 International Journal of Epidemiology 2002;**31**:72–76

Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses

Toshiaki A Furukawa,[a] Gordon H Guyatt[b] and Lauren E Griffith[b]

Background Meta-analyses summarize the magnitude of treatment effect using a number of measures of association, including the odds ratio (OR), risk ratio (RR), risk difference (RD) and/or number needed to treat (NNT). In applying the results of a meta-analysis to individual patients, some textbooks of evidence-based medicine advocate individualizing NNT, based on the RR and the patient's expected event rate (PEER). This approach assumes constant RR but no empirical study to date has examined the validity of this assumption.

Methods We randomly selected a subset of meta-analyses from a recent issue of the Cochrane Library (1998, Issue 3). When a meta-analysis pooled more than three randomized controlled trials (RCT) to produce a summary measure for an outcome, we compared the OR, RR and RD of each RCT with the corresponding pooled OR, RR and RD from the meta-analysis of all the other RCT. Using the conventional *P*-value of 0.05, we calculated the percentage of comparisons in which there were no statistically significant differences in the estimates of OR, RR or RD, and refer to this percentage as the 'concordance rate'.

Results For each effect measure, we made 1843 comparisons, extracted from 55 meta-analyses. The random effects model OR had the highest concordance rate, closely followed by the fixed effects model OR and random effects model RR. The minimum concordance rate for these indices was 82%, even when the baseline risk differed substantially. The concordance rates for RD, either fixed effects or random effects model, were substantially lower (54–65%).

Conclusions The fixed effects OR, random effects OR and random effects RR appear to be reasonably constant across different baseline risks. Given the interpretational and arithmetic ease of RR, clinicians may wish to rely on the random effects model RR and use the PEER to individualize NNT when they apply the results of a meta-analysis in their practice.

Keywords Meta-analysis, odds ratio, risk ratio, risk difference, number needed to treat, evidence-based medicine

Accepted 25 June 2001

Today's meta-analyses summarize their results in several ways. When the outcome is dichotomous, some authors prefer the number needed to treat (NNT), because it expresses the efforts that clinicians and patients must expend in order to accomplish the desired treatment target. The NNT is calculated as the inverse of the risk difference (RD), which is an absolute measure of effectiveness. However, many meta-analyses continue to utilize relative measures of effectiveness such as odds ratio (OR) and risk ratio (RR).

It is the ultimate aim of evidence-based medicine (EBM) to individualize group data from clinical research, in order to satisfy each individual patient's values and preferences.[1] Therefore, many EBM theorists note that since event rates vary, often dramatically, across patients, a single NNT is unlikely to be applicable to all patients. They therefore advocate individualizing the NNT, depending on estimates of RR obtained from group studies and on each patient's expected event rate (PEER).[2] This approach is based on the assumption of constant RR, i.e. 'the relative benefits and risks of therapy are the same for persons

[a] Department of Psychiatry, Nagoya City University Medical School, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan.

[b] Departments of Medicine, and Clinical Epidemiology Biostatistics, McMaster University, 1200 Main St West, Hamilton, Ontario L8N 3Z5, Canada.

Correspondence: Toshiaki A Furukawa, Department of Psychiatry, Nagoya City University Medical School, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan. E-mail: furukawa@med.nagoya-cu.ac.jp

with high or low PEERs'.[3,p.120] Basically the argument goes that, assuming a constant RR, the absolute benefit of the treatment, usually expressed as NNT, is smaller among low risk patients than among high-risk patients. A classical example where this holds is the treatment of hypertension.[4]

But is this premise true for a wide variety of health interventions? Sackett *et al*. themselves wrote, 'this (constancy of RR) is a big assumption'.[5,p.170]

To the best of the present authors' knowledge, no empirical study to date has directly examined how applicable or generalizable various effect measures of meta-analyses are in actual practice. One study showed that RR and OR are more independent of the baseline risk than RD for a wide range of randomized controlled trials (RCT),[6] but these investigators examined individual trials and do not tell us which summary effect measure was better for meta-analysis.

Another study compared fixed as well as random effects model OR and RD for 125 meta-analyses and found that RD tended to be more heterogeneous among trials.[7] They noted that random effects estimates often showed wider CI than fixed effects models and concluded that, since formal tests of heterogeneity are often underpowered, it might be appropriate to assume that systematic differences among trials are always present and to use a random effects model. This study did not include RR in their comparisons. Moreover, neither of these two studies tells us how often a summary effect measure, OR or RR or RD, is constant over a range of baseline risks. The constancy of an effect measure is an important factor to consider in deciding if we can individualize NNT and which summary measure to utilize if we want to do so. The present report presents the results of an empirical examination of the generalizability of the most commonly used measures of association for summarizing treatment effects in meta-analyses.

Methods

We included all the meta-analyses in the field of psychiatry as well as a randomly selected subset of all the other meta-analyses in other branches of medicine contained in a recent issue of the Cochrane Library.[8] When a meta-analysis pooled more than three RCT to produce a summary measure for one outcome, we compared the OR, RR and RD of each of the included RCT with the pooled OR, RR and RD, respectively, from the other RCT. At minimum, a meta-analysis of three RCT would contribute three comparisons because, for a single outcome, we would compare each of the three RCT to a meta-analysis of the other two. Another meta-analysis of three RCT would contribute nine comparisons if the study pooled the three RCT for three discrete outcomes, such as acceptability of treatment, response, and side effects. Furthermore, for each outcome, the number of comparisons was equal to the number of RCT: a meta-analysis of four RCT would yield four comparisons, five RCT five comparisons, and so on.

Using methods described by Fleiss,[9] if the individual OR (or RR or RD) and the pooled OR (or RR or RD) were statistically significantly different at a conventional *P*-value of 0.05, we regarded them as discordant and, if not, as concordant.

We calculated pooled estimates using both a fixed effects model (Mantel-Haenszel) and a random effects model (DerSimonian and Laird). Theoretically neither may be entirely satisfactory because the latter is only exchanging the questionable homogeneity assumption of the former for a fictitious random distribution of effects.[10,11] In practice, when pooling non-heterogeneous studies, investigators have found that both agree rather well, but the random effect model tends to be more conservative, and often yields wider CI.[7,12]

Because it is conceivable that the rate of concordance could be artificially inflated due to small sample size of some of the RCT involved (that is, we fail to reject the hypothesis that the point estimate from the individual RCT differs from the pooled estimate because of inadequate precision, and therefore excessively wide CI), we conducted a sensitivity analysis restricted to comparisons in which both individual RCT and the corresponding meta-analyses produced statistically significant results. In order to examine the consistency of treatment effectiveness indices when the control event rate (CER) differs substantially, we conducted another sensitivity analysis limited to instances where the results of RCT and the meta-analysis were statistically significant and in addition the CER of individual RCT was less than half or more than twice of that of the weighted average of the other studies from that meta-analysis. We further examined concordance rates when the results of the individual RCT and the meta-analysis were statistically significant and the CER of individual RCT was three-times different from that of the weighted average of the other studies.

Our results showed that the fixed effects OR, random effects OR and random effects RR all produced potentially acceptable concordance rates between one RCT and the meta-analysis of similar RCT (see Results). In order to individualize NNT, we would apply these indices of treatment effectiveness to PEER by the following formulae:

$$NNT = \frac{1}{PEER \times (1 - RR)} \text{ and } NNT = \frac{1 - PEER + OR \times PEER}{PEER \times (1 - OR) \times (1 - PEER)}.$$

We therefore next examined the extent to which these models would produce similar individualized NNT across the range of baseline risks in which clinicians would typically apply the method of individualizing the NNT. For this analysis, we used meta-analyses which produced statistically significant fixed effects model OR. For each meta-analysis that met this criterion, we calculated the NNT assuming patient expected event rates of 0.1, 0.2, 0.3, 0.4, and 0.5.

To determine the extent of agreement, we needed to define a range of NNT in which the clinical implications are likely to be very similar. We chose the following (inevitably somewhat arbitrary) criteria: for NNT of 1–5 differences of $\leqslant 3$; for NNT of 6–10, differences of $\leqslant 4$; for NNT of 11–50, differences of $\leqslant 15$; for NNT 51–100, differences of $\leqslant 30$; for NNT over 100, $< 0.3 \times$ NNT. Using these criteria, we calculated agreement of the individualized NNT based on fixed or random effects OR and random effects RR.

Because the results were very similar between psychiatry and general medicine, we present the combined results. Because of lack of independence of effect measures in these sets of comparisons, we were unable to calculate 95% CI for the concordance rates or to examine if the differences in the concordance rates were statistically significant. The results, however, show us how often, in absolute terms, we can expect the pooled effects of meta-analyses to apply to separate groups of patients.

We did not exclude the comparisons where statistical heterogeneity was noted if the studies were combined and the summary measures were reported in the original meta-analyses. Nor did we consider the impact of switching from the absence to the presence of the selected outcome event for the RR (for instance, from death to survival, or persistent disease to cure), although the RR of event and the RR of no event can make a substantial difference in the estimated effect size, its 95% CI and observed heterogeneity (in contrast to RR, OR and RD are symmetrical around 1 and 0, respectively, if we switch the selected event, and therefore do not present such problems).[13] We aimed to examine the generalizability of pooled results as they are currently practised and reported.

Results

We made 1843 comparisons between OR, RR or RD of an individual RCT and the pooled OR, RR or RD of meta-analyses of all the other comparable RCT for various outcome variables extracted from 55 meta-analyses in the Cochrane Library (16 from psychiatry and 39 from general medicine). These included such diverse topics as antenatal thyroxin releasing hormone (TRH) prior to preterm delivery, antibiotics in salmonella, anticoagulation following non-embolic stroke, clozapine for schizophrenia and pharmacotherapy for dysthymia.

In terms of the total sample of comparisons made, all effect measures appeared to be reasonably and satisfactorily generalizable at around 90% concordance rates, but random effects model OR and RR produced the highest concordance rates (92%) (Table 1).

When we limited the comparison to those instances in which both individual RCT and their corresponding meta-analysis produced statistically significant outcomes, 412 comparisons were possible for each effect measure. Here fixed or random effects model OR and random effects model RR had concordance rates which were still close to 90%, while the risk difference demonstrated an appreciable drop in concordance.

When we further limited the comparisons to instances in which the CER of individual RCT was less than half or more than twice of that of the corresponding meta-analyses random effects model OR had the highest concordance rate (88%), closely followed by fixed effects model OR and the random effects model RR (87% and 84%, respectively). The concordance rate of the RD, both fixed and random effects model, showed marked declines from the values obtained for the total sample. The results were consistent when we examined more extreme cases where the CER was three-times different (Table 1).

We noted no particular clinical area where either the OR or the RR showed more than occasional inconsistency across studies. Out of 412 comparisons in which both RCT and meta-analysis had significant results, in only 17 instances (4%) was the discrepancy qualitative, i.e. an RCT produced an RR in the opposite direction from the random effects model RR of the meta-analysis of the other RCT. Instances in which qualitative differences were noted for a number of outcomes were very diverse and showed no common features that we can discern: antibiotics for treating salmonella gut infection, prophylactic surfactant in preterm infants, amodiapine versus chlorquinine in symptomatic patients with malaria, and clozapine versus typical antipsychotics in schizophrenia.

For a range of patient's expected event rates (PEER), point estimates of the individualized NNT calculated from fixed effects OR, random effects OR and random effects RR all produced good to excellent agreement, and were unlikely to lead to differing clinical decisions (Table 2).

Discussion

In this study, the random effects model OR showed the greatest consistency across RCT within meta-analyses, closely followed by the fixed effects model OR and random effects model RR. All of these measures of effect showed individual RCT results consistent with those of the other trials addressing the same question 82% or more of the time, even when the baseline risk differed substantially. On the other hand, the random or fixed effects model RD proved substantially less generalizable.

This degree of concordance for some measures of associations is surprising and encouraging, given that trials differ in the patients recruited, the way the interventions are administered, and the way the outcomes are measured, all of which can influence the size of the treatment effect. Publication bias could have inflated the concordance rates. This could occur if negative RCT, which would likely have RR qualitatively discrepant from those

Table 2 Agreement of point estimates of individualized number needed to treat (NNT) based on fixed effects odds ratios (OR), random effects OR and random effects risk ratios (RR)

PEER[a]	Fixed effects OR versus random effects OR	Fixed effects OR versus random effects RR	Random effects OR versus random effects RR
0.1	97%	82%	83%
0.2	98%	92%	93%
0.3	99%	90%	91%
0.4	99%	89%	90%
0.5	99%	80%	82%

[a] Patient's expected event rate.

Table 1 Concordance between each individual randomized controlled trial (RCT) and meta-analytic results of the remaining RCT among Cochrane meta-analyses

		Odds ratio		Risk ratio		Risk difference	
	No. of comparisons	Fixed	Random	Fixed	Random	Fixed	Random
All comparisons	1843	90%	92%	89%	92%	85%	88%
When both RCT and meta-analysis significant	412	86%	88%	82%	86%	72%	78%
When both significant and CER[a] doubly different	91	87%	88%	78%	84%	56%	65%
When both significant and CER three-times different	50	84%	84%	76%	82%	54%	64%

[a] Control event rate.

in positive RCT, were not published. In addition, systematic reviews that suggested substantial heterogeneity may not have been performed or published. We cannot know the extent of this bias.

Demonstrating similar treatment effects across differing groups of patients within a series of trials would provide the strongest support for assuming a constant RR or OR. For instance, assume that pooling results in the low, moderate, and high-risk patients who participated in a group of trials showed a similar magnitude of effect. Such a finding would provide very powerful evidence for applying a single OR or RR in calculating the likely benefit in all such patients. Unfortunately, such data are seldom available. The results of this study provide somewhat weaker, but still compelling, evidence that we may safely assume a similar magnitude of treatment effect when we want to individualize treatment decisions in separate groups of individuals or in an individual patient who may have a varying baseline risk.

On the other hand, we found that no effect measure was 100% applicable to all the possibly similar groups of patients. Our results apply only to the range of baseline risks seen in the studies that we included. The largest range was a 30-fold difference but the large majority was up to 5-fold difference; in 81% the difference was no greater than 2-fold, in another 10% no greater than 3-fold, and in another 5% no greater than 5-fold. Applying our results to greater differences in baseline risk is less secure. Examples are recently accumulating where OR and RR do appear to differ materially among subgroups of patients with differing baseline risks. They include anti-arrhythmic drugs after myocardial infarction,[14] carotid endarterectomy,[15] and human immunodeficiency virus infection.[16] Our results suggest that these cases represent exceptions, and that across various health interventions in humans, fixed effects model OR and random effects model OR or RR would be correct in eight to nine out of ten instances when applied to separate groups of individuals.

On the basis of our results, the best summary measure for a meta-analysis might be the random effects model OR. Clinicians could then use this OR to calculate PEER-adjusted NNT. Moreover, OR has some theoretical advantages over RR, because (1) it is symmetric around unity, (2) it does not predict impossible event rates if measure is assumed constant, (3) efficient estimation in small samples is available, (4) it can be easily expanded to a model with multiple factors and multiple levels, and (5) it can be estimated from any of the basic three epidemiological study designs (retrospective, cross-sectional or prospective).[17]

However, along with these mathematical properties, there are other factors to consider when recommending a summary measure for meta-analyses, such as the ease of interpretation and communication.[18] Clinicians find the OR difficult to interpret[19] and repeated examples show that even the most prestigious journals misinterpret OR as if they were RR.[20,21] This difficulty appears even greater when the result is to be used to obtain 'informed consent' from a patient.[22] The difference between OR and RR is large when the CER is moderate to high and/or when the OR and RR are much greater or smaller than 1.0, and misinterpreting OR as RR often ends up overestimating the benefits or harms of an intervention.[13] Furthermore, calculation of NNT from OR and PEER is arithmetically complicated.[23] On the other hand, our analyses suggested that point estimates of individualized NNT agree well if we calculate them from OR or RR.

Our results, and the additional considerations we have outlined, suggest the following approach to individualizing estimates of treatment benefit. First, the clinician should examine the available results to ensure that there is no evidence that relative risk varies substantially across risk groups. In the absence of such evidence, the clinician can safely use the random effects model RR to estimate PEER-adjusted NNT for individual patients they treat.

KEY MESSAGES

- When applying the results from a meta-analysis to individual patients, numbers needed to treat (NNT) are often calculated using the combined relative risk and the patient's expected event rate. This approach assumes that relative risks are constant across individual trials. This study examined the constancy of different summary effect measures.
- The odds ratio from fixed effects and random effects models and the relative risk from random effects models were reasonably constant across different baseline risks.
- The risk difference showed considerably less constancy, independent of the model used to combine the data.
- Given the interpretational ease of the relative risk, clinicians may want to rely on the relative risk from random-effects models.

References

[1] Glasziou P, Guyatt GH, Dans AL, Dans LF, Straus S, Sackett DL. Applying the results of trials and systematic reviews to individual patients. *Evidence-Based Medicine* 1998;**3:**165–66.

[2] Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd Edn*. Boston/Toronto/London: Little, Brown and Company, 1991.

[3] Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM. 2nd Edn*. Edinburgh: Churchill Livingstone, 2000.

[4] Mulrow CD, Cornell JA, Herrera CR, Kadri A, Farnett L, Aguilar C. Hypertension in the elderly. Implications and generalizability of randomized trials. *JAMA* 1994;**272:**1932–38.

[5] Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice & Teach EBM*. New York: Churchill Livingstone, 1997.

[6] Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998;**17:**1923–42.

[7] Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000;**19:**1707–28.

[8] Cochrane Collaboration. *Cochrane Library* [database on disk and CDROM]. Oxford: Update Software, Issue 3, 1998.

[9] Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993;**2:**121–45.

[10] Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140:**290–96.

[11] Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998;**351:**123–27.

[12] Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989;**8:**141–51.

[13] Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001.

[14] Boissel JP, Collet JP, Lievre M, Girard P. An effect model for the assessment of drug benefit: example of antiarrhythmic drugs in postmyocardial infarction patients. *J Cardiovasc Pharmacol* 1993;**22:**356–63.

[15] Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995;**345:**1616–19.

[16] Ioannidis JP, Cappelleri JC, Schmid CH, Lau J. Impact of epidemic and individual heterogeneity on the population distribution of disease progression rates. An example from patient populations in trials of human immunodeficiency virus infection. *Am J Epidemiol* 1996;**144:**1074–85.

[17] Walter SD. Choice of effect measure for epidemiological data. *J Clin Epidemiol* 2000;**53:**931–39.

[18] Deeks J. What is an odds ratio? *Bandolier* 1997;**2:**6–7.

[19] Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol* 1994;**47:**881–89.

[20] Hayes RJ. Odds ratios and relative risks [letter]. *Lancet* 1988;**ii:**338.

[21] Altman DG, Deeks JJ, Sackett DL. Odds ratio should be avoided when events are common. *Br Med J* 1998;**317:**1318.

[22] Feinstein AR. Indexes of contrast and quantitative significance for comparisons of two groups. *Stat Med* 1999;**18:**2557–81.

[23] Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence-Based Medicine* 1996;**1:**164–66.

International Journal of Epidemiology 2002;**31**:76–77

Commentary: Relative treatment effects are consistent across the spectrum of underlying risks … usually

Finlay A McAlister

To practice evidence-based therapeutics, clinicians have to integrate measures of efficacy and safety from the literature with their patient's unique risks and values.[1] To do this requires two assumptions. First, we assume that we can accurately estimate our patient's underlying baseline risk (or the 'expected event rate' referred to by Furukawa and colleagues).[2] This is no simple matter: while clinicians appear to be reasonably accurate in estimating the relative risks of different patients, even experienced clinicians perform poorly when estimating any one individual's absolute risk.[3] Methods for estimating the expected event rate for a particular patient have recently been reviewed and will not be considered further in this commentary.[1] The second common assumption in extrapolating from trials or meta-analyses to individual patients (who typically are at different risks from the 'average patient' in these studies), is that the relative effects of therapy are similar for patients at different risks. The study by Furukawa and colleagues[2] in this issue of the *International Journal of Epidemiology* adds to the emerging evidence supporting the validity of this assumption.

In discussing underlying or baseline risks in this commentary, I am not referring simply to the control event rate, but rather to patient characteristics (such as age, gender, disease aetiology, concomitant conditions, or disease status) present at baseline and known to impact prognosis for that particular disease. It is well recognized that any analyses relating treatment effects to control event rates in the same dataset will demonstrate a relation even if none exists (since the control event rate factors into both the expression for baseline risk and the expression for treatment effect).[4]

The best evidence for deciding whether treatment responsiveness differs across a spectrum of underlying risks would arise from individual patient data meta-analyses where the relative treatment effects in subgroups with widely varying risks can be directly compared. Although some such analyses have been conducted, they are few and far between and we are left to consider the second best approach to address this question: comparison of the relative treatment effects in different trials testing the same intervention (assuming that patients in different trials

The Division of General Internal Medicine, University of Alberta and The Institute of Health Economics, Edmonton, Alberta, Canada.

Correspondence: Dr F McAlister, 2E3.24 WMC, University of Alberta Hospital, 8440 112 Street, Edmonton, Alberta, Canada T6G 2R7. E-mail: Finlay.McAlister@ualberta.ca

with the same condition will have different risk profiles). The problem is how to find these groups of related trials—the easiest solution (employed by Furukawa and colleagues) is to review the reference lists of the rigorously conducted systematic reviews included in the Cochrane Collaboration database. I have described this approach as second best as selection bias may very well distort the findings (since a group of trials reporting widely divergent treatment effects are unlikely to be pooled to provide a single summary measure due to excessive heterogeneity). However, at this time it is the best we have.

Schmid *et al*.[5] provided the first such systematic approach by examining the relationship between effect measures and baseline risk in 115 meta-analyses (using a hierarchical model to account for the functional correlation between observed rates discussed earlier and random error in measurement of the control rate). They found that while the risk difference was significantly related to underlying risk in 31% of cases, in 87% of cases the relative risk (RR) (and in 86% of cases the odds ratio [OR]) did not vary significantly with the control event rate. Fukurawa and colleagues extend this work in a separate dataset by demonstrating high rates of concordance for both the OR and the RR in 1843 comparisons between individual trials and the summary effect measures derived from pooling all other trials in that topic area.[2] In particular, they found that the concordance rates were high for both OR and RR even when control event rates differed substantially (up to threefold) between trials. They found a qualitative discrepancy (where an individual trial reported an RR [or OR] in the opposite direction from the summary RR [or OR] for the other trials) in only 4% of cases but could not discern any features of these trials that suggested why such a discrepancy may occur. Thus, while both studies suggest that relative effect measures are constant across the usual spectrum of underlying risks in the vast majority of cases, neither study has advanced our understanding of when this assumption is unlikely to hold.

Sackett has hypothesized that relative treatment effects will be constant over the usual range of underlying risks for 'risk factor' interventions designed to slow the progress of disease, but that they will rise with increasing baseline risk for interventions designed to reverse the consequences of a disease process.[6] This latter situation would seem to apply particularly when the intervention has both positive and negative effects on the outcome of interest (for example, surgical procedures to prevent certain outcomes in the long-term usually expose patients to an increased risk of these same outcomes in the immediate peri-operative period). Thus, we would expect the relative risk reduction (RRR) associated with treatments such as angiotensin converting enzyme inhibitors in heart failure, beta-blockers in myocardial infarction, or thiazides for hypertension to be similar in patients with different underlying risks—indeed this is exactly what is seen.[7–9] On the other hand, we would expect the RRR associated with interventions such as carotid endarterectomy or coronary artery bypass grafting to be higher in patients at higher risk—again, exactly what is seen.[10,11] However, a word of caution: the validity of this hypothesis depends on the outcomes examined and is unlikely to hold for combined endpoints where the risk factor intervention impacts on only one of these endpoints.[12] For example, consider the example of cholesterol lowering agents. Although these drugs produce a consistent RRR in 'coronary events' and 'cardiac mortality' across different risk strata, their RRR for the combined endpoint of 'all-cause mortality' will vary and indeed be greater in 'high-risk' patients (such as those with established coronary disease) in whom a greater proportion of all deaths will be cardiac.[13]

Nevertheless, on the basis of the studies by Drs Schmid and Furukawa, it now seems reasonable to accept the assumption that relative treatment effects are consistent across the spectrum of underlying risks … usually.

Acknowledgements

FM is a Population Health Investigator of the Alberta Heritage Foundation for Medical Research. The author thanks D Sackett, W Taylor, R Roberts, B Haynes, PJ Devereaux, A Laupacis, S Walter, S Yusuf, J Sinclair, S Connolly, T Louis, and B Djulbegovic for their active participation and thoughtful insights in an e-mail discussion group on this issue and M Egger for helpful editorial comments.

References

[1] McAlister FA, Straus SE, Guyatt GH, Haynes RB, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature. XX. Integrating research evidence with the care of the individual patient. *JAMA* 2000;**283:**2829–36.

[2] Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002;**31:**72–76.

[3] Grover SA, Lowensteyn I, Esrey KL *et al*. Do doctors accurately assess coronary risk in their patients? Preliminary results of the coronary health assessment study. *Br Med J* 1995;**310:**975–78.

[4] Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *Br Med J* 1996;**313:**735–38.

[5] Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statist Med* 1998;**17:**1923–42.

[6] Sackett DL. Why RCTs fail but needn't. II. Failure to apply physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001;(in press).

[7] Flather MD, Yusuf S, Kober L *et al*. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. *Lancet* 2000;**355:**1575–81.

[8] The Beta-Blocker Pooling Project Research Group. The Beta-Blocker Pooling Project (BBPP): subgroup findings from randomized trials in post infarction patients. *Eur Heart J* 1988;**9:**8–16.

[9] Collins R, Peto R, MacMahon S *et al*. Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990;**335:**827–38.

[10] Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995;**345:**1616–19.

[11] Yusuf S, Zucker D, Peduzzi P *et al*. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists' Collaboration. *Lancet* 1994;**344:**563–70.

[12] Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. *Br Med J* 1994;**308:**72–74.

[13] Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *Br Med J* 1993;**306:**1367–73.