NNT for studies with long-term follow-up

The first article in the series for learners of evidence-based medicine discusses the concept of number needed to treat (NNT).[1] However, studies involving patients who need long-term follow-up, such as those with cancer or chronic cardiac conditions, commonly use time-to-event or survival analysis. It is important to realize that the NNT is not calculated in the same way for these studies.

NNT from survival analysis data should be estimated by the hazard ratio[2] and is not based on the difference in event rates between treatment groups at the end of follow-up.[3]

Mario L. de Lemos
British Columbia Cancer Agency
Vancouver, BC

References

1. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ* 2004;171(4):353-8.
2. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319(7223):1492-5.
3. de Lemos ML. How to survive the survival plots [letter]. *Lancet* 2002;360(9337):954.

DOI:10.1503/cmaj.1041426

[Three of the authors and a colleague respond:]

Mario de Lemos advises that for trials in which survival analysis is used, clinicians should ideally calculate the NNT from the hazard ratio.[1] We agree, but would emphasize that more important than the small differences created by the choice of method to calculate NNT are the very large differences consequent on different baseline risks. In this letter we review issues related to the calculation of NNT directly from trial data and illustrate what we believe is the appropriate approach, taking into account patients' baseline risk.

Consider 2 women. One is a tall, slightly overweight 50-year-old recently postmenopausal woman, who exercises regularly and who has normal bone mineral density. The second is a small 75-year-old woman who does not exercise and has a history of 4 vertebral fractures. The question we address here is how much hormone replacement therapy would reduce fracture risk in these women.

Before going any further in our consideration of these particular women, however, we will look at the "average" patient, using data from the Women's Health Initiative (WHI) trial, a large randomized trial of hormone replacement therapy,[2] which reported both event rates and a survival analysis.

NNT from event rates at the end of follow-up: Our first analysis is the "crude" or naïve approach that de Lemos criticizes. As described in our paper,[3] clinicians can calculate the NNT as the inverse of the difference in event rates (or absolute risk reduction) at the end of the study follow-up. According to the WHI data, among the 8506 women who were randomly assigned to receive active treatment, 44 had a hip fracture; in the placebo (control) group, 62 of 8102 had a hip fracture by the end of the study (after an average of 5.2 years of follow-up). These data are shown in Table 1, together with the NNT of 403, obtained by taking the reciprocal of the absolute risk reduction. In other words, this analysis suggests that we would need to treat 403 women with hormone replacement therapy over 5.2 years to prevent one hip fracture. Table 1 highlights the fact that the NNT is different over different time frames. For example, per year, we would have to treat approximately 2000 women to prevent one hip fracture. This can be calculated most easily by multiplying the NNT by 5 (403 × 5) and a little more tediously by calculating the event rates per year in treatment and control groups (Table 1). Clearly, the time frame is critical for NNT, and clinicians should insist on knowing the time frame associated with any NNT.

NNT from trials reporting survival analysis: In the paper cited by de Lemos, Altman and Andersen[2] outlined 2 methods (methods 1 and 2 below) for calculating NNT from trials that report the results of survival analyses; one method uses the difference in estimated survival probabilities between the treatment and control groups, and the other uses the hazard ratio and the survival probability in the control group. The rationale for using a survival analysis (i.e., time-to-event

Table 1: Number needed to treat for hip fracture, calculated from event rates and absolute risk reduction over different times*

	Group; event rate (ER)				
Time	Treatment (ER_t)	Control (ER_c)	ARR ($ER_c - ER_t$)	NNT (inverse of ARR)	Comments
Over 5.2 yr	44/8506 = 0.52% (0.00517 × 100)	62/8102 = 0.77% (0.00765 × 100)	0.248% (0.765% – 0.517%)	403 (100%/0.248%)	Based on number of events at end of follow-up, average 5.2 yr[3]
Over 1 yr	0.10% (0.52%/5.2 yr)	0.15% (0.77%/5.2 yr)	0.05% (0.15% – 0.10%)	2000 (100%/0.05%)	From annualized event rates[3]

Note: ARR = absolute risk reduction, NNT = number needed to treat.
*Data from the Women's Health Initiative trial.[2]

methods) is that it adjusts for censoring (the loss of at-risk study participants over various amounts of time since enrolment because of the termination of data collection or because of competing events such as death from other causes). In the WHI trial, follow-up ranged up to 8.5 years, with an average of 5.2 years.

Method 1, using survival probabilities in treatment and control groups: We can calculate the NNT from the inverse of the difference in survival probabilities between the treatment and control groups (Table 2). Apart from adjustment for censoring, this is exactly the same method as outlined above for event rates; it's just a matter of how the information is framed (event or nonevent, i.e., survival without an event). However, the survival probabilities may not be reported, in which case you might have to read them from the survival curves, which is a little tedious. For the WHI trial, this method suggests an NNT of 357 at 5.2 years.

Method 2, using the hazard ratio and the survival probability in the control group: Survival analysis produces hazard ratios (HRs). Although for many purposes HRs can interpreted as if they were rate ratios (or relative risks), the calculations that produce them are different, being based on complex statistical methods. In any case, you can calculate the NNT from the HR and the survival rate (probability) in the control group at a specified time point.[1] The calculation is based on the following equation:

$$\text{NNT} = 1/\{[\,S_c\,(t)^{\text{HR}}] - S_c(t)\}$$

where $S_c(t)$ is the survival probability in the control group at a specified time t. To do this calculation, you need the HR and the survival probability at your chosen time; again, this value is unlikely to be provided explicitly in published reports, but you can read it off the survival curve. This method is illustrated in Table 3, which uses the HR for hip fracture and estimates of survival probabilities at a variety of time points, obtained by reading them off the Kaplan–Meier curve in the WHI report.[2] This approach has the advantage that clinicians willing to deal with the formula shown above can readily calculate the NNT at any specified time point during the follow-up period. Because it is based on the survival analysis, it is adjusted for censoring. Again, it is clear that the time point chosen has a major impact on the numeric value of NNT. This method gives an NNT of 421 at 5.2 years.

Table 2: Number needed to treat for hip fracture, calculated from survival probabilities in treatment and control groups

Time point	Group; survival probability* (S)		$S_t - S_c$	NNT (inverse of $S_t - S_c$)
	Treatment (S_t)	Control (S_c)		
5.2 yr	0.9958	0.9930	0.0028	357

*Estimated from Fig. 3 (Kaplan–Meier curves for hip fracture) in the Women's Health Initiative trial.[2]

Table 3: Number needed to treat for hip fracture, calculated from hazard ratio and survival probabilities in control group at various time points

Time point, yr	Hazard ratio for hip fracture	Survival probability in control group*	NNT†
1	0.66	0.9990	2942
3	0.66	0.9965	841
5.2	0.66	0.9930	421
7	0.66	0.9885	257

*Estimated from Fig. 3 (Kaplan–Meier curves for hip fracture) in the Women's Health Initiative trial.[2]
†NNT = $1/\{[S_c(t)^{\text{HR}}] - S_c(t)\}$, where S_c is survival probability in control group at time t and HR is the hazard ratio.

These 2 methods are effectively the same and should give the same NNT for a given time point, because both are based on the results of the survival analysis. The difference observed in our example (357 v. 421 at 5.2 years) may relate to the difficulty and likely error in reading very small probabilities off the survival curves; also, the published curves[2] are stepped rather than smooth, whereas the hazard ratio is constant.

Comments: All of these methods rely on the same underlying principle. The NNT is based on the inverse of the difference between the event rates (or their complement, the survival rates) in the treatment and control groups. The most important difference between them is that the results of a survival analysis allow for censoring. In trials that may be substantially affected by censoring, estimates of NNT may be inaccurate if event rates are used.

The approaches we have illustrated so far assume that the particular patient before us has a baseline risk of hip fracture corresponding to the average of the women enrolled in the WHI. This is certainly not true for the 2 patients described above. The first patient has a risk that is probably about half of the baseline risk of women in the trial, or about 0.4% over 5 years. Using a crude estimate of the relative risk calculated from the event rates (0.52/0.77 or 0.68) and a relative risk reduction of 0.32 (1.00–0.68), we estimate a risk difference of 0.4% × 0.32 or about 0.13%. The NNT is therefore 100/0.13 or 769. We could have arrived at the same answer (with a slight difference because of rounding) by multiplying the NNT in Table 1 by 2 (403 × 2 = 806).

Alternatively (and, in theory, preferably) we could use an approach based on the hazard ratio from the survival analysis. Ideally, we would use the formula given above and in Table 3, substituting the patient's probability of hip fracture (0.4) for the control hip fracture rate (0.77). With a risk half that of the control group, the NNT would be double that in Table 3: 421 × 2 = 842.

Consider now the second patient, whose risk of hip fracture is approximately 20 times that of the average in

the WHI (20 × 0.77% = 15.4%). We could use both the approaches described above. For the crude approach, the risk difference is now approximately 15.4% × 0.32 or 4.93% and the NNT 100/4.93 or just slightly above 20. Using the hazard ratio approach for this patient also yields an NNT of just over 20.

As we have shown here, differences between naïve approaches to calculating NNT based on event rates and more sophisticated approaches based on survival analysis may not be large enough to change clinical decisions. We suggest that clinicians who are interested in using the NNT to help guide their practice should not be overly concerned about inaccuracies that may arise from estimating the NNT from event rates, especially when using data from large, randomized trials with high rates of follow-up. What they must avoid is applying NNTs from trial data without considering how their patient's baseline risk may differ from that of the patients in the trial. That mistake could lead to serious miscalculations of the NNT that *would* have implications for clinical decision-making.

Alexandra L. Barratt
School of Public Health
University of Sydney
Sydney, Australia
Peter C. Wyer
Columbia University College of Physicians and Surgeons
New York, NY
Gordon Guyatt
Departments of Medicine and of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, Ont.
Judy M. Simpson
School of Public Health
University of Sydney
Sydney, Australia

References

1. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319:1492-5.
2. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
3. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al; for the Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ* 2004;171(4):353-8.

DOI:10.1503/cmaj.1041709

How to diagnose diabetes

In their commentary on the impact of new guidelines for glucose tolerance testing, Andrew Lyon and associates[1] argue against increased use of the oral glucose tolerance test (OGTT) on the grounds of poor reproducibility, cumbersomeness and questionable cost-effectiveness. They rightly conclude that devoting resources to programs that can help patients to modify their risk for diabetes is preferable to performing more OGTTs. However, it would have been appreciated if they had considered the simplified or abbreviated version of the glucose tolerance test[2,3] in terms of its suitability for detecting new cases of diabetes; this form of the test would be both reliable and less expensive.

Gurusamy Sivagnanam
Asian Institute of Medicine, Science and Technology
Kedah, Malaysia

References

1. Lyon AW, Larsen ET, Edwards AL. The impact of new guidelines for glucose tolerance testing on clinical practice and laboratory services. *CMAJ* 2004;171(9):1067-9.
2. Posner NA, Silverstone FA, Breuer J, Heller M. Simplifying the intravenous glucose tolerance test. *J Reprod Med* 1982;27(10):633-8.
3. Forrest RD, Jackson CA, Yudkin JS. The abbreviated glucose tolerance test in screening for diabetes: the Islington Diabetes Survey. *Diabet Med* 1988;5(6):557-61.

DOI:10.1503/cmaj.1041707

Andrew Lyon and associates[1] point out that the Canadian Diabetes Association's new clinical guidelines may increase the burden on laboratories because of increased use of the OGTT. I would like to add that the diagnosis of diabetes is mainly initiated by family doctors, but they may be too busy to implement any screening or to follow up appropriately once diabetes has been identified.[2] The burden on family practitioners to initiate mass testing and