MiniReview

Number-Needed-to-Treat (NNT) – Needs Treatment with Care

Palle Mark Christensen1 and Ivar Sonbo Kristiansen2

1Research Unit of Clinical Pharmacology and Research Unit for General Practice, University of Southern Denmark, Odense, Denmark and 2Research Unit of General Practice, University of Southern Denmark, Institute of Health Management and Economics, University of Oslo, Norway

(Received November 17, 2005; Accepted January 18, 2006)

Abstract: The “number-needed-to-treat” (NNT) was introduced about 15 years ago and has gained widespread use. It has been claimed to be “easy to understand” and gives “intuitive meaning”. When used to measure the effectiveness of interventions targeting chronic disease processes e.g. atherosclerosis and osteoporosis, NNT (as well as relative and absolute risk reduction) does not capture the crucial time component, a fact that has important consequences: NNT varies over time, it may not mean that adverse events (fractures, myocardial infarctions etc.) are avoided, but simply that they are postponed. Finally, empirical studies indicate that lay people and doctors misunderstand NNT. We recommend that NNT be used with considerable care. There is probably no single effect measure that is able to convey all necessary information.

Number Needed-to-Treat (NNT) in evidence-based medicine.

NNT has been promoted by many scholars, especially by advocates of evidence-based medicine (McCormack & Levine 1993; Black et al. 1995; Jaeschke et al. 1995; Sackett et al. 2002; Schulzer & Mancini 1996) claiming for example that ‘number needed to be treated (NNT) to prevent one event is the most useful measure of clinical effort ... patients must expend in order to help them avoid bad outcomes’. The concept of NNT has been expanded to encompass for example harm (‘number needed to harm’ – NNH), screening (‘number needed to screen’) (Rembold 1998), education (‘number needed to educate’) (Gallefoss & Bakke 2001), and offence (‘number needed to offend’) (Stone et al. 2000).

In this review, we will first explore the characteristics of NNT and second review how NNT is understood by patients and doctors.

Definition of some quantitative measures of therapeutic effects.

Assuming that \(D_c\) (frequently called CER in evidence-based medicine) denotes the proportion of adverse events (deaths, heart attacks, hip fractures etc.) in a control group and \(D_i\) (EER in evidence-based medicine) denotes the proportion of adverse events in an intervention group, then the quantitative effects of the therapy can be expressed in several ways:

1. \(D_c - D_i = \text{Absolute Risk Reduction (ARR)}\)
2. \((D_c - D_i)/D_c = \text{the relative risk reduction (RRR)}\)

In 1988, Laupacis et al. suggested to use the reciprocal of the absolute risk reduction as a measure of therapeutic effectiveness. They called this measure the number-needed-to-treat – NNT.

3. \(1/(D_c - D_i) = \text{Number-needed-to-treat (NNT)}\)

Quantitative measurements of therapeutic effect and timing of adverse events.

When thinking in terms of NNT it is important to distinguish between two clinical situations depending on the timing of the event. If the effect of the intervention is immediate, e.g. when treating ventricular fibrillation the patient will either be cured within minutes or die, the use of NNT poses no major problem (Kristiansen et al. 2002). However, in situations in which the intervention is aimed at continuous disease processes such as atherosclerosis or osteoporosis, the use of NNT may cause several problems, which will be the scope of this review.

To the extent that interventions slow or halt the disease processes such as atherosclerosis (causing adverse events such as myocardial infarction and stroke) or osteoporosis (causing bone fractures) the adverse consequences are postponed. If the postponement is longer than the remaining
life span, the outcome is totally “avoided”. Even though postponement and the time dimension are crucial for the understanding of the intervention effects, it is impossible to measure postponement of death or other adverse events directly. What we can do is to observe the timing of adverse event in either treated or untreated groups of patients and then compare the proportions of individuals with the adverse event in the two groups, i.e. Dc and Di as mentioned above. Fig. 1 depicts the results of a clinical trial aimed at exploring survival benefits from an ACE inhibitor in patients who have sustained a heart attack (Torp-Pedersen & Kober 1999). The risk of fatal outcomes at for example five years of follow-up is about 44% in the intervention group and 51% in the control group. The absolute risk reduction is then 7% (51%–44%=7%), or more precisely 6.4% according to the publication giving an NNT of 1/0.064=15.6.

Pitfalls of NNT

The main of problem of using NNT (as well as with absolute and relative risk reduction) stems from the fact that interventions for chronic diseases have a crucial time dimension that is not captured by metrics measured at one single point in time (Kristsiansen et al. 2002). In fig. 1, the vertical distance between the two survival curves represents ARR. NNT is 31, 18, 11, 15, 16 and 16 after 1, 2, 3, 4, 5 and 6 years of follow-up. Which one of these NNTs is then best suited as the basis of medical decisions? The end of a clinical trial may be seen as a “natural” measuring point, but it is still arbitrary. When NNT is presented for a therapy, it is crucial to know when in the course of the therapy NNT is measured, and whether NNT varies in time. Survival curves often vary over time and thus the vertical distance between the curves varies, and hence NNT, varies in time.

NNT has been interpreted as a measure of the probability of benefit from therapy (Laupacis et al. 1988; Sackett et al. 2002) (e.g. a value of NNT=10 has been interpreted as the likelihood of benefit from therapy, 1/10). This is unlikely to be true for chronic disease processes. Imagining a theoretical example in which every 10 patients in a control group sustains a hip fracture one by one, at one month intervals (fig. 2). After 10 month, all 10 patients had sustained a fracture. In an otherwise identical intervention group, the hip fractures are postponed by one month such that at any time one less patient in the intervention group has sustained a fracture. This means that the absolute risk reduction is 1/10 and hence NNT 10 whenever it is measured. This example illustrates that although all patients experienced an effect, the NNT was 10, which by some was interpreted as if only 1/10 had effect and the rest 9/10 did not have an effect. In other words, we cannot infer the proportion of patients who (dis-)benefit from a therapy by using NNT or other metrics measured at a single point in time. From a biological point of view, it seems plausible that a higher proportion than 1/NNT benefit from the intervention. For instance an angiographic study has shown that about 50% of patients, there was an effect on the coronary arteries of lipid lowering therapy (Brown et al. 1990). In experienced a study of osteoporosis, 79% had stable or increasing bone mineral density when using a bisphosphonate (Hochberg et al. 1999).

Expressing the 95% confidence interval of NNT is problematic, especially if the confidence interval of ARR encompasses zero. If for example the confidence interval for ARR is [−0.01, +0.02], the confidence interval for NNT is not [−100, 50]. Firstly, NNT cannot take values in the interval [−1, 1] because at least one patient should be treated to

---

**Fig. 1.** Survival in patients with reduced left ventricular function after an acute myocardial infarction (heart attack). The dotted line (black) represents patient who received active drug (trandolapril) while the other represents a similar control group. The vertical lines (red) line in the upper panel indicate a point in time (two years from start of therapy) at which NNT might be estimated. The publication (Torp-Pedersen & Kober 1999) presents data for 7 years follow-up. Reprinted from Torp-Pedersen & Kober (1999) with permission from Elsevier.

**Fig. 2.** A hypothetical extension of the study in fig. 1 until all patients are dead is indicated by red solid and dotted lines. The area between the survival curves for the treatment group and the control group represents the gain in mean survival time (life expectancy).
avoid or induce one event. Secondly, when the ARR approaches zero from the positive end of the scale, the value of NNT approaches $+\infty$, and $-\infty$ when ARR approaches 0 from negative end of the scale. Thus in the above mentioned example the confidence interval would be one at $[-\infty, -100]$ and another at $[50, \infty]$, which seems difficult to interpret.

All these aspects of NNT require a change in the commonly used interpretation of NNT. An NNT of for example 10 does not mean that “10 patients would have to be treated with treatment A for three years to prevent one event”. Rather, the correct interpretation is that “on average 10 patients would have to be treated for three years with treatment A to observe one fewer adverse event after three years” (Hutton 2000). NNT does not say anything about the proportion of patients affected by the therapy, nor does it say whether outcomes are avoided forever or not.

Finally it should be noted that comparing two interventions may produce the same RRR but different ARR (and thus NNT), thus giving rise to different interpretations of the two interventions. For example, in a theoretical study where 1000 patients would receive placebo and 1000 patients would receive a new drug, an event rate of 100 and 20, respectively, or of 20 and 4, respectively, would yield the same RRR of 80% but a NNT of 12.5 (12 or 13) in the first case and of 62.5 (62 to 63) in the second case.

**Alternatives to NNT**

All quantitative measures of effectiveness according to equations (1–3) (RRR, ARR, and NNT) are based on the proportions surviving in an intervention and a control group at a specific time-point. These proportions can be measured by drawing a vertical line in the survival plot (see dotted line in fig. 1), and these effectiveness measures may be called vertical measures. An alternative is to use horizontal effect measures, i.e. to measure along a horizontal line that divides the patient groups in two parts. Such a measure is the gain in median survival time (the difference in survival time for the first 50% who die in the control group and in the intervention group). If the median survival time is 30 months in a control group, and 35 months in the intervention group, then the average gain in median survival time (the difference in survival measures neither in a clinical nor in a scientific setting.

The problem with the above-mentioned studies on persuasiveness of the different risk formats is that they do not examine how the risk reduction formats were understood. The NNT format has been claimed to be easily understandable and promoted in evidence based medicine (Riegel- man & Schroth 1993; McAlister et al. 2000; Sackett et al. 2002). However, only a limited number of studies have addressed how NNT is understood.

Based on a MEDLINE search from 1970 to June 2005 using “NNT” as a search term combined with going over the reference list of relevant papers, we found seven studies using a randomised design to survey the understanding of NNT among medical doctors (Nexo et al. 2002; Halvorsen et al. 2003), patients (Sheridan & Pigmone 2003), medical students (Sheridan et al. 2002) and layman (Kristiansen et al. 2002; Christensen et al. 2003; Halvorsen & Kristiansen 2005) have been published.

Nexo et al. (2002) mailed a questionnaire to 1,500 Danish general practitioners presenting them with a scenario concerning a hypothetical drug that could prevent premature death. The authors concluded that in order to advise pa-
tients, doctors need to take all available measures of risk reductions into consideration. Halvorsen et al. (2003) reviewed a representative sample of Norwegian doctors (n=1,616) using a postal questionnaire. The doctors were presented with the same clinical scenario as the study mentioned above (Nexo et al. 2002). The authors conclude that medical doctors were sensitive to magnitude of NNT but that many doctors seem to misinterpret NNT.

Kristiansen et al. (2002) undertook a study in which a random sample of 675 non-institutionalized Danes aged 20–74 years were interviewed face-to-face and were offered a hypothetical therapy for heart attack. The respondents were randomized to receive the benefit of the drug as NNT of 10, 25, 50, 100, 200 and 400 and the proportions consenting to therapy were high, and surprisingly similar for all levels of NNT; namely 83%, 87%, 85%, 85%, 81% and 74%, respectively. In a study with a similar design (Christensen et al. 2003), the respondents were offered a hypothetical osteoporosis intervention with NNT also in the range from 10 to 400. Both studies concluded that lay people were insensitive to differences in NNT, which might be consistent with the hypothesis that lay people have difficulties in understanding NNT. This is supported by the finding that almost 43% indicated that they felt uncertain about the meaning of NNT (Christensen et al. 2003).

Sheridan et al. (2003) undertook a randomized, cross-sectional survey of 350 patients at a university internal medicine clinic. The authors concluded that NNT is often misinterpreted by patients and should not be used alone to communicate risks to patients.

In another study by Sheridan & Pignone (2002), 62 first-year medical students' ability to interpret effectiveness data was studied. Randomized to effect format, the students were asked to calculate how much one drug reduces risk of disease. The authors concluded that NNT makes students less able to undertake correct risk calculations than do ARR and RRR.

Recently Halvorsen & Kristiansen (2005) describe another randomized trial in which a representative sample of the Norwegian population (n=1,201) was allocated to scenarios with random combinations of a disease to be prevented, drug treatment costs and effect size in terms of NNT. They concluded that lay people's decisions were influenced by the type of disease to be prevented and the cost of the intervention, but not by the effect size in terms of NNT.

All of the above-mentioned studies indicate that doctors and lay people vast may have difficulties in understanding risk reduction including NNT. This is not surprising as the literature on numeracy skills indicate that such skills are often limited. For instance, it has been found that death rates of 1,286 out of 10,000 were found more risky than 24.14 out of 100 by 52 undergraduate students (Yamagishi 1997). Another study examined whether risks were better understood when presented as rates (e.g. 2.6 versus 8.9 per 1000) or proportions (e.g. 1 in 112 versus 1 in 384) in a survey of approximately 650 patients (Grimes & Snively 1999). Both formats were difficult for the patients to understand; but 73% correctly stated that 8.9 per 1,000 was a greater risk than 2.6 per 1,000, whereas 56% correctly stated that 1 in 112 is a higher risk than 1 in 384 (Grimes & Snively 1999).

To understand effectiveness in terms of postponement of the undesirable outcome, this study indicates that lay people seem to be able to discriminate between different effects when presented in a simple postponement format (Christensen et al. 2003). When using more complex formats, the perception of postponement remains unknown. Whether the ability to discriminate between different postponement formats can be interpreted as understanding of the outcome also remains to be shown.

**Concluding remarks**

Why NNT has been adopted with such optimism, one can only guess. Perhaps the name of the metric: "number-needed-to-treat" may seem attractive for a clinician investigating patients each day. One may only speculate whether a more correct abbreviation such as "RARR-APIT" (Reciprocal Absolute Risk Reduction – Arbitrary Point In Time) would become nearly as popular as NNT is.

The problems and limitations of NNT are not unique to this metric. All metrics that capture effect at a single point in time (ARR, NNT, RRR, odds ratio, etc.) have the same time-dependent limitations, and there is no evidence that any of them help people better than NNT in making clinical decisions. Postponement of adverse event has been proposed as an alternative to single point measures as NNT and RRR (Christensen et al. 2002). Everyone is used to the concept of time and have experience in distinguishing between for example one month and one year of duration. One would therefore expect that people understand what is meant by one month of life extension as opposed to for example one year. Indeed, in a study of layman's perception of osteoporosis therapies, the respondents were sensitive to the magnitude of the postponement when considering a hypothetical therapy (Christensen et al. 2003). However, so far there is no direct evidence that postponement of adverse event really help people make better than NNT or other metrics.

In conclusion, NNT poses important pitfalls, and is not easily understood. We suggest that NNT be treated with considerable care. There is probably no single effect measure that is able to convey all necessary information.

**References**


