PROGRESS IN CLINICAL NEUROSCIENCES: Measuring the Benefit of Therapies for Neurological Disorders

Miguel Bussière, Samuel Wiebe

ABSTRACT: *Background:* Clinicians in the neurosciences need to interpret and apply a growing body of evidence about therapy. *Methods:* Using a clinical scenario about painful diabetic neuropathy and evidence about one treatment option, we review the advantages, limitations, and the clinical interpretation of commonly reported measures of effectiveness, emphasizing their application to the care of individual patients. *Results:* Absolute differences between treatment and control groups (e.g., absolute risk difference) are clinically intuitive and preferable to relative measures (e.g., relative risk). The number needed to treat is particularly useful and clinically applicable. Continuous measures are best interpreted using thresholds of clinically important change, which provide information about the number of patients experiencing meaningful improvement or worsening. *Conclusions:* Using simple principles of evidence based care, clinicians can correctly interpret the common measures of treatment effectiveness and apply them to the care of individual patients.

RÉSUMÉ: L'évaluation du bénéfice dans le traitement des maladies neurologiques. *Introduction:* Les cliniciens en neurosciences doivent interpréter et appliquer de plus en plus de données sur le traitement des maladies neurologiques. *Méthodes:* Au moyen d'un scénario clinique décrivant une neuropathie diabétique douloureuse et des données sur une option thérapeutique, nous avons revu les avantages, les limites et l'interprétation clinique de mesures d'efficacité fréquemment rapportées, en mettant l'accent sur leur application au traitement de patients individuels. *Résultats:* Les différences absolues entre le groupe sous traitement et le groupe témoin (e.g. la différence absolue du risque) sont intuitives en clinique et préférables à des mesures relatives (e.g. le risque relatif). Le nombre de sujets à traiter (NNT) est particulièrement utile et il est applicable en clinique. La meilleure interprétation de mesures continues est l'utilisation de seuils de changements importants en clinique qui donnent de l'information sur le nombre de patients dont l'état s'améliore ou empire de façon significative. *Conclusions:* En utilisant des principes simples de thérapie factuelle, les cliniciens peuvent interpréter correctement les mesures courantes d'efficacité thérapeutique et les appliquer aux soins de patients particuliers.

In the last decade, there have been tremendous advances in our understanding and treatment of neurological disorders. Indeed, the 1990s were labeled as the "decade of the brain". Clinicians have to make sense of the results of over 9,000 randomized trials dealing with therapeutic interventions in the neurosciences, published in the last decade alone.[1] By applying evidence-based care principles, clinicians can evaluate the best available evidence and apply it to the care of individual patients.

Our aim is to review the advantages and limitations of measures commonly used to describe the efficacy of treatments emphasizing the "number needed to treat" (NNT), a clinically useful measure of the effectiveness of interventions.

MEASURES OF THE EFFECTIVENESS OF INTERVENTIONS/ TREATMENTS

Randomized controlled trials (RCTs) and systematic reviews express the efficacy of treatments in several ways.[2-6] An understanding of these measures allows us to appropriately assess and apply the information provided in the literature. Table 1 summarizes the calculations described throughout the text. Readers are also referred to several electronic references which provide resources simplifying calculation of the various measures of effectiveness mentioned in the text below (University of Western Ontario Evidence Based Neurology - http://www.uwo.ca/cns/ebn/; Oxford Centre for Evidence Based

From the Department of Clinical Neurosciences (SW), University of Calgary, Calgary, Alberta, Canada; Department of Clinical Neurological Sciences (MB), University of Western Ontario, London, Ontario, Canada
Reprint requests to: Miguel Bussiere, Division of Neurology, Department of Clinical Neurological Sciences, University of Western Ontario, 339 Windermere Road, Rm 7-GE6, London, Ontario, N6A 5A5, Canada

Medicine - http://www.cebm.net/; University of Toronto Centre for Evidence Based Medicine - http://www.cebm.utoronto.ca/).

Clinical Scenario

Prompted by a patient with painful diabetic neuropathy who asks you about using Gabapentin, you assess a multi-centered RCT evaluating the effect of Gabapentin on painful diabetic neuropathy.[7] The trial involved 165 diabetic patients with painful neuropathy of 1-5 years duration, of whom 84 randomly received gabapentin and 81 received placebo. Patients were evaluated eight weeks later using standardized pain scales. When compared to the start of the trial, 52 patients in the gabapentin group and 30 in the placebo group had a moderate to marked improvement in their daily pain scores. The remainder of patients (32 in the gabapentin group and 51 in the control group) experienced an unfavourable outcome consisting of worsening, no change or minimal improvement. The authors report a NNT of 3.8 with 95% confidence intervals (CI) from 2.4 to 8.7.[8] The investigators also found that the mean improvement in pain scores in the gabapentin group was 1.1 points higher on an 11-point Likert scale than in the placebo group (p < 0.001). How can you interpret these and other measures of treatment effect?

Relative Versus Absolute Measures

We can group dichotomous (yes/no) measures of effectiveness into those that provide a relative estimate versus those that provide an absolute estimate of benefit. Examples of dichotomous outcome events are death, stroke or, as in our scenario, improvement or no improvement in daily pain severity. For consistency, all outcomes in this article are framed as adverse events (e.g., death, stroke or worsening, no change or minimal improvement in daily pain scores), even though the same statistical calculations can be applied with positive outcome events (e.g., moderate or marked improvement in daily pain scores).

Relative Measures

Two common ways to express the impact of treatment on dichotomous outcomes are the relative risk (RR) and the relative risk reduction (RRR). The RR is the ratio of the probability of an adverse event in the treatment group (experimental event rate or EER) and the probability of an adverse event in the control group (control event rate or CER), that is EER/CER. Therefore, the EER is 32/84 or 0.38, and the CER is 51/81 or 0.63. Consequently, the RR is 0.38/0.63 or 0.60. The RRR is calculated as 1-RR or 1-EER/CER, thus 1-0.60 or 0.40. A RR or RRR between zero and one indicates that a treatment is beneficial and reduces the risk of an adverse event, whereas a RR>1 or a negative RRR indicates that a treatment causes harm (Table 2).

The RR or the RRR are relatively easy to grasp clinically. The RRR is the proportion of baseline risk that is removed by the therapy. If a treatment does not alter the risk of an adverse event, the RR is one and the RRR is zero. As the ability of a treatment to lower the risk of an adverse event increases, the RR approaches zero and the RRR approaches one and vice versa (Table 2). The main disadvantage of the RR or RRR is that they do not reflect the baseline risk of adverse events without therapy. That is, the same RR and RRR can be obtained with different CER and EER. For example (Table 3), the same RR and RRR that was calculated from our clinical scenario can be obtained if in a hypothetical study, 0.18% of patients treated with gabapentin and 0.3% of those treated with placebo had an unfavourable outcome (EER=0.0018, CER=0.003, RR=EER/CER=0.6). Thus, the RR or RRR will overestimate the effect of therapy if the baseline rate of an adverse event is small (small CER) i.e. adverse events are rare. Conversely, if adverse events are common or frequent (large CER), the RR or RRR will underestimate the magnitude of the effect of therapy.

Another frequently used relative measure of efficacy is the odds ratio (OR). The odds ratio is the odds of an adverse event in the treatment group relative to that in the control group. The odds of an adverse event is calculated as the ratio of the number of patients in the group experiencing an adverse event vs. the number of patients in the group not experiencing this adverse event. Thus, in our example the OR=(32/52)/(51/30)=0.36. An OR<1 means that a treatment is beneficial or that it reduces the risk of an adverse event, whereas an OR>1 means that a treatment causes harm or increases the risk of an adverse event (Table 2).

Since the OR is a relative measure of treatment effect, like the RR or RRR, it does not reflect the baseline risk of an adverse event, as described above. In addition, the OR is not intuitively applicable to clinicians. For example, in our scenario, what does an OR of 0.36 mean to a clinician treating a patient with diabetic neuropathy? Clinicians may be tempted to substitute the RR, a more intuitive measure, for the OR which is difficult to interpret. However, it should be stressed that these two relative measures approximate each other only if the event rate in the treatment and the control groups is low, and if the treatment effect is small. If the event rate is high in either group, or if the treatment effect is large, the OR will be larger than the RR.[9] It should be acknowledged that some scientific journals may require the

Table 1: Definition of commonly used measures of therapeutic efficacy

	Adverse event	
Treatment	Present	Absent
Yes	a	b
No (control)	c	d

Control event rate	CER = c/c+d
Experimental event rate	EER = a/a+b
Relative risk	RR = EER/CER
Relative risk reduction	RRR = 1-RR
Absolute risk difference	ARD = CER-EER
Number needed to treat	NNT = 1/ARD
Control odds	CO = c/d
Experimental odds	EO = a/b
Odds ratio	OR = EO/CO or ad/bc

Table 2: Comparison of commonly used measures of therapeutic efficacy

Measure	No effect	Beneficial	Harmful
RR	1	Between 0 - 1	> 1
RRR	0	Between 0 - 1	Negative RRR
OR	1	< 1	> 1
ARD	0	Positive ARD (ARR)	Negative ARD (ARI)
NNT	Infinity	Small, positive NNT	Negative NNT (NNH)

Interpretations provided in this table assume an adverse outcome event. Absolute risk difference (ARD), absolute risk increase (ARI), absolute risk reduction (ARR), number needed to treat (NNT), number needed to harm (NNH), odds ratio (OR), relative risk (RR), relative risk reduction (RRR).

results of trials or systematic reviews to be reported in a specific way making it difficult for researchers to express results in the most clinically meaningful manner based on the arguments and literature presented in this article.

Absolute Measures

The absolute risk difference (ARD) is the difference between the risk or probability of an adverse event in the treatment group and that in the control group. It is, therefore, calculated as the CER-EER and, therefore, in the above example the ARD=0.63-0.38=0.25. Since the ARD is an absolute measure, it accounts for the baseline risk of an adverse event. Treatments that reduce the risk result in an absolute risk reduction (ARR), those that increase the risk result in an absolute risk increase. Treatments that reduce the risk of infrequent adverse events have a small ARR, and those that reduce the risk of common adverse events have a larger ARR. Using the hypothetical study results, if the EER were 0.0018 and the CER were 0.003 the RR would be 0.6 but the ARR would only be 0.0012 (Table 3).

Although the ARD is clinically more useful, it may be difficult to remember because it is expressed as a decimal. Fortunately, using the ARD, we can easily derive a much more clinically useful measure of effectiveness, the number needed to treat (NNT).[2-6] The NNT, first described by Laupacis et al.,[4] is defined as the number of patients that need to be treated over a certain time period in order to prevent one adverse event. This value is specific for the intervention, comparison, outcome and time frame used in the original research. The NNT is obtained by calculating the inverse of the ARD (1/ARR). NNTs can have a positive or negative value. Negative NNTs are obtained from absolute risk increases and indicate that a therapy is harmful because it increases the risk of an adverse event. A negative NNT can also be called a number needed to harm (NNH).

In our scenario the NNT=1/ARR=1/0.25=4, meaning that, as compared to placebo, gabapentin therapy for eight weeks moderately or markedly improved daily pain severity scores in one out of four patients with painful diabetic neuropathy. The NNT is therefore more easily remembered and directly applicable to the clinical environment. Similar to the ARD, the NNT reflects the baseline risk of an adverse event. In our hypothetical study, if the EER were 0.0018 and the CER were 0.003, the RR would be 0.6 but the ARR would only be 0.0012 and the NNT would be huge at 833 (Table 3). Empirical evidence demonstrates that the method used to report the results of clinical trials alters the response of clinicians to this information.[10,11] For example, clinicians judge interventions as significantly less effective when the results of the same trial are presented as absolute (e.g., ARR, NNT) rather than relative measures (RR, RRR, OR).[10,11] In our hypothetical study, it is readily apparent how clinicians would be more likely to consider as effective a therapy whose effect is reported as a RR=0.6, and RRR=0.4, than one with an ARR=0.0012, and NNT=883; even though all measures refer to the same results of one study. This reflects the inability of relative measures to take into account the baseline risk of adverse events, resulting in overestimates of effectiveness.

Table 3: The baseline risk of an adverse event affects absolute but not relative measures of therapeutic efficacy

	Baseline risk		Relative measures		Absolute measures	
	CER	EER	RR	RRR	ARD	NNT
Scenario 1	63%	38%	0.6	0.4	25%	4
Scenario 2	0.3%	0.18%	0.6	0.4	0.12%	833

Relative measures of effectiveness (RR, RRR) do not reflect the baseline risk of adverse events without therapy, whereas absolute measures vary with the baseline risk. Absolute risk difference (ARD), control event rate (CER), experimental event rate (EER), number needed to treat (NNT), relative risk (RR), relative risk reduction (RRR).

CALCULATING NNTS FROM OR, RR OR RRR

Although NNTs are increasingly used in clinical trials and systematic reviews of therapy, results are often still expressed in terms of RR, RRR or ORs. If NNTs are not provided, they can often be calculated from available information. If enough information is available to calculate the CER and EER or these rates are provided, the NNT can be calculated as described above. If the CER and either the RR or RRR are provided, the NNT can be estimated using a nomogram such as that provided by Chatelier et al.[12] Alternatively, an estimate of the NNT can be derived from the CER and OR using tables such as those provided by McQuay and Moore[6] or Sackett et al.[13] Simple formulas for calculating NNTs are shown in the appendix.

CALCULATING 95% CIS FOR NNTS

Since the NNT is an estimate of the benefit of a treatment it must be accompanied by the precision of its value, often expressed as the 95% confidence interval (CI). This interval gives us the range of values where we can be 95% certain the true NNT lies. The 95% CI for the NNT is calculated simply by taking the reciprocals of the 95% CI for the ARR. The 95% CI for the ARR are calculated using the following formulas:

$$95\% \text{ CI for ARR} = ARR \pm 1.96 \sqrt{\frac{CER \times (1 - CER)}{n_{controls}} + \frac{EER \times (1 - EER)}{n_{exp}}}$$

In our clinical scenario, the ARR is 0.25 and its 95% CI is 1.96 multiplied by the square root of {0.63(1-0.63)/81 + 0.38(1-0.38)/84}, which equals 0.15. Thus, the 95% CI for our ARR of 0.25 is (0.10, 0.40) and the 95% CI of our NNT of 4 is 1/0.10 to 1/0.40, or 2.5 to 10. In other words the best estimate of the NNT is 4, and we can be 95% confident that the NNT lies between 2.5 and 10.

Confidence intervals for NNTs can take positive and negative values. Consider an NNT of 15 with a 95% CI ranging from -5 to 10. This indicates that we cannot be sure whether this treatment is beneficial (positive NNT) or harmful (NNH). That is, we can be 95% certain that the effect of the intervention ranges from a harmful effect (NNH=5) to a beneficial effect (NNT=10). It also means that the effect of the intervention is not statistically significant because it encompasses harm and benefit.

EXTRAPOLATING NNTS FOR DIFFERENT LENGTHS OF FOLLOW-UP

It is paramount to remember that NNTs are always associated to a specific time period. That is, outcomes are assessed at specific times during a trial, and NNTs are calculated for outcomes at these predetermined times. Sackett et al.[14] suggest a simple formula to compare NNTs for different durations of follow-up: $NNT_S = NNT_T \times T / S$, where NNT_S = the NNT at the desired duration of follow-up (S) and NNT_T = the NNT for the duration of follow-up in the trial (T). In our clinical scenario, if we wanted to know the NNT at six months, instead of two months as in the RCT, we would obtain an $NNT_S = 4 \times 2/6 = 1.3$ (or 2, if we round upwards). Note that the longer the time, the smaller the NNT. This calculation assumes that the baseline risk of an adverse event is constant over time. This assumption may not always be true. For example, the risk of seizure recurrence after stopping an anti-epileptic medication decreases as a function of time[15] and the efficacy of anti-migrainous drugs is different at 0.5 and 3 hours.[16]

EXTRAPOLATING NNTS FOR DIFFERENT BASELINE RISKS

It is also possible to estimate a NNT for an individual patient that may be at higher or lower risk for an adverse event than patients participating in a RCT. A simple formula is used to estimate this NNT: $NNT_F = NNT_T / F$, where NNT_F = the NNT for an individual patient with an estimated baseline risk of "F" and NNT_T is the NNT determined from a trial. Estimates of baseline risk can be derived from several sources, including cohort studies describing the prognosis of similar patients, clinical prediction guides, or from a subgroup most similar to the patient in a RCT or systematic review.[17] In some situations, baseline risks are known or estimated from local clinical sources. In our clinical scenario, if we wanted to calculate the NNT for a patient who has an estimated baseline risk two times greater than patients in the RCT, we would obtain the $NNT_F = 4 / 2 = 2$. Note that the higher the estimated baseline risk, the smaller the NNT and vice versa. This calculation assumes that the relative risk of an adverse event is constant with varying baseline risk. Again, this assumption may not always be true. For example, the effect of carotid endarterectomy for stroke prevention differs in patients with varying degrees of carotid stenosis.[18] Although these adjustments to NNTs make them more applicable to clinical practice, clinicians must be wary when making such adjustments.

LIMITATIONS OF NNTS

Like other measures of therapeutic benefit, NNTs have several limitations. As described, some relate to the time and baseline risk differences between patients involved in the index studies and those to whom one is applying the results. Second, NNTs are population, disorder, treatment and outcome specific and therefore should not be directly compared across different disease conditions. Third, pooled NNTs derived from meta-analyses must be interpreted with caution since the baseline risk, clinical setting and outcomes assessed may vary appreciably among trials included in the analysis.[19] Fourth, there is no clear notion of the threshold NNT at which specific therapies become worthwhile or worthless. Like other measures of effectiveness, NNTs do not take into account the side effects, cost, values or preferences for any interventions. Therefore, some commentators propose measures that encompass both the benefit and harm of an intervention, such as the "likelihood of being helped versus harmed" or the "threshold number needed to treat." The latter is an approach to determine the magnitude of a NNT below which one should use the therapy, and above which one should not.[17,20]

CONTINUOUS OUTCOMES

Heretofore we have focused on dichotomous (yes/no) outcomes. However, many clinical outcomes are measured along a continuum. Examples include the time free of illness or relapses, the number of disease attacks, most laboratory values, and the scores of scales assessing severity of symptoms, pain, psychosocial outcomes, quality of life, and physical function. Most RCTs report the impact of interventions on continuous measures as the mean difference between treatment and control groups, and assess whether the differences are statistically significant. Alternatively, investigators may report the effect

Table 4: Obtaining the NNT in parallel RCTs with three outcome categories (improved, unchanged, worsened)

	Treatment		
Control	% Improved (x)	% Unchanged (y)	% Worsened (z)
% Improved (d)	*dx*	*dy*	*dz*
% Unchanged (e)	*ex*	*ey*	*ez*
% Worsened (f)	*fx*	*fy*	*fz*

Each cell is the product of multiplying the proportions in the corresponding headings. For example, cell dx is obtained by multiplying proportion (d) by proportion (x). Patients along the diagonal (clear cells) are unchanged. To obtain the NNT add up the cells of those who improved (gray cells, *ex+fx+fy*), subtract the cells of those who deteriorated (dark cells, *dy+dz+ez*), and divide 1 by the result. For a spreadsheet that calculates this NNT visit http://www.uwo.ca/cns/ebn/ebntools.xls and click on the tab labeled "NNT-3 category." Modified from Guyatt et al.[22]

size, which transforms outcomes into standard deviation units, and is obtained by dividing the mean between-group difference by the pooled standard deviation at baseline. Using generally accepted levels of small (0.2), medium (0.5) and large (0.8) effect sizes,[21] clinicians can assess the results of a RCT and compare them with those of other interventions. However, it is very difficult to make clinical sense of results reported as group means or effect sizes because these statistical measures convey no information about how many individuals actually improved, worsened, or remained unchanged. For example, if the mean differences between treatment groups are not statistically significant, clinicians may erroneously conclude that the treatment has no important effect. Yet, it has been shown that small mean differences may conceal clinically important changes occurring in a substantial number of individuals.[22] Similarly, large between-group differences may be due to large changes occurring in a relatively small number of individuals, while the minority remains unchanged.[23]

To make sense of continuous outcomes clinicians need to have a notion of the amount of change in a given measure that is clinically meaningful to individual patients. Once this threshold is known, therapeutic effects can be evaluated by comparing the number of individuals in the treatment (EER) and control groups (CER) that achieved this value. In our example, investigators determined that a moderate or marked improvement on a continuous outcome (i.e., pain score) would constitute a worthwhile effect of Gabapentin. They then reported the proportion of patients in the treatment (EER) and control groups (CER) who achieved this threshold, and calculated the ARD and NNT, which are easily interpretable clinically. Had these investigators decided to report their results only as a mean difference between treatment groups (1.1 points), we would be unable to assess its clinical meaning, regardless of its impressive statistical significance (p < 0.001).

When using continuous outcomes, such as a pain scale, to assess the effect of interventions in clinical practice or in therapeutic trials, a more realistic scenario is that patients respond not in two but in three ways. They may be improved, unchanged or worse. Guyatt et al.[22] reported a simple and intuitive method for obtaining NNTs in parallel randomized trials when there are three outcome categories (improved, unchanged, worsened) (Table 4). For a user-friendly spreadsheet that calculates this NNT visit http://www.uwo.ca/cns/ebn/ebntools.xls and click on the tab labeled "NNT-3 category."

Clinicians seeking to obtain NNTs from continuous outcomes will encounter two main hurdles. First, meaningful thresholds for many outcomes and symptom severity scales in the neurosciences do not exist or they are being developed.[23] Second, even if thresholds for specific outcome measures are known, authors may choose to continue reporting only the traditional, but clinically uninformative mean group differences. These hurdles can be overcome by increasing research efforts to identify clinically meaningful thresholds for continuous outcome measures, and by raising the awareness of authors and journal editors about reporting results in clinically meaningful ways. This will enable clinicians to interpret the relevance of therapeutic trials, apply them to clinical decisions about individual patients, and to move towards evidence-based care. On the other hand, there is some evidence that in many circumstances, the meaningful threshold of change can be approximated by using one half (0.5) of the standard deviation of the instrument assessing the outcome of interest (e.g., pain scale, quality of life instrument, life satisfaction scale, etc.), provided that this is described in the literature.[24]

DISCUSSION

Clinicians in the neurosciences need to make sense of a large growing body of evidence about treatment. By acquiring an understanding of and applying simple evidence based care principles, clinicians can correctly interpret the measures of treatment effectiveness described in the scientific literature and apply them to the care of individual patients.

REFERENCES

1. Wiebe S, Demaerschalk B. Evidence based care in the neurosciences. Can J Neurol Sci 2002; 29: 115-119.
2. Schulz KF, Moher D, Altman DG. Interpreting the number needed to treat. JAMA 2002; 288: 831-832.

3. Moher D, Altman DG, Schulz KF. Reporting the clinical importance of randomized controlled trials. CMAJ 2002; 166: 711-712.

4. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988; 318: 1728-1733.

5. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995; 310: 452-454.

6. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997; 126: 712-720.

7. Backonja M, Beydoun A, Edwards KR, et al. Gabapentin for the symptomatic treatment of painful neuropathy in patients with diabetes mellitus: a randomized controlled trial. JAMA 1998; 280: 1831-1836.

8. Backonja MM. Use of anticonvulsants for treatment of neuropathic pain. Neurology 2002; 59(5 Suppl 2): S14-S17.

9. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ 1998; 316: 989-991.

10. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med 1992; 117: 916-921.

11. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. BMJ 1994; 309: 761-764.

12. Chatelier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. BMJ 1996; 312: 426-429.

13. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Chapter 5. Therapy. In: Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (Eds). Evidence-Based Medicine. How to Practice and Teach EBM. (2nd ed.) New York: Churchill Livingstone 2000: 136-137.

14. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Chapter 5. Therapy. In: Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (Eds). Evidence-Based Medicine. How to Practice and Teach EBM. (2nd ed.) New York: Churchill Livingstone 2000: 116-117.

15. Medical Research Council Antiepileptic Drug Withdrawal Study Group. Prognostic index for recurrence of seizures after remission of epilepsy. BMJ 1993; 306: 1374-1378.

16. Smith LA, Oldman AD, McQuay HJ, Moore RA. Eletriptan for acute migraine (Cochrane Review). In: The Cochrane Library Issue 3. 2004. Oxford: Update Software.

17. McAlister FA, Straus SE, Guyatt GH, Haynes RB. Users' guides to the medical literature: XX. Integrating research evidence with the care of the individual patient. Evidence-Based Medicine Working Group. JAMA 2000; 283: 2829-2836.

18. Cina CS, Clase CM, Haynes RB. Carotid endarterectomy for symptomatic carotid stenosis (Cochrane Review). In: The Cochrane Library Issue 3. 2004. Oxford: Update Software.

19. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses - sometimes informative, usually misleading. BMJ 1999; 318: 1548-1551.

20. Sinclair JC, Cook RJ, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used? Derivation of the threshold number needed to treat and the minimum event rate for treatment. J Clin Epidemiol 2001; 54: 253-262.

21. Cohen J. The t test for means. In: Cohen J, (Ed). Statistical Power Analysis For The Behavioral Sciences. Hillsdale: Lawrence Erlbaum Associates, 1988: 19-74.

22. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ 1998; 316: 690-693.

23. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003; 56: 52-60.

24. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. Medical Care 2003; 41: 582–592.

APPENDIX

If the CER and RR or RRR are provided, the NNT can be calculated using the following formulas:

$$\text{if the RR<1 then NNT} = \frac{1}{(1\text{-}RR) \times CER}$$

$$\text{if the RR>1 then NNT} = \frac{1}{(RR\text{-}1) \times CER}$$

If the CER and OR are provided, the NNT can be calculated using the following formulas:

$$\text{if the OR<1 then NNT} = \frac{1\text{-}[CER \times (1\text{-}OR)]}{[(1\text{-}CER) \times CER \times (1\text{-}OR)]}$$

$$\text{if the OR>1 then NNT} = \frac{1\text{+}CER \times (OR\text{-}1)}{[CER \times (OR\text{-}1) \times (1\text{-}CER)]}$$