Number needed to treat (NNT): estimation of a measure of clinical benefit†

S. D. Walter*,†

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, L8N 3Z5

SUMMARY

The number needed to treat (NNT) is becoming increasingly popular as an index for reporting the results of randomized trials and other clinical studies. It represents the expected number of patients who must be treated with an experimental therapy in order to prevent one additional adverse outcome event (or, depending on the context, to expect one additional beneficial outcome), compared to the expected event rates under the control therapy. Although NNT is a clinically useful measure, little work has been done on its statistical properties. In this paper, alternative NNT-type measures are defined for use with discrete or continuous data. Estimators and their variances are obtained for these measures in cross-over or parallel group designs. The ideas are illustrated with data on quality of life in asthma patients. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

In clinical trials and other biomedical studies, results are often expressed in terms of the proportions $\pi_1$ and $\pi_2$ on a control or experimental treatment (respectively) who experience an outcome event, such as death. Traditionally, the comparison of the treatment benefit relative to control has been summarized using the relative risk $RR = \pi_2/\pi_1$, the absolute risk reduction $\pi_1 - \pi_2$, the odds ratio $\pi_2/(1 - \pi_2)/[\pi_1/(1 - \pi_1)]$, or the relative risk reduction $(\pi_1 - \pi_2)/\pi_1$ [1]. More recently, the number needed to treat (NNT) measure has been proposed [2–7]. NNT has been defined as

$$NNT = \frac{1}{\pi_1 - \pi_2} \quad (1)$$

NNT represents the expected number of patients who must be placed on the experimental treatment in order to prevent one adverse event. If, rather than reducing deleterious outcomes,
the goal of the therapeutic intervention is to improve the rate of beneficial outcomes (for example, successful pregnancy, or improvement in quality of life), then the roles of $\pi_1$ and $\pi_2$ are reversed, and NNT then represents the expected number of patients required in order to gain one beneficial outcome event.

As a numerical illustration, suppose the treatment leads to fewer adverse outcomes, with $\pi_1 = 0.4$ and $\pi_2 = 0.2$. Then $RR = 0.2/0.4 = 0.5$, and $NNT = 1/(0.4 - 0.2) = 5$. Thus one has to treat 5 patients to avoid an expected one additional adverse event, compared to the expected event rate in the control group. As a second example, let $\pi_1 = 0.02$ and $\pi_2 = 0.01$. Then $RR = 0.5$ as before, but now $NNT = 100$. Thus even though $RR$ would characterize the treatment benefit identically in both examples, in the second case only 1 in 100 patients might expect to achieve actual benefit from the treatment. This occurs because of the much lower event rates. Other features (for example, cost and side-effects) being equal, one would assess the second treatment as less valuable than first.

The expression of treatment benefit through NNT has considerable clinical utility [3, 8–10], and a nomogram and software for its calculation have been published [11–13]. Treatments may sometimes appear to have substantial benefit when using $RR$ or the odds ratio (such as in the previous examples, where $RR = 0.5$ would be considered a large treatment effect in most circumstances), but they may be seen in a different light when using NNT. The NNT measure is useful for comparing the benefits achieved by different interventions; other factors being equivalent, one would prefer treatments with lower NNT values. A different perspective also emerges when a mean difference between treatment groups for a continuous outcome is contrasted with an analysis of NNT [4, 10]. Finally, NNT is useful in economic analyses intended to calculate the cost of the total therapeutic effort required per unit gain (one adverse event prevented or one beneficial outcome added) in a series of patients.

NNT has usually been applied to data on deleterious outcomes such as stroke or death, but similar ideas can be applied to beneficial events such as improvement in physical function, remission of symptoms, return to work after rehabilitation, or improved quality of life. The same principle has also been suggested in a wide variety of other circumstances, including the following: (i) in screening, where the number needed to screen (NNS) [14] indicates how many individuals must be screened in order to prevent an expected one adverse outcome; (ii) in prevention, where the ‘Intervention Index’ [15] assesses the number of people whose risk exposure must change in order that an expected one adverse outcome is prevented; (iii) in clinical medicine, to evaluate the number of patients needed to be assessed and the number of missed opportunities to prevent an adverse outcome [6]; (iv) in a population context, where the disease impact number generalizes the concept of NNT to all persons with a given disease, and the population impact number generalizes NNT to all persons in a population [16]; (v) where the data are obtained as in the form of a time to event [17]; and (vi) in case-control studies [18]. Incorporation of different beneficial and adverse outcomes has also been considered [19].

Despite these diverse applications, rather little work has been done on the statistical properties of NNT. Applications have primarily been confined to binary data, and there has been little discussion of the sample variability of NNT estimates. In this paper, we develop estimators for NNT where the outcome variable is discrete or continuous, and estimation methods are proposed for the parallel group and cross-over designs. We show some illustrative examples using data on quality of life measures in studies of asthma.
2. METHODS – OVERVIEW

We note first from expression (1) that NNT is a function of only the absolute risk reduction \( \pi_1 - \pi_2 \). Hence in order to estimate NNT, it is sufficient to obtain an estimate of \( \pi_1 - \pi_2 \) and then apply the reciprocal transformation. This idea has been adopted by Altman [5] and Daly [20] to construct confidence intervals (CIs) for NNT. In this context, \( \pi_1 - \pi_2 \) essentially measures the benefit of treatment relative to control, with benefit being assessed according to the proportions of occurrences of the outcome event. We will also develop analogous ideas when the outcome is a continuous variable.

In cross-over studies, one can directly observe the responses to treatment or control in the same subjects. This is a useful approach for studies such as on quality of life in asthmatics, where subjects can be randomly assigned to one treatment, and subsequently transferred to the other. Of course, the cross-over design is not appropriate with outcomes such as death, or where transfer between treatments is not possible (for example, surgical studies). For situations where only one treatment can be applied to a given patient, the within-subject joint distribution of responses under treatment and control is unobservable. Only separate marginal data on the responses to each treatment can then be obtained. This places strictures on the estimation of NNT, as discussed later.

We now consider the estimation of NNT for the various scenarios in turn, involving discrete or continuous data, and in the context of cross-over or parallel group designs.

3. ESTIMATION OF NNT WITH DISCRETE DATA

3.1. Cross-over design

Data from a two-period cross-over study with a binary outcome can be displayed in a \( 2 \times 2 \) table with an underlying probability distribution as shown in Table I. Reflecting the nature of the numerical example to be discussed in Section 5, we will now take the goal of treatment to be an increase in the rate of beneficial outcomes, relative to the outcome rates for patients in the control group. Thus we will be assuming that \( \pi_2 \geq \pi_1 \). The proportion \( \pi_1 \) who have beneficial outcomes in the control under the treatment is \( \lambda_1 \), and the corresponding proportion \( \pi_2 \) in the treatment group is \( \lambda_1 \) (where ‘.’ indicates summation). We denote the net proportion of subjects who benefit under treatment relative to control by \( \theta \), which is therefore \( \theta = \lambda_1 - \lambda_1 = \lambda_{21} - \lambda_{12} \). In this context \( \theta \) is equivalent to \( \pi_2 - \pi_1 \), the absolute risk reduction.

<table>
<thead>
<tr>
<th>Control</th>
<th>Treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improved</td>
<td>Not improved</td>
</tr>
<tr>
<td>Improved</td>
<td>( \lambda_{11} )</td>
<td>( \lambda_{12} )</td>
</tr>
<tr>
<td>Not improved</td>
<td>( \lambda_{21} )</td>
<td>( \lambda_{22} )</td>
</tr>
</tbody>
</table>
If there are \( N \) subjects (each with a pair of responses), and the sample data have frequencies \( n_{ij} \) corresponding to \( \hat{\lambda}_{ij} \), then

\[
\hat{\theta}_{CO} = \hat{\lambda}_{21} - \hat{\lambda}_{12}
\]  

(2)

where \( \hat{\lambda}_{ij} = n_{ij}/N \), is an unbiased estimator of \( \theta \) from the cross-over design. For the variance, we use the multinomial relationship of \( n_{12} \) and \( n_{21} \) to obtain

\[
\text{var}(\hat{\theta}_{CO}) = \text{var}(\hat{\lambda}_{21}) + \text{var}(\hat{\lambda}_{12}) - 2 \text{cov}(\hat{\lambda}_{12}, \hat{\lambda}_{21})
\]

\[
= [\hat{\lambda}_{21}(1 - \hat{\lambda}_{21}) + \hat{\lambda}_{12}(1 - \hat{\lambda}_{12}) + 2\hat{\lambda}_{12}\hat{\lambda}_{21}]/N
\]

\[
= [\hat{\lambda}_{21} + \hat{\lambda}_{12} - (\hat{\lambda}_{21} - \hat{\lambda}_{12})^2]/N
\]  

(3)

An approximate 95 per cent confidence interval (CI), \((\hat{\theta}_{L}, \hat{\theta}_{U})\), can thus be obtained as \( \hat{\theta}_{CO} \pm 1.96 \text{SE}(\hat{\theta}_{CO}) \), and one can then use the reciprocal transformation to obtain a corresponding interval for NNT. Alternatively [19], one can use the approximation \( \text{var}(N\hat{\theta}_N) = \text{var}(\hat{\theta})/\hat{\theta}^4 \).

Altman [5] shows how, when the CI for \( \hat{\theta} \) includes the value 0, it has two components; the first component \((0, \hat{\theta}_{U})\) corresponds to potential benefit from the treatment, and the second component \((\hat{\theta}_{L}, 0)\) corresponds to potential harm from treatment, both relative to the event rate in the control group. The corresponding components for NNT are \((N\hat{\theta}_{NU}, \infty)\) and \((\infty, N\hat{\theta}_{NL})\), respectively, where \( N\hat{\theta}_{NU} = 1/\hat{\theta}_{U} \) and \( N\hat{\theta}_{NL} = -1/\hat{\theta}_{L} \).

### 3.2. Parallel group design

In the parallel group design, suppose that \( N_1 \) subjects are randomized to the control and \( N_2 \) to the treatment group. As before, we denote the proportions of subjects who improve by \( \pi_1 \) and \( \pi_2 \). In this case the estimator \( \hat{\theta}_{PG} \) is simply a difference of two independent binomial variables, so

\[
\hat{\theta}_{PG} = \hat{\pi}_2 - \hat{\pi}_1
\]  

(4)

with

\[
\text{var}(\hat{\theta}_{PG}) = \frac{\pi_2(1 - \pi_2)}{N_2} + \frac{\pi_1(1 - \pi_1)}{N_1}
\]  

(5)

Note that if we take \( \pi_1 = \hat{\lambda}_{i,} \) and \( \pi_2 = \hat{\lambda}_{i,} \) from the cross-over design, and assume \( N_1 = N_2 = N \), then we have from (2) and (4) that

\[
\hat{\theta}_{PG} = \hat{\lambda}_{i,1} - \hat{\lambda}_{i,2} = \hat{\lambda}_{21} - \hat{\lambda}_{12} = \hat{\theta}_{CO}
\]  

(6)

so in this case the parallel group design provides the same estimator as would have been obtained if within-subject data had been available.
4. ESTIMATION OF NNT WITH CONTINUOUS DATA

4.1. Cross-over design

We now consider the continuous data version of NNT, by examining $\theta = \Pr(Y > X + c)$, where $Y$ and $X$ are the treatment and control responses of a given subject and $c$ is a difference that has been declared to be clinically important on a continuous outcome scale. (Again, to reflect the situation in the numerical example to be discussed later, we will assume that higher response values correspond to desirable outcomes.) The value of $c$ is known as the ‘minimally important difference’ (MID) or the ‘minimal clinically important difference’ [10, 21, 22].

With cross-over data, $\theta$ can be estimated empirically, without making distributional assumptions about $Y$ and $X$. Let $n$ be the number of subjects for which $Y > X + c$, and let $N$ be the total number of patients, as before. Then

$$\hat{\theta}_{CO} = \frac{n}{N} \quad (7)$$

is a suitable estimator of $\theta$, with

$$\text{var}(\hat{\theta}_{CO}) = \theta(1 - \theta)/N \quad (8)$$

An alternative approach is to consider the joint probability distribution function (PDF) of $(X, Y)$. As we will see later, this allows greater insight into the behaviour of $\hat{\theta}$ and provides a basis for comparing the cross-over design with the parallel group design (in which distributional assumptions are required anyway). Let $h(x, y)$ be the joint PDF of $(X, Y)$. Then

$$\theta = \int_{-\infty}^{\infty} \int_{X+c}^{\infty} h(x, y) \, dy \, dx \quad (9)$$

This bivariate area of integration for $\theta$ is above the line AB (defined by the equation $Y = X + c$) shown in Figure 1.

![Figure 1. Joint distribution of responses under treatment ($Y$) and control ($X$) and identification of area of relative benefit from treatment (the region above the line AB corresponds to subjects who benefit by more than the minimally important difference $c$).](image)
If we now make the additional assumption that \((X, Y)\) have a bivariate normal distribution, then, as shown in detail in the Appendix, we can most conveniently express \(\theta\) in terms of the standard normal distribution. Specifically we obtain

\[
\theta = \Phi \left( \frac{\mu_2 - \mu_1 - c}{\left( \sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2 \right)^{1/2}} \right) \tag{10}
\]

where \(\Phi(.)\) denotes the standard normal CDF, \(\mu_1(\mu_2)\) and \(\sigma_1(\sigma_2)\) are the mean and standard deviation of \(X(Y)\), and \(\rho\) is the correlation of \(X \) and \(Y\). An estimate \(\hat{\theta}_{CO}\) is obtained by substituting the usual estimates of \(\mu_1, \mu_2, \sigma_1, \sigma_2\) and \(\rho\) into (10).

Note that \(\theta\) is an increasing function of \(\mu_2 - \mu_1\), so, as expected intuitively, a larger net proportion of subjects benefits when the mean effect of treatment is large. Also, \(\theta\) is a decreasing function of \(c\), so a smaller proportion of subjects benefits when a stringent definition (that is, large \(c\) value) of a clinically important difference is adopted. Finally, \(\theta\) is an increasing function of \(\rho\) if \(\mu_2 - \mu_1 - c > 0\), and a decreasing function of \(\rho\) otherwise. Thus studies in which the responses on treatment and control are highly correlated and where the group mean difference exceeds the MID will tend to indicate greater benefit (in terms of smaller NNT) for given values of \(\sigma_1\) and \(\sigma_2\).

As is also shown in more detail in the Appendix, \(\text{var}(\hat{\theta}_{CO})\) may be obtained by using Taylor series expansions of (10) to obtain

\[
\text{var}(\hat{\theta}_{CO}) = \phi^2 \left( \frac{\mu_2 - \mu_1 - c}{\left( \sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2 \right)^{1/2}} \right) \left( 1 + \frac{[\mu_2 - \mu_1 - c]^2}{2[\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2]} \right) / N \tag{11}
\]

where \(\phi(.)\) is the standard normal PDF. In the special case \(\sigma_1 = \sigma_2 = \sigma\), we have from (10) and (11) that

\[
\theta_{CO} = \Phi \left( \frac{\mu_2 - \mu_1 - c}{\sigma[2(1 - \rho)]^{1/2}} \right)
\]

and

\[
\text{var}(\hat{\theta}_{CO}) = \phi^2 \left( \frac{\mu_2 - \mu_1 - c}{\sigma[2(1 - \rho)]^{1/2}} \right) \left( 1 + \frac{(\mu_2 - \mu_1 - c)^2}{4\sigma^2(1 - \rho)} \right) / N \tag{12}
\]

4.2. Parallel group design

In the parallel group design, we obtain independent estimates of \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\) from the two treatment groups; however, in contrast to the cross-over design, the within-patient correlation \(\rho\) is not estimable because different patients are observed in the two groups. We assume, as before, that independent samples of size \(N_1\) and \(N_2\) are observed, and substitute estimates of \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\) into (11) to estimate \(\theta_{PG}\). Lacking any other information, we may use a variety of different assumed values of \(\rho\), and we can then investigate the sensitivity of \(\hat{\theta}\) to the unknown correlation value.
After some derivation (see Appendix for details), we obtain

\[
\text{var}(\hat{\theta}_{\text{PG}}) = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \phi^2 \left( \frac{[\mu_2 - \mu_1 - c]}{[\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2]^{1/2}} \right) \times \left[ \left( \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \right) + \frac{(\mu_2 - \mu_1 - c)^2}{2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} \right] \left( \frac{\sigma_1^2}{N_1} (\sigma_1 - \rho\sigma_2)^2 + \frac{\sigma_2^2}{N_2} (\sigma_2 - \rho\sigma_1)^2 \right)
\]

(13)

In the special case \( \sigma_1 = \sigma_2 = \sigma \) and \( N_1 = N_2 = N \), this result simplifies to

\[
\text{var}(\hat{\theta}_{\text{PG}}) = \frac{1}{(1 - \rho)} \phi^2 \left( \frac{\mu_2 - \mu_1 - c}{\sigma[2(1 - \rho)]^{1/2}} \right) \left[ 1 + \frac{(\mu_2 - \mu_1 - c)^2}{8\sigma^2} \right] / N
\]

(14)

5. EXAMPLE

To illustrate the various methods for estimating \( \theta \) and NNT, we will use data from a randomized double-blind cross-over study to compare two treatments for asthma, salmeterol (the experimental treatment) and salbutamol (the standard treatment, to be referred to here as the control) [4, 23]. Subjects reported their quality of life (QL) during the last 14 days of each treatment, using a 32-item questionnaire relating to five domains: symptoms; activity limitations; emotional function; environment exposures, and sleep. Each item was scored on a seven-point Likert scale, and an overall QL score was computed as an average of the five domain scores. Baseline measurements after treatment on a blinded placebo were also available. Higher scores correspond to better QL outcomes. The minimal important difference (MID) of clinical interest on this scale had previously been demonstrated to be \( c = 0.5 \) [4, 21]. Very high reliability in the QL scores had been demonstrated previously; the intraclass correlation was 0.92 for the overall QL measure used here, and similarly high for its component domains [24].

For the purpose of this example, we will analyse the total QL scores on the treatment and control relative to the baseline levels. We will first use the data in their original cross-over format (that is, using the paired responses of each participant), and then use their marginal distributions, for illustration of NNT estimation in parallel group studies. Analysing the same data in these two ways permits a convenient comparison of the two designs. In practice, of course, one would analyse cross-over data with only a cross-over analysis.

Table II shows the paired responses on treatment and control from the 139 study participants. These data are in binary format, indicating whether or not the subject experienced improvement in QL that exceeded the MID. The results show that 61/139 (44 per cent) of patients in the treated group and 34/139 (24 per cent) in the control groups experienced such improvement. Hence \( \theta \), the difference in these response rates, is 19.4 per cent. The cross-over and parallel group estimates of \( \theta \) are shown in Table III, together with their confidence intervals and corresponding estimates and confidence intervals for NNT. The cross-over estimates use the \( 2 \times 2 \) cell frequencies with equations (2) and (3), while the parallel group estimates use only the marginal data in (4) and (5). The confidence limits for \( \theta \) are computed using the SEs and large sample approximations; corresponding limits for NNT are obtained after
Table II. Results of a cross-over study of asthma therapy: binary outcome variable.

<table>
<thead>
<tr>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Improved</td>
<td>27</td>
</tr>
<tr>
<td>Not improved</td>
<td>34</td>
</tr>
</tbody>
</table>

Table III. Estimates of proportion of subjects who benefit ($\theta$) and number needed to treat (NNT): discrete data analysis.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>$\theta$</th>
<th>SE($\theta$)</th>
<th>95 per cent CI</th>
<th>NNT</th>
<th>95 per cent CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-over</td>
<td>0.194</td>
<td>0.043</td>
<td>0.11, 0.28</td>
<td>5.15</td>
<td>3.59, 9.10</td>
</tr>
<tr>
<td>Parallel group</td>
<td>0.194</td>
<td>0.056</td>
<td>0.09, 0.30</td>
<td>5.15</td>
<td>3.30, 11.76</td>
</tr>
</tbody>
</table>

Reciprocal. Because of the reciprocation, the estimate of the upper confidence limit for NNT is more variable than the lower limit. As expected from (6), the point estimates from the two designs are identical, but the SE of the parallel group estimator is larger.

Note that the use of SE from (5) reflects the usual assumption made in the analysis of parallel group studies that the response rates in the treated and control groups are independently estimated. However, at the level of individual patients in a cross-over study, responses to the two interventions are actually correlated. The magnitude of the correlation cannot be estimated with parallel group data, but it is in effect assumed to be 0 in the standard analysis. The difference in the standard errors of $\theta$ between the parallel group and cross-over designs is partly explained by this correlation. For comparison with the continuous data analysis that follows, note that the value of the correlation $\rho$ as calculated from the data in the binary format was 0.407.

Figure 2 shows the data in their original paired continuous format. The distribution appears to be unimodal and reasonably symmetric in both treatment groups. The estimates of the means, variances and correlation are $\hat{\mu}_1 = 0.128$, $\hat{\sigma}_1^2 = 0.481$, $\hat{\mu}_2 = 0.497$, $\hat{\sigma}_2^2 = 0.689$ and $\hat{\rho} = 0.402$.

Based on these numerical estimates, we may use equations (10) and (11) for the cross-over estimates and their SEs: the results are shown in Table IV. For the parallel group design we use (10) and (13) with a range of assumed values for $\rho$, including the observed value, which in the continuous data format takes the value 0.402. Note that $\hat{\theta}_{PG}$ is rather insensitive to the assumed value of $\rho$, until $\rho$ becomes relatively large. Consequently, the estimates from the cross-over and parallel group designs are reasonably similar. (In particular, the parallel group estimates using $\rho = 0.402$ equal the cross-over estimates). The variance of $\hat{\theta}_{PG}$ progressively increases with $\rho$.

Additional estimates of $\theta$ are shown under the assumption that $\sigma_1 = \sigma_2 = \sigma$, with $\hat{\sigma}^2 = (\hat{\sigma}_1^2 + \hat{\sigma}_2^2)/2$, and using $\rho = 0.402$ for the parallel group design. These differ very little from the general estimates, presumably because $\sigma_1 \simeq \sigma_2$ in these data.

All of the previous estimates of $\theta$ from the continuous data are based on the assumption of normally distributed data. In the context of the cross-over design one can instead use...
Figure 2. Distribution of paired responses to treatment and control in a cross-over trial of asthma therapy.

the numbers of subjects who have treatment minus control differences $\geq c$, to evaluate the empirical estimate and its SE, from (7) and (8). The numbers of subjects in this category in the asthma study with $c = 0.5$ was 53, leading to the empirical estimates shown in Table IV.

6. DISCUSSION

In the numerical example, we found that the cross-over and parallel group estimates of $\theta$ were reasonably close with either discrete or continuous data analyses, but their standard errors were somewhat smaller for the cross-over design. These conclusions may not always pertain. The close agreement of the discrete and continuous data methods might here be due to the symmetry in the data (see Figure 2) and the similar values of $\rho$ in both formulations (0.407 and 0.402 in the discrete and continuous analyses), conditions which may not occur in other data. In general, one should be aware of the potential loss of information caused by grouping a continuous variable into discrete categories.

In practice, one would presumably use only a cross-over estimator from a cross-over design, and adopt the discrete or continuous methods to correspond to the type of data observed. In a parallel group design, the value of $\rho$ remains unobserved, so one can only condition on its value and inspect the sensitivity of $\hat{\theta}$ to that value (see Table IV). In the example used here, we found only small effects on $\hat{\theta}$ and its SE unless a large value of $\rho$ is involved.
Table IV. Estimates of proportion of subjects who benefit (θ) and number needed to treat (NNT): continuous data analysis.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>θ</th>
<th>SE(θ)</th>
<th>95 per cent CI</th>
<th>NNT</th>
<th>95 per cent CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-over*</td>
<td>0.438</td>
<td>0.034</td>
<td>0.37, 0.50</td>
<td>2.28</td>
<td>1.98, 2.68</td>
</tr>
<tr>
<td>assuming equal variances</td>
<td>0.438</td>
<td>0.034</td>
<td>0.37, 0.50</td>
<td>2.28</td>
<td>1.98, 2.69</td>
</tr>
<tr>
<td>Parallel group:*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ = 0</td>
<td>0.452</td>
<td>0.034</td>
<td>0.39, 0.52</td>
<td>2.21</td>
<td>1.93, 2.59</td>
</tr>
<tr>
<td>ρ = 0.2</td>
<td>0.446</td>
<td>0.038</td>
<td>0.37, 0.52</td>
<td>2.24</td>
<td>1.92, 2.68</td>
</tr>
<tr>
<td>ρ = 0.4</td>
<td>0.438</td>
<td>0.043</td>
<td>0.35, 0.52</td>
<td>2.28</td>
<td>1.91, 2.82</td>
</tr>
<tr>
<td>ρ = 0.402</td>
<td>0.438</td>
<td>0.043</td>
<td>0.35, 0.52</td>
<td>2.28</td>
<td>1.91, 2.83</td>
</tr>
<tr>
<td>ρ = 0.6</td>
<td>0.425</td>
<td>0.052</td>
<td>0.32, 0.53</td>
<td>2.35</td>
<td>1.90, 3.10</td>
</tr>
<tr>
<td>ρ = 0.8</td>
<td>0.397</td>
<td>0.071</td>
<td>0.26, 0.54</td>
<td>2.52</td>
<td>1.87, 3.89</td>
</tr>
<tr>
<td>assuming equal variances†</td>
<td>0.438</td>
<td>0.043</td>
<td>0.35, 0.52</td>
<td>2.28</td>
<td>1.91, 2.83</td>
</tr>
<tr>
<td>Empirical‡</td>
<td>0.381</td>
<td>0.041</td>
<td>0.30, 0.46</td>
<td>2.62</td>
<td>2.17, 3.33</td>
</tr>
</tbody>
</table>

*Assuming data are normally distributed.
†Using observed value ρ = 0.402.
‡No distributional assumptions.

ρ = correlation between treatment and control responses.

Additional complications in the cross-over design include the possibility of carry-over and period effects. Appropriate adjustment for these effects would imply that potentially different estimates would then be obtained in cross-over and parallel group designs. However, note that the identity of the two estimators described in Section 3.2 still holds true if a given set of data is analysed in both ways, even if carry-over and period effects exist.

The empirical estimate of θ with continuous data was also reasonably close to the other estimates that involved the normal assumption, although its SE was slightly larger than the cross-over estimate. No general advantage can be claimed here, so the assumption of normality may or may not yield higher precision for ˆθ in a given data set.

Many of the estimated SEs and CIs have involved large sample approximations. Further work is needed to examine the adequacy of those approximations, especially as they might affect the confidence limits for NNT. Recent work by Lesaffre and Pledger [25] suggests that it is generally preferable to estimate θ and then to reciprocate the confidence limits to provide a CI for NNT, rather than using the sample distribution of NNT directly. Methods for data where the distribution of the treatment and control responses is other than the normal might also provide further insight into the properties of the estimators, although their practical utility will be limited by the non-observability of the joint distribution in the parallel group design.

Hutton [26] has shown that convergence to normality of the NNT estimator may be slow, especially if the absolute risk reduction is small or close to one. In fact, she comments that the variance of NNT is not formally defined, because of a singularity in the sample distribution when θ = 0. This is similar to the problem of potentially infinite values of the odds ratio or relative risk when zero cell frequencies are observed. In that context, the problem of singularities has been addressed by adding a small positive constant to each data cell, so that an infinite estimate cannot then arise. Further research is therefore suggested, to develop an analogous estimator of NNT which would similarly avoid its singularity. However, as indicated earlier, it does seem preferable to approach the problem of confidence interval construction by using the distribution of θ in the first place.
We have considered NNT estimation for a single study. As described through the numerical examples in the Introduction, NNT will tend to vary with the underlying event rate, with high NNT values associated with rarer outcomes. For this reason, it is generally inadvisable to consider to attempt any comparisons of NNT values between studies where the event rate may be expected to vary, such as in meta-analyses [27, 28]. Another typical situation where event rates might vary is when different lengths of follow-up have been used to measure the outcome; longer follow-up will tend to increase the event rate. Others have argued that use of NNT in meta-analysis is useful in certain circumstances, with appropriate caution [29].

The issue of baseline risk is also relevant within the context of single studies. Proponents of NNT have argued that it is a useful measure to communicate risk to clinicians, especially if the NNT can be regarded as constant over a spectrum of patients with different baseline risks. This situation arises if one can demonstrate that the relative risk is constant for patients at different baseline risk; this can be easily seen by re-expressing equation (1) as

$$\text{NNT} = \frac{1}{\pi_i (1 - \text{RR})}$$

As has been discussed elsewhere [30, 31], this assumption may or may not be correct in particular data sets, which has led to a debate concerning the most appropriate measure to summarize data from clinical trials and epidemiologic studies.

In response to Hutton’s criticism of the NNT index, Altman and Deeks [32] and Lesaffre and Pledger [33] have argued that the NNT is indeed a useful communication tool. These commentators are themselves very critical of Hutton’s negative stance on NNT, which they claim is largely based on an inappropriate discussion of the technical statistical difficulties, rather than focusing on NNT as a way of communicating among clinicians. Indeed, they indicate that NNT is ‘a way of presenting results, but not of analyzing data’ [32].

A reviewer of this paper has pointed out that measurement error in the outcome variable leads to particular concern with respect to the interpretation of NNT, specifically that the measurement error can lead to bias in the NNT estimate. This is in contrast to the situation where a continuous outcome measure is used; in that case, random measurement error does not bias the estimated treatment effect, although there will be a deleterious effect on its precision. We have described the phenomenon of bias in NNT induced by measurement error in more detail elsewhere, focusing on the impact of measurement error in practice [34]. In brief, the bias in NNT can in general be either positive or negative, but numerical results suggest that the bias may be limited to tolerable levels as long as the reliability of the measurement is at least 80 per cent. This condition was satisfied in the quality of life example presented here [24].

In contrast, if the reliability in the data is poor, the NNT will be substantially biased. If the liability of the data is unknown, then (as pointed out by the reviewer) one cannot sensibly determine what proportion of patients have indeed benefited at a certain level. If, for example, the mean difference between the treated and control groups is substantially greater than the MID, one cannot distinguish the possibilities that (i) there is no measurement error, and most treated patients have benefited by more than the MID, (ii) there is substantial measurement error, and only some small fraction of patients have truly benefited at this level, or (iii) an intermediate situation. A similar conclusion pertains if the mean treatment effect is less than the MID; in that case it might be that no single patient has benefited from treatment at the level of MID or greater, or some fraction of patients may indeed have benefited because of between-patient variability of response.

Cox et al. [35] discuss many of these issues. They approach the question by considering components of variance related to treatments, patients, treatment by patient interaction.
and random error. They point out that reliability in the data (in the sense of the relative magnitudes of between-patient variation to random variation) is dependent on the particular study in question, and the extent to which the sampled patients are heterogeneous or homogeneous. They also mention some of the practical difficulties in obtaining estimates of these variance components, in particular in distinguishing between treatment by patient interaction and random error.

These considerations lead to two main conclusions. First, it will likely be hazardous to draw conclusions at the level of individual patients based on NNT calculations. It appears preferable to regard NNT as a population parameter, representing the clinical effort that is required by treatment in order to prevent one adverse event, compared to the expected response in the control group. However, it is not an index through which one can make statements about which particular patients will or will not benefit from treatment. The second conclusion is that the findings of our earlier investigation [34] are reinforced, specifically that high reliability in the data is required in order to carry out NNT calculations without risk of serious bias. For data with only moderate reliability, an appropriate substudy might lead to suitable estimates of Cox’s variance components, for instance by subtracting the estimated variance for random error [35]. However, a more satisfactory solution is to strive for highly reliable data in the first place.

The reviewer has also commented that there may be a considerable loss of information associated with the dichotomization of individual observations originally made on a continuous scale. I agree with that sentiment, and would emphasize that if continuous data are available, then the corresponding form of NNT estimation (based on the continuous approach described in Section 4) should be adopted. That said, and once again recognizing that clinicians and statisticians may communicate in different ways, they may prefer to use different summaries of the data [32, 33]. Specifically, clinicians may prefer to report their data in discrete categories, defined by their chosen value of MID. Once again, the issue of data reliability is important here. If measurement error is substantial and the reliability is poor, then the categorization of the continuous data will be error-prone and the NNT estimate will be biased. In contrast, if the data are very reliable, then the categorization will be largely error-free, and the NNT estimates will be trustworthy.

Finally, recall that although this paper has focused on the use of NNT in the interpretation of benefits from clinical therapy, the same ideas can be applied in other areas such as screening, preventive intervention and case-control studies, for example. The NNT index we have discussed here becomes the NNS (number needed to screen) [14], the ‘Intervention index’ [15], and the NNTH (number needed to harm) [18], respectively, in these applications. With suitable modifications, the estimates and standard errors presented here carry over to these other situations without difficulty. Further extension of this work to incorporate different outcome measures of benefit and harm [19] or net benefit over harm [34] could also be contemplated.

APPENDIX: DERIVATIONS FOR NNT WITH CONTINUOUS VARIABLES

A1. Cross-over design

To evaluate \( \theta \) as given in (9), we first transform from \((X, Y)\) values to standardized variables

\[
z_1 = \frac{(X - \mu_1)}{\sigma_1} \quad \text{and} \quad z_2 = \frac{(Y - \mu_2)}{\sigma_2}
\]

(A1)
where \( \mu_1 = E(X) \), \( \mu_2 = E(Y) \), \( \sigma^2_1 = \text{var}(X) \) and \( \sigma^2_2 = \text{var}(Y) \). The line AB in Figure 1 becomes the line \( z_2 = (z_1 \sigma_1 + \mu_1 - \mu_2 + c)/\sigma_2 \). Next, we rotate the distribution about the origin in Figure 1 through an angle \( -\psi \), where \( \tan \psi = \sigma_1/\sigma_2 \) is the slope of the line AB in the \((z_1, z_2)\) co-ordinates. We denote the response co-ordinates after rotation by \((z'_1, z'_2)\). In the new co-ordinate system, the expression for \( \theta \) becomes

\[
\theta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h'(z'_1, z'_2) \, dz'_1 \, dz'_2
\]  

where \( h'(z'_1, z'_2) \) is the joint PDF of \((z'_1, z'_2)\), and \( z'_2(A) \) is the ordinate of the point A after rotation. Rotation through \( -\psi \) transforms \( z_2 \) to \( z'_2 \) according to

\[
z'_2 = z_2 \cos \psi - z_1 \sin \psi
\]

The \((z_1, z_2)\) co-ordinates of A are defined by \([z_1(A) \sigma_1 + \mu_1 - \mu_2 + c]/\sigma_2 = 0 \) and \( z_2(A) = 0 \). Hence

\[
z'_2(A) = (\mu_1 - \mu_2 + c)/(\sigma^2_1 + \sigma^2_2)^{1/2}
\]

To evaluate (A2) we will require \( \text{var}(z'_1) \). Using Taylor series on (A3) we may obtain the asymptotic variance

\[
\text{var}(z'_1) = \sin^2 \psi \, \text{var}(z_1) + \cos^2 \psi \, \text{var}(z_2) + 2 \cos \psi \sin \psi \, \text{cov}(z_1, z_2)
\]

Noting that the transformation (A1) yields \( \text{var}(z_1) = \text{var}(z_2) = 1 \), we find that

\[
\text{var}(z'_1) = 1 + 2 \cos \psi \sin \psi \, \rho
\]

where \( \rho \) is the correlation between \( Y \) and \( X \). Making the substitutions \( \sin \psi = -\sigma_1/(\sigma^2_1 + \sigma^2_2)^{1/2} \) and \( \cos \psi = \sigma_2/(\sigma^2_1 + \sigma^2_2)^{1/2} \) into (A5) gives

\[
\text{var}(z'_1) = \frac{\sigma^2_1 + \sigma^2_2 - 2 \rho \sigma_1 \sigma_2}{\sigma^2_1 + \sigma^2_2}
\]

When we assume that \((X, Y)\) have a bivariate normal distribution, then using results (A2), (A4) and (A6), we have that

\[
\theta = \Pr \left( \frac{z'_2}{(\sigma^2_1 + \sigma^2_2)^{1/2}} > -\frac{(\mu_2 - \mu_1 + c)}{(\sigma^2_1 + \sigma^2_2)^{1/2}} \right)
\]

which leads to result (10).

To obtain \( \text{var}(\hat{\theta}_{CO}) \) we first note that \((\hat{\mu}_1, \hat{\mu}_2)\) and \((\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})\) are independent sets of estimators [36]. Denoting \( \eta = \mu_2 - \mu_1 - c \) and \( \nu = (\sigma^2_1 + \sigma^2_2 - 2 \rho \sigma_1 \sigma_2)^{1/2} \), we can use Taylor series expansions of (10) to obtain

\[
\text{var}(\hat{\theta}_{CO}) = \left( \frac{\eta}{\nu} \right)^2 \phi^2 \left( \frac{\eta}{\nu} \right) \left[ \frac{1}{\eta^2} \text{var}(\hat{\eta}) + \frac{1}{\nu^2} \text{var}(\hat{\nu}) \right]
\]

where \( \phi(.) \) is the standard normal PDF. We have

\[
\text{var}(\hat{\eta}) = \text{var}(\hat{\mu}_1) + \text{var}(\hat{\mu}_2) - 2 \text{cov}(\hat{\mu}_1, \hat{\mu}_2)
\]
where \( \text{var}(\hat{\mu}_i) = \sigma_i^2/N \) (for \( i = 1, 2 \)) and \( \text{cov}(\hat{\mu}_1, \hat{\mu}_2) = \rho \sigma_1 \sigma_2/N \). Hence (A8) simplifies to \( \text{var}(\hat{\eta}) = \nu^2/N \). To obtain \( \text{var}(\hat{\nu}) \) we first expand it as

\[
\text{var}(\hat{\nu}) = \frac{1}{4\nu^2} \left[ \text{var}(\hat{\sigma}_1^2) + \text{var}(\hat{\sigma}_2^2) + 4 \text{cov}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) - 4 \text{cov}(\hat{\sigma}_1^2, \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2) - 4 \text{cov}(\hat{\sigma}_2^2, \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2) \right]
\]  

(A9)

We require the following results from standard theory for the normal distribution [36]:

\[
\text{var}(\hat{\sigma}_i^2) = 2\sigma_i^4/N
\]

\[
\text{cov}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) = 2\rho^2 \sigma_1^2 \sigma_2^2/N
\]

and

\[
\text{cov}(\hat{\sigma}_i^2, \hat{\rho}) = \rho(1 - \rho^2) \sigma_i^2/N
\]

for \( i = 1, 2 \). Substituting these results into (A9) gives \( \text{var}(\hat{\nu}) = \nu^2/2N \). Finally, substituting all these results into (A7) gives

\[
\text{var}(\hat{\theta}_{CO}) = \phi^2 \left( \frac{\eta}{\nu} \right) \left( 1 + \frac{\eta^2}{2\nu^2} \right)/N
\]

which is equivalent to result (11).

A2. Parallel group design

We again use approximation (A7), but now

\[
\text{var}(\hat{\nu}) = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}
\]  

(A10)

and

\[
\text{var}(\hat{\nu}) = \frac{1}{4\nu^2} \left[ \text{var}(\hat{\sigma}_1^2) + \text{var}(\hat{\sigma}_2^2) + 4\rho^2 \text{var}(\hat{\sigma}_1^2, \hat{\sigma}_2) - 4\rho \text{cov}(\hat{\sigma}_1^2 + \hat{\sigma}_2^2, \hat{\sigma}_1\hat{\sigma}_2) \right]
\]  

(A11)

Using Taylor series for the variances and covariances in (A11), and taking \( \rho \) as a constant, we obtain after simplification that

\[
\text{var}(\hat{\nu}) = \frac{1}{2\nu^2} \left[ \frac{\sigma_1^2}{N_1} (\sigma_1 - \rho \sigma_2)^2 + \frac{\sigma_2^2}{N_2} (\sigma_2 - \rho \sigma_1)^2 \right]
\]  

(A12)

which then leads to result (13) for \( \text{var}(\hat{\theta}_{CO}) \).

ACKNOWLEDGEMENTS

This work was partly supported by the Natural Sciences and Engineering Research Council. The author thanks Drs Gordon Guyatt, Tony Perkins and David Sackett for input on preliminary versions of this paper, and Dr Elizabeth Juniper and Lauren Griffith for provision of the data used in the example.
REFERENCES


17. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. British Medical Journal 1999; 319:1492–1495.

18. Bjerre LM, Lelorier J. Expressing the magnitude of adverse effects in case-control studies: ‘the number of patients needed to be treated for one additional patient to be harmed’. British Medical Journal 2000; 320:503–506.


34. Walter SD, Irwig L. Estimating the number needed to treat (NNT) index when the data are subject to error. *Statistics in Medicine* 2001; 20:893–890.
