ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 444−450

Journal of Clinical Epidemiology

# Considerable variation in NNT - A study based on Monte Carlo simulations

Torbjørn Wisløff[a,b,*], Odd O. Aalen[c], Ivar S. Kristiansen[b,d]

[a]*The Norwegian Knowledge Centre for the Health Services, Oslo, Norway*
[b]*Institute of Health Management and Health Economics, University of Oslo, Oslo, Norway*
[c]*Institute of Basic Medical Sciences, Department of Biostatistics, University of Oslo, Oslo, Norway*
[d]*Institute of Public Health, University of Southern Denmark, Odense, Denmark*

Accepted 22 April 2010

## Abstract

**Objective:** The aim of this analysis was to explore the variation in measures of effect, such as the number-needed-to-treat (NNT) and the relative risk (RR).

**Study Design and Setting:** We performed Monte Carlo simulations of therapies using binominal distributions based on different true absolute risk reductions (ARR), number of patients ($n$), and the baseline risk of adverse events ($p_0$) as parameters and presented results in histograms with NNT and RR. We also estimated the probability of observing no or a negative treatment effect, given that the true effect is positive.

**Results:** When RR is used to express treatment effectiveness, it has a regular distribution around the expected value for various values of true ARR, $n$, and $p_0$. The equivalent distribution of NNT is by definition nonconnected at zero and is also irregular. The probability that the observed treatment effectiveness is zero or negative when the true value is positive depends on $n$, $p_0$, and the true ARR. In some cases, this probability is even higher than 50%.

**Conclusion:** For realistic values of true ARR, $n$, and $p_0$, the observed NNT varies much more than the observed ARR and RR. Clinicians should use NNT cautiously when expressing treatment benefits. 

*Keywords:* Number-needed-to-treat; Binomial distribution; Relative risk; Treatment effectiveness; Opposite results; Absolute risk reduction

## 1. Introduction

Defined as the reciprocal of the absolute risk reduction (ARR), the number-needed-to-treat (NNT) has been increasingly used in the medical literature and even advocated through the CONSORT statement [1]. In most cases, NNT is reported as a single value although a confidence interval is sometimes included [2]. An estimated NNT of, for example, 48 has been interpreted in the following manner: "only 1 of 48 patients will benefit from treatment," which is a clear oversimplification [3]. NNT is known to have several undesirable statistical and other properties [4−11]. A common way of reporting NNT is: "we would have to treat seven patients to avoid one extra death" [12]. Such interpretation of NNT is at best imprecise and may be misleading in the context of chronic diseases. Interpretations like those previously mentioned could also leave the impression that there is no uncertainty connected to the effect estimate, which obviously is not the case. When the confidence interval of ARR covers zero, the confidence interval of NNT is nonconnected because NNT is not defined when ARR is zero. A proposed way of displaying the confidence interval of NNT is then "NNTB 13.0 to ∞ to NNTH 119" where NNTB means NNT(benefit) and NNTH means NNT(harm) [5].

If the outcome of interest (e.g., death) is unavoidable in the long run, adverse outcomes are not truly avoided but only postponed. NNT in its usual form does not fully take this into account. There have however been proposed alternatives, such as reporting NNT for specific time periods or calculating NNT from survival models [13−15]. To not report the time in relation to NNT values when time is a crucial factor is not recommended, compared with relative risk (RR), which may be similar over time [14].

The objective of this study was to explore the random variation of the observed number of patients who seem to benefit from a therapy when NNT is known with certainty. In practice, NNT is not known with certainty, and this raises the issue of estimation of confidence intervals for NNT, but this issue will not be addressed here.

* Corresponding author. Institute of Health Management and Health Economics, University of Oslo, Oslo, Norway. Tel.: (+47) 92867976; fax: (+47) 23255010.

*E-mail address*: tow@nokc.no (T. Wisløff).

0895-4356/$ - see front matter 
doi: 10.1016/j.jclinepi.2010.04.025

**What is new?**

What this article adds:

- The observed number of health events varies greatly because of chance. Hence, doctors should be aware that it is not always the case that what is observed in clinical practice represents the underlying "truth."
- Observed number-needed-to-treat (NNT) does not have a regular distribution around the true value of the treatment effect, as is the case for relative risk (RR).
- Baseline risk is important for the interpretation of NNT and not only for RR.
- The use of single NNT values may mislead patients to think that there is no uncertainty. Hence, the random aspect of NNT should be emphasized when informing patients.
- NNT is a measure of effect that should be used with caution.

## 2. Background and notation

In an intervention study, the results can be presented as in Table 1.

If $p_0 = P$(adverse event given control) and $p_1 = P$(adverse event given treatment), then $p_0$ and $p_1$ (the true proportions) can be estimated by

$$\widehat{p}_0 = \frac{X_0}{X_0 + Y_0} \text{ and } \widehat{p}_1 = \frac{X_1}{X_1 + Y_1}.$$

Throughout this article, we will handle problems where the treatment always has a true positive effect compared with the control group (i.e., $p_0 > p_1$). The NNT is given as the reciprocal of the ARR, defined as $p_0 - p_1$. Hence

$$\text{True } NNT = \frac{1}{p_0 - p_1}.$$

The RR is defined as $p_1/p_0$.

$X_0$ and $X_1$ can be seen as two binomial random variables, where $n$ is the total number of trials, which in this case corresponds to $X_0 + Y_0$ or $X_1 + Y_1$. Hence, we assume that equal number of patients in each arm of the trial and total sample size is then $n + n$.

To denote the absolute risk of an adverse event in the control group ($p_0$), we use the term "baseline risk" throughout this article. We use NNT (and RR and ARR) to denote the true (but usually not known with certainty) values. We use the term "observed NNT" to denote the observed outcome when patients undergo treatment with a known NNT. The observed NNT will depend on randomness and may not be equal to the true NNT. This *observed NNT* is conceptually different from the *estimated NNT*, which is calculated on the basis of specific studies.

Table 1
Outcomes of a typical trial

| Intervention | Number of adverse events | Number of nonadverse events |
|---|---|---|
| Treatment | $X_1$ | $Y_1$ |
| Control | $X_0$ | $Y_0$ |

The aim of this study was to explore variation in the observed NNT because of random variation in outcomes and compare this variation with similar variation in the observed RR. We will do this for different values of the true effect, baseline risk, and number of patients treated ($n$).

## 3. Material and methods

To perform Monte Carlo simulations of ranges of NNT that are relevant to clinical practice, we searched Medline (PubMed) for randomized controlled trials (RCTs) with the terms "number-needed-to-treat" or "NNT" mentioned in the title or in the abstract. We searched on December 18, 2008 with no boundaries back in time. We found 305 hits, of which 56 were not original RCTs or not clinical trials reporting NNT and baseline risk. The remaining 249 abstracts were not only searched for values of NNT but also for $p_0$ and $n$. In cases where not all three values were available in the abstract, we ordered the article to extract the remaining values. In articles with several reported NNTs, we chose NNT for the primary outcome. If there were more reported NNTs for the primary outcome, we chose the first NNT mentioned. We did not exclude trials on the basis of quality, and we used the reported statistics whether they were based on intention-to-treat analysis or not. Our search indicates a steady increase in the use of NNT during the period 1995–2002. In the years after 2002, the increase in use of NNT seems to have leveled off.

We simulated observed NNTs based on different values of the true NNT, $n$, and $p_0$, assuming the number of patients in each arm of the trials was equal. Based on the binomial distribution, we performed Monte Carlo simulations of number of adverse outcomes in treatment and control groups. From these numbers of adverse outcomes, we calculated the observed NNT and RR for each iteration and plotted them in histograms to evaluate the variation in NNT and RR. For each set of true NNT, $n$, and $p_0$, we noted the probability of observing no effect or a negative effect, given that the true treatment effect is positive.

To demonstrate that this probability is not negligible, it is subjected to a Monte Carlo analysis, illustrated by a series of graphs.