The number needed to treat: problems describing non-significant results

The number needed to treat (NNT) and its counterpart, the number needed to harm (NNH), are commonly used measures of effect in studies of mental health treatments. The NNT is an absolute measure, which conveys a more direct and intuitive impression of the effect on an individual than the absolute risk reduction (ARR).

To illustrate this, consider an imaginary trial comparing the effects of cognitive behavioural therapy versus interpersonal psychotherapy for people with bulimia. If 5% of people receiving cognitive behavioural therapy relapsed after treatment and 25% of people receiving psychotherapy relapsed, we could express this as an absolute risk reduction:

ARR = absolute difference in event rates (25–5%) = 20% (interpretation: cognitive behavioural therapy reduces the absolute risk of relapse by 20% compared with psychotherapy)

Alternatively, we could describe the result as a number needed to treat:

NNT = 1/ARR (1/20%) = 5 (interpretation: treat five people with cognitive behavioural therapy rather than psychotherapy to prevent one relapse)

The NNT allows us to dispense with percentage risks and helps describe effects in terms of real people. Confidence limits can also be provided for the NNT, which, loosely interpreted give us an idea of the range of effect we might expect most of the time. When the confidence interval includes a situation where intervention has no effect, the result is said to be non-significant.

So far, so good. The NNT appears to be a simple measure that can intuitively convey both the magnitude and the certainty of a trial result. There is a complication when dealing with non-significant results, however. To understand how this complication comes about, let's look at how NNT behaves with respect to the ARR.

The NNT is a reciprocal function; the inverse of absolute risk reduction. This relationship may be represented graphically (see webextra). The graph shows that the function is not continuous. As the ARR crosses the "line of no effect" from positive to negative, the NNT crosses from plus infinity to minus infinity. When the ARR is non-significant, by definition its confidence interval includes or touches zero. The confidence interval for the corresponding NNT will straddle plus and minus infinity.

When the NNT describes a significant result, the interpretation is straightforward. Suppose our earlier result (ARR=20%) was conventionally significant at the 5% level. By definition, the 95% confidence interval for the ARR would not include zero. Suppose the confidence limits for the ARR were 10% and 30%. In this case, we invert the confidence limits for the ARR to yield the corresponding confidence limits for the NNT. We can state the entire result as NNT 5, 95% CI 3.3 to 10. The result suggests we need to treat anywhere from 3 to 10 people with cognitive behavioural therapy to prevent one relapse.

But look what happens when the result is non-significant: Suppose the results were a non-significant ARR of 5%, with 95% CI –5% to 10%. To obtain the corresponding NNTs, we invert each percentage. Thus we obtain NNT 20, 95% CI –20 to +10. This looks rather odd. The range described by the confidence interval of the NNT does not seem to contain the mean value.

To understand what is going on, consider the interpretation of the NNT. A negative NNT is equivalent to the number needed to

harm – the number of people needed to treat with cognitive behavioural therapy rather than psychotherapy to *cause* one relapse. Interpreting the non-significant result, the upper confidence limit suggests that you will need to treat just a few people (10) with cognitive behavioural therapy to avoid one relapse. The lower confidence limit suggests that by treating 20 people with cognitive behavioural therapy, you will cause one relapse. The mean estimated effect, between the two extremes, is a more moderate conclusion. It suggests that we need to treat a greater number of people (20) to avoid one relapse. Note that the range does include NNTs of both plus and minus infinity (as shown by the figure). These are the situations where you would need to treat an infinite number of people to avoid (plus infinity) or cause (minus infinity) one relapse - in other words, when the treatment has no effect.

In the journal of *Evidence-based Mental Health* we are faced with a difficult choice when writing up non-significant results. Do we write a strange set of numbers to describe each NNT, without any explanation? Or do we describe each as we have done in this example, at the cost of several extra pages of text? Many journals avoid the dilemma by omitting non-significant NNTs. This may deprive us of the intuitive "feel" of the NNT in a situation where we really need a feel for the certainty of the result. Alternatively, we might present the mean NNT alone without its confidence intervals. This risks interpreting NNT at its mean value, without any regard for uncertainty around the estimate.

Another method for presenting non-significant NNTs is to use the concepts of "number needed to treat in order to benefit one person" (NNTB) and the "number needed to treat in order to harm one person" (NNTH).[1] In our example, we might write NNTB 20, 95% CI NNTH 20 to ∞ to NNTB 10. This method makes explicit the scenario where treatment has no effect (the need to treat an infinite number of people to cause or avert one relapse) and avoids negative NNTs. The method is not without its problems, however. First, the mean value still appears to the uninitiated to sit outside its confidence limits. One way of getting around this is simply to reverse the confidence limits, as follows: NNTB 20, 95% CI NNTB 10 to ∞ to NNTH 20. The second drawback of this notation is that it includes two "new" effect measures and the off-putting concept of infinity. Arguably, the complexity of the notation runs the risk that we go away remembering the mean value of the NNT and turn a blind eye to the tricky confidence intervals.

Evidence-based Mental Health is a succinct digest of the latest high quality evidence for busy practitioners. Non-significant NNTs exemplify the trade off that we sometimes face between providing as much detail as possible and making information readable and interpretable. For the moment, our position is in favour of readability at the expense of detail: we do not present non-significant NNTs.

Journal processes, like the NNT, should be reciprocal. We would welcome your feedback about whether and how you would like to see non-significant NNTs represented. (To provide your feedback contact the Medical Director, Dr Vivek Muthu, at vivek.muthu@bazian.com)

DR VIVEK MUTHU
MEDICAL DIRECTOR

Relationship between NNT and ARR

1 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; **319**: 1492–1495.