Number Needed to Treat (or Harm)

Martin R. Tramèr, M.D., D.Phil., Bernhard Walder, M.D.

Division of Anesthesiology, Department APSIC, Geneva University Hospitals, Geneva, Switzerland

Abstract. The effect of a treatment versus controls may be expressed in relative or absolute terms. For rational decision-making, absolute measures are more meaningful. The number needed to treat, the reciprocal of the absolute risk reduction, is a powerful estimate of the effect of a treatment. It is particularly useful because it takes into account the underlying risk (what would happen without the intervention?). The number needed to treat tells us not only whether a treatment works but how well it works. Thus, it informs health care professionals about the effort needed to achieve a particular outcome. A number needed to treat should be accompanied by information about the experimental intervention, the control intervention against which the experimental intervention has been tested, the length of the observation period, the underlying risk of the study population, and an exact definition of the endpoint. A 95% confidence interval around the point estimate should be calculated. An isolated number needed to treat is rarely appropriate to summarize the usefulness of an intervention; multiple numbers needed to treat for benefit and harm are more helpful. Absolute risk reduction and number needed to treat should become standard summary estimates in randomized controlled trials.

The effect of an experimental intervention versus that of a control intervention may be expressed in relative or absolute terms. Relative measures are relative risk, relative risk reduction, and the odds ratio. Absolute measures are absolute risk reduction and number needed to treat. This article concentrates on the number needed to treat (or harm).

Relative and Absolute Estimates of Efficacy and Harm

In randomized studies patients are randomly allocated to an experimental intervention (that is thought to have some beneficial effects) and to a control intervention. In a placebo-controlled trial, the control intervention is a placebo; in an active-controlled trial, the control intervention is another active treatment. The incidence of an event occurring with the experimental intervention may be called the experimental event rate (EER); accordingly, the incidence of the event occurring with the control intervention is the control event rate (CER).

Beneficial or harmful effects of the experimental intervention are usually measured by comparing the probabilities of events in the experimental and control group, for instance, using the relative risk (RR) which is the ratio of experimental and control event rates (RR = EER/CER). Another, related measure is the relative risk reduction (RRR), which is derived by subtracting the relative risk from 1 (RRR = 1 − RR) or by dividing the absolute difference between the control and experimental event rates by the control event rate (RRR = [CER − EER]/CER). The odds ratio is not discussed here because it is less helpful in clinical decision-making [1].

There is a main disadvantage, though, of using relative measures of treatment effects in clinical decision-making, as relative measures do not reflect the magnitude of the risk without therapy [2]. The risk without therapy, also called baseline risk or underlying risk, may have a major impact on the effect of a treatment. For instance, when the underlying risk is extremely low, even an effective therapy has no scope to show efficacy; and when the underlying risk is extremely high, the efficacy of an intervention may be exaggerated. Therefore to take into account the underlying risk in a study population, the absolute risk reduction (ARR) is often considered an additional measure of clinical effectiveness. The absolute risk reduction is the numerical difference between the control and experimental event rates (ARR = CER−EER). This number provides an idea about the clinical relevance of the effect of a treatment. However, even the absolute risk reduction is problematic as a measure of efficacy or harm, as it is a dimensionless, abstract number that may be difficult to incorporate into clinical practice [3]. The number needed to treat (NNT), which is simply the reciprocal of the absolute risk reduction (NNT = 1/ARR = 1/[CER−EER]), has the advantages of the absolute risk reduction and additionally provides a way of expressing the effect of a treatment in clinical terms. The number needed to treat is the number of patients who must be treated with an experimental intervention to achieve a particular result (beneficial or harmful) in one of them which would not have been the case had they all received the control intervention [4]. Table 1 summarizes the relations between experimental and control event rates, relative risk, relative risk reduction, absolute risk reduction, and number needed to treat. In analogy to the quantification of benefit (number needed to treat) and harm (number needed to harm), the number needed to screen has been

Correspondence to: Martin R. Tramèr, M.D., D.Phil., e-mail: martin.tramer@hcuge.ch

Table 1. Relation between risk (of efficacy or harm) with treatment and control, relative risk, relative and absolute risk reduction, and number needed to treat. An alternative way to calculate the relative risk reduction is 1 − RR.

Risk with treatment EER	Risk with control CER	Relative risk	Relative risk reduction	Absolute risk reduction	Number needed to treat
0.2	0.5	0.4	0.6	0.3	3.3
0.02	0.05	0.4	0.6	0.03	33.0
0.002	0.005	0.4	0.6	0.003	333.0

EER: experimental event rate; CER: control event rate; RR: relative risk; RRR: relative risk reduction; ARR: absolute risk reduction.

proposed as a further use of the number needed to treat concept [5]; this is the number of patients who must be screened for a disease to prevent one adverse outcome.

It becomes obvious from Table 1 that across different control event rates, an intervention may have a constant degree of efficacy in relative terms; however, in absolute terms it makes a huge difference if the intervention decreases the event rate from 0.5 to 0.2 or from 0.005 to 0.002. In the first case, about three patients need to be treated for one to show the desired result; in the latter case, more than 300 patients need to receive the treatment for one to benefit. Consequently, a number needed to treat must be interpreted in relation with a control event rate, or, in the ideal situation, with the true baseline or underlying risk of a study population or of an individual patient. However, the true underlying risk is rarely known. In placebo-controlled trials, we may assume that the incidence of events in the untreated control group (i.e., the control event rate) reflects to some extent the underlying risk. Sometimes indicators of the underlying risk are known; then a close relation between control event rates that are reported in individual trials and underlying risks in the study populations may be identified [6].

Specific terms have been used to summarize beneficial or harmful effects of interventions. For instance, to summarize beneficial effects, relative *benefit*, relative and absolute risk *reduction* (or *decrease*), and number needed to *treat* may be chosen. For harm, the corresponding terminology is then relative *risk*, relative and absolute risk *increase*, and number needed to *harm*. Number needed to treat *to benefit* and number needed to treat *to harm* have also been suggested [7].

It has been shown repeatedly that the way research results are presented may have an impact on decision-making by health care professionals. For instance, when trial results were presented as numbers needed to treat rather than as relative risk reductions, clinicians rated treatments less effective [8], and they showed less willingness to prescribe a treatment [9]. Health policymakers were also shown to be more impressed by measurements that reported relative benefits compared with those that were based on an index of absolute benefit [10].

Confidence Interval of the Number Needed to Treat

As with other estimates, the uncertainty in the estimated number needed to treat is accompanied by a confidence interval [7]. If the upper limit of the 95% confidence interval (CI) around the number needed to treat (which puts the intervention in the least favorable light) still lies within what we would regard as a clinically relevant effect, the intervention is definitely useful. A 95% CI for the number needed to treat can be constructed by simply inverting and exchanging the limits of a 95% CI for the absolute risk reduction [1]. Problems arise, however, when the difference
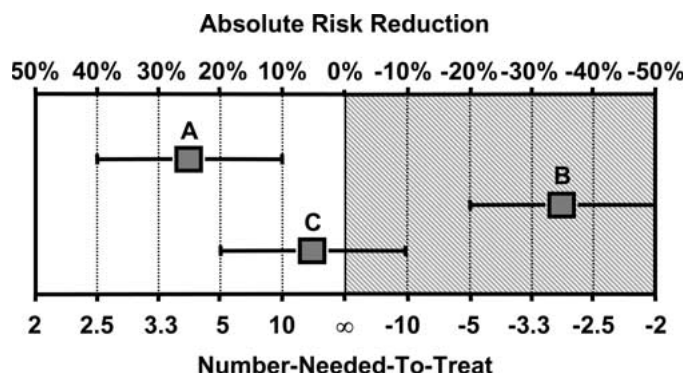

Fig. 1. Confidence interval for absolute risk reduction and number needed to treat. Horizontal bars are 95% confidence intervals (Adapted from Altman [7]).

between experimental and control intervention is not statistically significant. Then, the 95% CI around the absolute risk reduction spans from a positive to a negative number and includes 0%. This result is compatible with three conclusions: First, the experimental intervention is more efficacious than the control intervention (i.e., there is a reduction in the risk of a bad outcome). Second, there is equivalence, and thus the absolute risk reduction is 0%. An absolute risk reduction of 0%, however, translates into a number needed to treat of infinity [7]. Third, the experimental intervention is less efficacious than the control intervention (i.e., there is a "negative reduction" or, accordingly, an increase in the risk of a bad outcome). This curious situation is illustrated in Figure 1.

In Figure 1, treatment "A" is significantly more efficacious than the control treatment. The 95% CI of both the absolute risk reduction and the corresponding number needed to treat includes positive numbers only; the values for the number needed to treat, for instance, expand from +2.5 to +10.0. This result is compatible with fewer than 3 patients to no more than 10 patients who need to be treated for one to show a beneficial outcome with treatment "A" compared with the control. For treatment "B," both the point estimate and the entire 95% CI are on the negative side of the axis of the absolute risk reduction. Accordingly, the 95% CI of the number needed to treat expands from −5 to −2. That result is compatible with a negative reduction in risk, which may be interpreted as harm in certain circumstances (see below). Treatment "C," finally, is not significantly different from the control; the 95% CI of the absolute risk reduction expands from +20% to −10% including 0%. The corresponding 95% CI of the number needed to treat goes from +5 to infinity to −10. Altman suggested that such a confidence interval should be interpreted as a continuum that ranges from benefit (i.e., positive absolute risk reduction and number needed to treat) to equality (i.e., absolute

578

World J. Surg. Vol. 29, No. 5, May 2005

risk reduction = 0 and number needed to treat = ∞) to harm (i.e., negative absolute risk reduction and number needed to treat or, accordingly, positive absolute risk increase and number needed to harm) [7]. McQuay and Moore proposed that in such cases the point estimate of the number needed to treat only should be reported [3]. As we will see, both interpretations are correct for specific circumstances.

Interpreting and Using Numbers Needed to Treat

A number needed to treat should always be interpreted in its proper clinical context. Information should be given about the experimental intervention (e.g., for drugs: dose, regimen, and route of administration) and about the control intervention against which the experimental intervention has been tested (placebo, no treatment, another active intervention). The length of the observation period should be defined; models to estimate numbers needed to treat based on survival data have been proposed [11]. An estimate of the underlying risk of the study population should be provided. Finally, the endpoint should be specified.

A highly effective intervention has a small number needed to treat, indicating that only a few subjects need to receive that intervention for one to profit. However, numbers needed to treat of less than five are rare in clinical practice. For instance, in patients at high risk for postoperative vomiting, the most effective single-dose regimens of antiemetic drugs have numbers needed to treat of about five to prevent vomiting within 24 hours after surgery compared with placebo [12]. In patients having moderate to severe pain after minor surgical procedures (e.g., wisdom tooth removal), an adequate single oral dose of an nonsteroidal antiinflammatory drug has a number needed to treat of two to three compared with placebo for at least 50% pain relief during a 6-hour period [13]. Obviously, the clinical relevance of the effect of an intervention depends not only on the numerical value of the number needed to treat but also on the severity of the outcome. For instance, when inserted for no more than 1 week, antibiotic-coated central venous catheters have a number needed to treat to prevent bloodstream infection of about 20 compared with untreated catheters [14]. This relatively high number needed to treat is likely to be of clinical relevance because, unlike postoperative vomiting or pain, bloodstream infection is potentially lethal.

Examples of summary estimates of efficacy and harm from randomized controlled trials from perioperative medicine are shown in Table 2 and and Figure 2 [15–21]. Smoking cessation before surgery [15], endovascular coiling of intracranial aneurysms [16], and perioperative β-blockade [17] are clearly associated with beneficial outcomes: numbers needed to treat for one successful outcome compared with the respective comparators range from 4 to 14, and all 95% CIs include positive numbers only.

The already mentioned impact of the underlying risk on the efficacy of an intervention is well illustrated by the study on perioperative β-blockade (Figure 2) [17]. This study suggested that about eight patients undergoing noncardiac surgery need to receive perioperative bisoprolol to prevent one cardiac death. In that trial, however, 50% of the randomized patients had a history of previous myocardial infarction, one-third had angina pectoris, one-third had limited exercise capacity, and 15% had diabetes mellitus. Of the patients in the control group who did not receive

β-blockade, 17% died within 30 days owing to cardiac causes. This unusually high control event rate places in question the applicability of the trial. The study population may not necessarily represent what most anesthesiologists and surgeons see in daily clinical practice. It is likely that only the selection of patients with an extremely high underlying risk made it possible to demonstrate such an extraordinary benefit with β-blockade. Indeed, in similar randomized trials that included populations with much lower underlying risks (and thus lower control even rates), the beneficial effect of perioperative β-blockade was less pronounced [22].

The fourth example in Figure 2 is about patients with sepsis who received 6% hydroxyethylstarch or 6% gelatin solution [18]. With hydroxyethylstarch there was a significantly higher rate of acute renal failure; the absolute reduction in the risk of renal failure was −18% (95% CI −2% to −34%), and the corresponding numbers needed to treat was −5.5 (95% .CI −44.0 to −2.9). This result may be interpreted as an absolute increase in the risk of having renal failure with hydroxyethylstarch of 18% (95%, CI +2% to +34%) and a corresponding number needed to harm of 5.5 (95% CI +2.9 to +44.0).

The three last examples in Figure 2—epidural anesthesia [19], restricted transfusion strategy [20], and pulmonary artery catheter-guided therapy [21]—report on results that are not statistically significant; the 95% CI around the absolute risk reduction includes 0%, and around the number needed to treat it includes infinity. These interventions are not significantly different from their active controls. The methodologically correct interpretation of the 95% CI around the estimate of efficacy is a continuum that ranges from benefit to equality to harm [7]. Thus the data of all three trials are compatible with the experimental intervention having a beneficial effect compared with the control intervention, or with equivalence, or with the experimental intervention actually increasing the risk of a bad outcome compared with the control intervention. In clinical practice, we would simply conclude that there is no evidence of any difference between the experimental and control interventions.

A limitation of this technically correct interpretation of the 95% CI interval around the number needed to treat when the comparison is not statistically significant is shown in Figure 3. The figure summarizes the dose-response of the antiemetic efficacy of droperidol when added to morphine in a patient-controlled analgesia device [23]. With the lowest regimen tested, 5 μg of droperidol added to 1 mg of morphine, the number needed to treat is 16, and the 95% CI ranges from +4.7 to −11.0; indicating that this dose is not statistically significantly different from placebo. Uncritical application of the rules that were discussed earlier (Table 2, Fig. 2) would lead to the conclusion that the upper limit of the confidence interval corresponds to a number needed to harm of +11, and therefore the antiemetic droperidol had the potential to cause emesis. This, obviously, does not make sense from a pharmacologic point of view. Therefore, for comparisons between an experimental intervention and a placebo control that do not show a statistically significant difference it may indeed be useful to provide the point estimate of the number needed to treat only and to ignore the 95% CI [3].

For a rational use of the number needed to treat, two further issues must be addressed. First, it is rarely adequate to summarize the usefulness of an intervention with a single number needed to treat. For instance, there is no benefit in survival in critically ill patients with Swan-Ganz catheter-guided therapy compared with

Table 2. Randomized controlled trials in surgery and perioperative medicine.

Reference	Experimental intervention	Control intervention	Observation period	Setting (underlying risk)	Endpoint	Experimental Control intervention intervention		Absolute risk reduction (95%CI)	No. needed to treat (95%CI)
						No. of patients with endpoint/total no. of patients			
						EER	CER		
15	Smoking cessation (counselling and nicotine replacement)	Standard care	Within 4 weeks of surgery	Adults undergoing hip and knee replacement	Wound related complication	3/56 (5.4%)	16/52 (30.8%)	25.4% (11.6% to 39.3%)	3.9 (2.5 to 8.7)
16	Endovascular coiling	Standard neurosurgical clipping	1 year	Adults with ruptured intracranical aneurysms	Dependence or death	190/801 (23.7%)	243/793 (30.6%)	6.9% (2.6% to 11.3%)	14.0 (8.9 to 39)
17	Perioperative β−blockade from −1 wk to +4 wks	Standard perioperative care	Within 30 days of surgery	Adults with positive dobutamine echocardiography undergoing vascular surgery	Cardiac death	2/59(3.4%)	9/53 (17%)	13.6% (2.5% to 25%)	7.4 (4.0 to 40)
18	6% Hydroxyethylstarch	3% Fluid-modified gelatin	During ICU stay	Adults with severe sepsis or septic shock	Acute renal failure	27/65 (41.5%)	15/64 (23.4%)	−18.1% (−2.3% to −34.0%)	−5.5 (−44 to −2.9)
19	Epidural anesthesia	Standard care without epidural anesthesia	Within 30 days of surgery	Adults undergoing major abdominal surgery	Death or at least one major complication	255/447 (57.0%)	268/441 (60.8%)	3.7% (10.2% to −2.7%)	27 (9.8 to −36)
20	Restricted strategy of red cell transfusion	Liberal transfusion strategy	Within 30 days after admission to the ICU	Critically ill patients	Death of all causes	78/418 (18.7%)	98/420 (23.3%)	4.7% (10.2% to −0.8%)	21 (9.8 to −120)
21	Goal-directed therapy guided by a pulmonary artery catheter	Standard care without pulmonary artery catheter	In hospital	High-risk patients scheduled for major surgery	Death of all causes	78/997 (7.8%)	77/997 (7.7%)	−0.1% (2.3% to −2.5%)	−997 (44 to −41)

EER: experimental event rate; CER: control event rate; ICU: intensive care unit; CI: confidence interval.

580

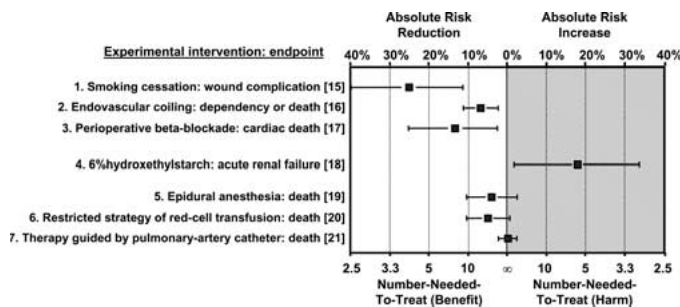World J. Surg. Vol. 29, No. 5, May 2005

Fig. 2. Absolute risk reduction/increase and number needed to treat/harm for seven active controlled randomized trials. Horizontal bars are 95% confidence intervals. See Table 2 for details.
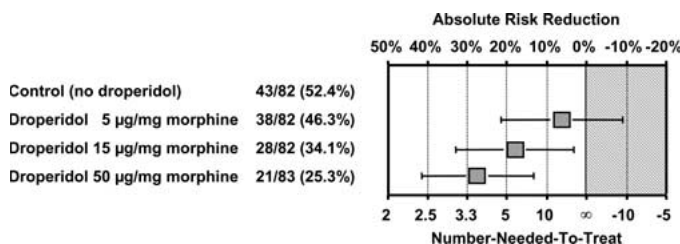

Fig. 3. Dose-response of the antiemetic efficacy of droperidol added to a patient-controlled analgesia (PCA) pump with morphine. Numbers are patients who were nauseous or who vomited within the first 24 hours postoperatively and the total number of patients. Horizontal bars are 95% confidence intervals (from Culebras et al. [23], with permission).

standard therapy (Table 2, Fig. 2) [21]. However, 8 of 997 (0.8%) patients who had a pulmonary catheter had a pulmonary embolism, and none of 997 controls had such a complication; this difference was statistically significant, the number needed to harm was 125 (95% CI 74–402). The combination of lack of efficacy and presence of major harm challenges even more the usefulness of pulmonary catheters. Reporting on multiple numbers needed to treat, for both beneficial and harmful outcomes, provides the most powerful tool for rational decision-making about the usefulness of medical interventions.

Second, to implement data from well designed clinical studies into clinical practice implies bringing summary estimates down to an individual patient level, which is a particular challenge [24]. Models that take into account estimates of efficacy and harm from randomized controlled trials and knowledge from observational studies on factors that predict the underlying risk have been proposed [25].

Conclusions

The effect of a treatment over a control may be expressed in relative or absolute terms. Both ways have advantages and disadvantages; however, for rational decision-making in daily clinical practice, absolute measures are more meaningful. The number needed to treat (and harm) is a powerful estimate of the effect of a treatment. It is particularly useful because it takes into account the underlying risk (what would happen without the intervention?). The number needed to treat not only tells us whether a treatment works but also how well it works. This number informs health care professionals about the effort that is needed to

achieve a particular outcome. Some simple rules must be considered, though, to avoid misuse of the number needed to treat. For instance, the number needed to treat should be correctly reported, and information about the clinical context should be provided. Also, a single number needed to treat is rarely appropriate for summarizing the usefulness of an intervention; multiple numbers needed to treat for distinct beneficial and harmful outcomes should be provided. Hopefully, the number needed to treat will become a standard summary estimate in randomized controlled trials [26].

References

1. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. B. M. J. 1995;310:452–454
2. Sackett DL, Haynes RB, Guyatt GH, et al. Clinical Epidemiology: A Basic Science for Clinical Medicine 2nd edition. Boston: Little, Brown, 1991
3. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann. Intern. Med. 1997;126:712–720
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. Rembold CM. Number needed to screen: development of a statistic for disease screening. B. M. J. 1998;317:307–312
6. Dumont L, Mardirosoff C, Tramèr MR. Efficacy and harm of pharmacological prevention of acute mountain sickness: quantitative systematic review. B. M. J. 2000;321:267–272
7. Altman D. Confidence intervals for the number needed to treat. B. M. J. 1998;317:1309–1312
8. 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
9. Lacy CR, Barone JA, Suh DC, et al. Impact of presentation of research results on likelihood of prescribing medications to patients with left ventricular dysfunction. Am. J. Cardiol. 2001;87:203–207
10. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. B.M.J. 1995;311:1056–1060
11. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. B.M.J. 1999;319:1492–1495
12. Tramèr MR. A rational approach to the control of postoperative nausea and vomiting: evidence from systematic reviews. Part I. Efficacy and harm of antiemetic interventions, and methodological issues. Acta Anaesthesiol. Scand. 2001;45:4–13
13. Moore A, Edwards J, Barden J, et al. Bandolier's Little Book of Pain: An Evidence-based Guide to Treatments. New York: Oxford University Press, 2003
14. Walder B, Pittet D, Tramèr MR. Prevention of bloodstream infections with central venous catheters treated with anti-infective agents depends on catheter type and insertion time: evidence from a meta-analysis. Infect. Control Hosp. Epidemiol. 2002;23:748–756
15. Møller AM, Villebro N, Pedersen T, et al. Effect of preoperative smoking intervention on postoperative complications: a randomised clinical trial. Lancet 2002;359:114–117
16. Molyneux A, Kerr R, Stratton I, et al. International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. Lancet 2002;360:1267–1274
17. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. N. Engl. J. Med. 1999;341:1789–1794
18. Schortgen F, Lacherade JC, Bruneel F, et al. Effects of hydroxyethylstarch and gelatin on renal function in severe sepsis: a multicentre randomised study. Lancet 2001;357:911–916
19. Rigg JR, Jamrozik K, Myles PS, et al. Epidural anaesthesia and analgesia and outcome of major surgery: a randomised trial. Lancet 2002;359:1276–1282
20. Hébert PC, Wells G, Blajchman MA, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in

critical care: transfusion .requirements in critical care investigators; Canadian Critical Care Trials Group. N. Engl. J. Med. 1999;340:409–417

21. Sandham JD, Hull RD, Brant RF, et al. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. N. Engl. J. Med. 2003;348:5–14

22. Stevens RD, Burri H, Tramèr MR. Pharmacologic myocardial protection in patients undergoing noncardiac surgery: a quantitative systematic review. Anesth. Analg. 2003;97:623–633

23. Culebras X, Corpataux JB, Gaggero G, et al. The antiemetic efficacy of droperidol added to morphine patient-controlled analgesia: a ran-domized, controlled, multicenter dose-finding study. Anesth. Analg. 2003;97:816–821

24. Chatellier G, Zapletal E, Lemaitre D, et al. The number needed to treat: a clinically useful nomogram in its proper context. B. M. J. 1996;312:426–429

25. Lee, A, Gin, T (2003) "From evidence to implementation" In: Tramèr, MR (editor), *Evidence-Based Resource in Anaesthesia and Analgesia*, 2nd edition, BMJ Books, London,

26. Nuovo J, Mehiikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. J. A. M. A. 2002;287:2813–2814