (PONV in the above example) in the control group (i.e. \(Pc = b/(b+d)\)).
A nomogram has been proposed to simplify calculation of NNT at the bedside for individual patients. The NNT defines the treatment-specific effect of an intervention or therapy. An NNT of 1 means that the treatment is effective in all patients. Most treatments or interventions will have an NNT higher than 1 because they are effective in some but not all the patients. An NNT of 2–3 indicates that a treatment or intervention is quite effective. However, NNT must not be used in isolation; the data from which it is derived should also be considered as heterogeneity of data alters NNT; and we need to consider age, epidemiological and others factors. In prophylactic interventions, such as adding aspirin to streptokinase to reduce the reinfarction rate following acute coronary syndromes, the NNT may be as high as 20–40 but may still be considered clinically effective.

**Number needed to harm (NNH)**

The NNH is a similar concept proposed by Sackett and colleagues to express the probability of additional adverse events occurring in clinical trials because of the treatment. The absolute difference in risk of harm between the experimental group and control group is known as the Absolute Risk Increase (ARI). The reciprocal of ARI is the “NNH” for that therapy. NNH is defined as the number of patients who must be treated with a therapy in order to have one additional patient suffer an adverse effect compared with the control treatment. However, undesirable effects often become apparent only after a treatment or intervention has become part of clinical practice. It may therefore be not possible to study them in a randomized controlled trial, making it necessary to resort to phase 4 studies or the less ideal case–control study. Bjerre and LeLorier proposed the term “number of patients needed to be treated for one additional patient to be harmed” (NNTH) which is derived from case–control studies. The NNTH is “the number of people exposed to a given treatment such that on average and over a given follow-up period one additional person experiences the adverse effect of interest because of the treatment.” It expresses the additional absolute risk of an adverse effect produced by the treatment and does not express the total risk attributable to the combination of the background risk and the risk due to exposure. It is calculated by the following formula:

$$\text{NNTH} = \frac{1}{(\text{OR} - 1)\text{UER}}$$

where OR is the odds ratio provided by the case–control study, and UER is the unexposed event rate. NNTH is a measure of absolute risk because it takes into account the background risk of the outcome occurring in unexposed people, unlike the OR that measures relative risk. NNTH is a composite measure that takes into account the OR and the unexposed event rate. Consequently, it provides a better estimate of risk that corresponds with the clinical situation (reality). A major limitation of NNTH is that it is not always possible to find a study that provides an approximate estimate of the unexposed event rate (which can be estimated from the controls in randomized controlled studies or the unexposed subjects in cohort studies).

**Number needed to screen (NNS)**

NNS is defined as the number of people that need to be screened for disease (for a given duration) to prevent one death or one adverse event. It can be calculated in trials that directly test the effectiveness of a screening strategy by finding the reciprocal of the ARR. It may also be estimated in clinical trials measuring the benefit of treating risk factors. In this case, the prevalence of unrecognized and untreated disease that can be detected must be known. NNS is calculated by dividing the NNT for treating risk factors by the prevalence of disease that was unrecognized or untreated.

**Time to the event**

Implicit in studies with binary outcomes is the dimension of time. The differences between event rates in experimental and control groups are specific for the period of follow-up. Consequently, the NNT for a study will vary depending on the length of follow-up. Comparisons of outcomes between treatments cannot be directly made if the follow-up times are different. Direct comparison of NNT for studies of different follow-up times assumes that the relative benefit of the treatment is constant over time (i.e. the RRR is constant). The NNT for at least one of the studies is then time-adjusted so as to conform with the time period of the study with which it is compared. This is achieved by applying the following formula:

$$\text{NNT}_{\text{(time adjusted)}} = \frac{\text{NNT}_{\text{(actual)}}}{(\text{Actual follow up time}) / (\text{comparative follow up time})},$$
where comparatively follow-up time is the follow-up time to which the NNT is to be adjusted.

Confidence intervals for NNT

A means of reporting the precision of study results is necessary when using NNT. This is usually expressed as the 95% confidence interval (95% CI) for the ARR and is given by

$$\text{ARR} \pm 1.96 \times \text{SE(ARR)},$$

where SE(ARR) is the standard error of the absolute risk reduction. The 95% confidence interval for NNT is the reciprocal of the values defining the confidence interval for ARR. For example, if in a trial the ARR is 10% and the 95% CI is 5–20%, the NNT becomes 10 (1/0.1) with 95% CI of 5–20 (1/0.2–1/0.05). The 95% confidence interval of NNT indicates that 19 out of 20 times the "true" value will be in the specified range. An NNT with an infinite confidence interval is a point estimate and includes the possibility of no benefit or harm.

Modifications of number needed to treat

A number of variations and extensions of NNT have been proposed to improve its utility. These include adjusting NNT for Utility and Timing of Benefits and Harms, NNT Unqualified Success and Unmitigated Failure, and "threshold NNT". However, these modifications are not used in anaesthesia, critical care medicine and pain medicine literature, and will not be described.

Advantages of NNT

There are numerous advantages in using NNT to express study outcomes. NNT expresses the efficacy of a treatment in a manner which incorporates baseline risk without therapy and the risk reduction with therapy. For example, a treatment which reduces mortality rate from an illness from 1 per million to 0.5 per million represents a RRR of 50% but a NNT of 2 million. Hence NNT quantifies the baseline risk as well as the RRR.

One of the strengths of NNT is that it is a simple and intuitive estimate measure of effectiveness of interventions and is easier to comprehend compared with other statistical descriptors. It has the advantage that it conveys both statistical and clinical significance. Furthermore, it can be used to extrapolate published findings to a patient when the risk reduction associated with treatment is constant for all levels of risk. Presentation of study data in terms of NNT as opposed to RRR results in a more conservative appraisal of treatment effectiveness. It is a more intuitive, relevant, and easily memorized approach for clinicians than ratios (e.g. RRR, ARR).

Limitations of NNT

NNT has limitations like all other measures of treatment effect. The ability of NNT to combine baseline risk and RRR into a single number can be a disadvantage. This is best illustrated with an example. A treatment to prevent an event with a low baseline risk of 20% has a RRR of 50%. Hence the NNT is 10. However, a treatment with a high baseline risk of 80% but a smaller RRR of 12.5% will also give a NNT of 10. An NNT of 10 tells us that, on average, 10 patients must be treated to prevent one additional event. This, however, does not provide any information about the fate of the other 9 patients. In the situation of the low baseline risk, far fewer of the 9 remaining patients are at risk of the event than the 9 in the situation of high baseline risk. It must be appreciated that NNT reflects the average number of patients that must be treated to prevent an event and it does not indicate the fate of the other patients or virulence of the disease process. Furthermore, NNT of 10 means that 9 of the 10 patients either do not require therapy or will not respond to treatment.

NNT can be calculated by pooling absolute risk differences from trials in meta-analyses. Such pooled NNT may be misleading because baseline risk may vary between the different trials considered. Changes in secular trends over time (example: changes in disease severity) may also influence the NNT generated, as can differences in the clinical settings in which the trials were conducted. In calculating the pooled NNT, it is suggested that a pooled estimate of ARR (or increase) is obtained first, from which NNT can then be calculated, rather than combining NNT values directly. This avoids meaningless results that may otherwise result.

NNT for a treatment is time dependent, and if the treatment produces a constant RRR over time, one can expect that NNT will decrease as the length...
of follow-up increases. Therefore, when comparing NNT for different studies, the length of follow-up must be standardized to allow direct comparison. Calculation of NNT for studies involving chronic diseases must take into account of the fact that mortality does not cluster in time, and that the duration of follow-up is rarely sufficiently long to record the outcome in all patients. The NNT calculated thus depends on the point in time at which the difference in risk is measured. In contrast, studies of acute conditions (e.g., acute myocardial infarction) where mortality does cluster in time (usually within the follow-up period). It is proposed that NNT should be uniformly calculated independently of duration of follow-up for chronic conditions. To achieve this, the NNT in such studies is calculated as the reciprocal of the hazard difference, where hazard is expressed as mortality per unit of person-time. The NNT then expresses the amount of person-time of follow-up required in each arm that results in exactly one less death in the treated arm. This NNT does not depend on the duration of follow-up as long as the hazard difference is constant across time. NNT has also been shown to be sensitive to crossover between treatment arms, and adjustments for such occurrences in trials have been proposed. NNT analysis of results may be misleading in studies that compare treatments which have effects on different subsets of the population, or where the treatments exert effects over different periods of time.

Other criticisms of NNT include the possibility that NNT may be higher in a study comparing treatment with placebo than when comparing treatment with no treatment due to the placebo effect itself. NNT and NNH do not capture the patients’ individual likelihood of benefit and harm. Methods have thus been described where the baseline risk of an event for a particular patient can be estimated from trial data, hence generating a patient-specific NNT and NNH. The patient’s likelihood of being helped versus harmed ratio (LHH) can be calculated as 1/NNT (or ARR):1/NNH (or ARI). This ratio tells the patient the likelihood that the treatment will help as opposed to harm him/her. The patient’s own values and preferences regarding benefits and harms can be incorporated into this ratio.

The variation in baseline risk has prompted suggestions that RRR is advantageous in outcome reporting (as opposed to NNT) as it is independent of baseline risk and thus has the same value in all patients. The issue of baseline risk has also been cited as limitation in terms of trial design. It is suggested that participants in studies are more likely to be healthier than the general population as a result of inclusion and exclusion criteria of the studies. This would have the effect of inflating the apparent NNT.

Examples of applications of number needed to treat

Anaesthesia

NNT has been a useful measure to report results of studies in prevention and treatment of PONV. One systematic review of prevention of vomiting after paediatric strabismus surgery reported NNTs for the efficacy of various anti-emetic medications based on randomized controlled trials. Adverse effects (such as extrapyramidal symptoms, sedation, drowsiness, and the oculocephalic reflex in the case of propofol) also had their corresponding NNHs reported. A systematic review of the efficacy of ondansetron in the treatment of established PONV reported a NNT of approximately 4, with no significant difference between doses of 1, 4, and 8 mg. A meta-analysis of randomized controlled trials to assess the efficacy of droperidol and 5-HT3-receptor antagonists alone and in combination for prophylaxis of PONV used pooled data to generate corresponding NNTs. In this case, there was no statistically significant improvement in antiemetic effect when droperidol and 5-HT3-receptor antagonists were combined compared with droperidol alone. This was reflected in the confidence intervals of the NNTs which included ∞. The authors noted the dependence of NNT on the baseline risk of the event studied and suggested that combinations of droperidol and 5-HT3-receptor antagonists be used only on patients at very high risk of PONV.

The concept of NNH was used in a review of the causation, frequency, and severity of bradycardia with propofol use. In controlled clinical trials, the NNH was 11.3 (95% CI 7.7–21). In other words, one would have to administer propofol to approximately 11 patients to cause one additional bradycardia which would not have occurred if another (control) anaesthetic agent had been used instead. The NNH for bradycardia in paediatric strabismus surgery was 4.1 (95% CI 3–6.7).

Pain management

Systematic reviews have been conducted to evaluate the efficacy of various analgesic agents in postoperative pain. The NNT for paracetamol
1000 mg to achieve at least 50% pain relief compared with placebo was 4.6. Paracetamol 600/650 mg plus codeine 60 mg had a NNT of 3.6. The authors noted that comparison with NNTs for other analgesics (obtained by quantitative systematic reviews) was possible and thus a ladder of analgesic efficacy could be devised. A systematic review with meta-analysis for single-dose ketorolac and pethidine in acute postoperative pain reported that the NNT to produce at least 50% pain relief was 2.9 (95% CI 2.3–3.9) with intramuscular pethidine 100 mg. The numbers needed to harm for pethidine at this dose for drowsiness and dizziness was 2.9 (2.2–4.4) and 7.2 (4.8–14), respectively. In the case of intramuscular ketorolac 30 mg, NNT was 3.4 (2.5–4.9), and for oral ketorolac 10 mg, NNT was 2.6 (2.3–3.1). The NNH for any adverse effect with oral ketorolac 10 mg was 7.3 (4.7–17).

Conclusion

NNT provides a means by which study data may be presented in a way considered by many to be more clinically useful than other methods. It can be extended to improve its utility and allow better planning and distribution of health resources. NNT can also be modified to guide treatment decisions for individual patients. It is being increasingly used in clinical research and in systematic reviews. However, an appreciation of its limitations and caveats is required in order to draw appropriate conclusions, and acceptance of NNT as a preferable means of study outcome presentation is by no means universal.

References