ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 151−157

Journal of Clinical Epidemiology

GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes

Gordon Guyatt[a,j,*], Andrew D. Oxman[b], Shahnaz Sultan[c], Jan Brozek[d], Paul Glasziou[e], Pablo Alonso-Coello[f], David Atkins[g], Regina Kunz[h,i], Victor Montori[j], Roman Jaeschke[k], David Rind[l], Philipp Dahm[m], Elie A. Akl[n], Joerg Meerpohl[o,p], Gunn Vist[b], Elise Berliner[q], Susan Norris[r], Yngve Falck-Ytter[r], Holger J. Schünemann[a]

[a]*Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[b]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs plass, 0130 Oslo, Norway*
[c]*Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Florida, Gainesville, FL, USA*
[d]*University of Oxford, Oxford, United Kingdom*
[e]*Iberoamerican Cochrane Centre, CIBERESP-IIB Sant Pau, Barcelona 08041, Spain*
[f]*QUERI Program, Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA*
[g]*Academy of Swiss Insurance Medicine (asim) University Hospital Basel Petergraben 4 CH-4031, Basel, Switzerland*
[h]*The Basel Institute of Clinical Epidemiology, University Hospital Basel Hebelstrasse 10, 4031 Basel, Switzerland*
[i]*Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA*
[j]*Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[k]*Harvard Medical School, UpToDate, Boston, USA*
[l]*Department of Urology, University of Florida, Gainesville, FL, USA*
[m]*Department of Medicine, State University of New York at Buffalo, Buffalo, NY, USA*
[n]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[o]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[p]*Technology Assessment Program, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, USA*
[q]*Oregon Health and Science University, Department of Medical Informatics and Clinical Epidemiology, Portland, OR 97239-3098, USA*
[r]*Division of Gastroenterology, Case and VA Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA*

Accepted 15 January 2012; Published online 27 April 2012

Abstract

GRADE requires guideline developers to make an overall rating of confidence in estimates of effect (quality of evidence—high, moderate, low, or very low) for each important or critical outcome. GRADE suggests, for each outcome, the initial separate consideration of five domains of reasons for rating down the confidence in effect estimates, thereby allowing systematic review authors and guideline developers to arrive at an outcome-specific rating of confidence. Although this rating system represents discrete steps on an ordinal scale, it is helpful to view confidence in estimates as a continuum, and the final rating of confidence may differ from that suggested by separate consideration of each domain.

An overall rating of confidence in estimates of effect is only relevant in settings when recommendations are being made. In general, it is based on the critical outcome that provides the lowest confidence.

Keywords: GRADE; Quality of evidence; Confidence in estimates; Guideline methodology; Systematic review methodology; Values and preferences

1. Introduction

In prior studies in this series devoted to exploring GRADE's approach to rating confidence in estimates of effect (quality of evidence) and grading strength of recommendations (guidance for practice) we have dealt with issues of framing the question [1]; introduced GRADE's

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the JCE Web site.

* Corresponding author. Gordon H. Guyatt, CLARITY Research Group, Department of Clinical Epidemiology & Biostatistics, Room 2C12, 1200 Main Street West Hamilton, ON, L8N 3Z5, Canada. Tel.: +905-525-9140; fax: +905-523-8781.

E-mail address: guyatt@mcmaster.ca (G. Guyatt).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.01.006

What is new?

Key points

GRADE requires a rating of confidence in effect estimates (quality of evidence) for each outcome.

Rating of confidence of evidence requires a gestalt that simultaneously considers all eight domains (risk of bias, precision, consistency, and so forth)

Guideline developers using GRADE will subsequently make an overall rating of confidence in effect estimates across all outcomes based on those outcomes they consider critical to their recommendation.

Optimal application of GRADE requires making the reasons for key judgments transparent.

conceptual approach to rating the confidence in a body of evidence [2]; and presented five reasons for rating down the confidence in effect estimates (risk of bias [3], imprecision [4], inconsistency [5], indirectness [6], and publication bias [7]) and three reasons for rating up the confidence in effect estimates [8] (a large magnitude of effect, a dose-response gradient, and a situation in which plausible biases, if present, would serve to increase our confidence in the effect estimate), as well as dealing with issues specific to resource use. This 11th article in the series will focus on (1) summarizing the confidence in effect estimates across a single outcome for each important or critical outcome and (2) determining the confidence in effect estimates across all critical outcomes.

2. Summarizing the confidence in effect estimates for individual outcomes

GRADE's approach to rating down (or not) with respect to each of five criteria and to rating up (or not) with respect to three others is sometimes straightforward and enhances the transparency of the system. Most commonly, authors will be comfortable with the rating of confidence in estimate of effect that results from considering each criterion separately. Not infrequently, however, if ratings are applied in a blanket or rote fashion without considering context and the relation of one criterion to another, the confidence rating could be problematic. Specifically, ratings of individual domains could result in an overall rating of confidence in effect estimates on a particular outcome that does not correspond well to an integrated assessment or the gestalt of confidence in estimates of effect. In such instances, an adjustment in the final rating based on that gestalt is required.

Consider a systematic review of randomized trials of flavonoids for the treatment of hemorrhoids that produced a pooled estimate of a relative risk of persisting symptoms (lack of improvement) of 0.42 (95% confidence interval [CI] 0.28−0.61) [9]. Table 1 presents an evidence profile summarizing the evidence regarding two outcomes: persisting symptoms and adverse effects of the intervention. The profile presents the number of studies and patients, considerations related to the five possible reasons for rating down confidence in effect estimates (summarized in the table with expansions in the associated footnotes), and the best estimates and CIs around relative and absolute effects.

Consider now the possible reasons for rating down confidence in effect estimates. In most studies, the published articles left uncertainty whether allocation was concealed (though blinding in most suggests the likelihood of concealment), and all studies used unvalidated measures of symptoms. Given these limitations, one could reasonably argue either for or against rating down for risk of bias.

Fig. 1 presents a forest plot depicting the results of the review. The point estimates from individual studies are quite variable, and some of the CIs overlap little. The test for heterogeneity is highly significant and the I^2 large. All these observations suggest rating down for inconsistency among studies. On the other hand, all point estimates suggest benefit, and one might argue that it is inappropriate to rate down for inconsistency when the only uncertainty appears to be whether the magnitude of the treatment effect is moderate or very large. For instance, if undesirable consequences of an intervention are minimal, even a modest treatment effect may warrant a strong recommendation in favor of that treatment. If, in such a circumstance, the basis of doubt is whether the true effect is modest or large, rating down for inconsistency may well be inappropriate.

All available randomized trials were of small or moderate size (from 40 to 234 patients), and all were industry funded. This is a situation that raises the possibility of publication bias. In addition, one could interpret the funnel plot as suggesting the possibility of publication bias, with three small, very positive studies and no corresponding studies with small or negligible effects (Fig. 2). This line of reasoning would suggest rating down confidence in the estimate for publication bias. On the other hand, the number of studies is insufficient to meet rigorous criteria for creating a funnel plot [10] and one could argue that the case for publication bias is speculative in which case one would not rate down.

Thus, for three of the five domains in which one might rate down confidence in effect estimates (risk of bias, inconsistency, and publication bias) one could reasonably make the case for rating down or for not doing so. The situation is further complicated by the magnitude of effect: the relative risk of persisting symptoms (0.41) is slightly less than 0.5, raising the possibility of rating confidence up for the magnitude of effect. A generous reviewer, who in each case is inclined to view the results favorably, would interpret the body of evidence from these flavonoid studies as high quality (i.e., would not rate down the quality). A less generous reviewer, who decides to rate down the

Table 1. GRADE Evidence Profile: flavonoids for patients with symptomatic hemorrhoids (question: flavonoids for patients with symptomatic hemorrhoids?; setting: outpatients)

Quality assessment						Summary of findings					
						No. of patients			Absolute risk		
No of studies (design)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	No treatment	Flavonoids	Relative risk (95% CI)	Control rate	Risk difference (95% CI)	Quality
Persisting symptoms/lack of improvement											
Nine (RCT)	Concealment not clear in most studies Outcome measures not validated[a]	P-value on test for heterogeneity < 0.0001 I^2 70.4[b]	No serious indirectness	No serious imprecision	All studies industry funded?[c]	218/384	93/398	RR 0.41 (0.27–0.62)	551/1,000	226 fewer per 1,000 (149–342)	Moderate quality because of publication bias[d]
Adverse effects											
13 (RCT)	Lack of concealment and unvalidated questionnaires[a]	No serious inconsistency	No serious indirectness	CI includes reduction to doubling of adverse effects[e]	All studies industry funded[c]	20/681	28/704	RR 1.22 (0.69–2.15)	60/1,000	Not significant	Low quality because of publication bias and imprecision

Abbreviations: CI, confidence interval; RR, relative risk; RCT, randomized controlled trial.

The table highlights the three questionable criteria in which reviewers might either rate down or not—study limitations, inconsistency, and publication bias—and how the final judgments could vary if one came to positive judgments on all three (e.g., for persisting symptoms, high-quality evidence) or negative judgments on all three (e.g., for persisting symptoms, very low-quality evidence).

[a] Allocation concealment unclear in most studies though blinding suggests the likelihood of concealment in most. The outcomes summarized here was failure to improve symptoms and side effects. These were measured by unvalidated questionnaires in each study. The questions, however, were simple and straightforward, final decision was not to rate down for risk of bias.

[b] Although the I^2 is large and the test for heterogeneity very highly significant, all studies but one suggest benefit, and uncertainty appears to be the magnitude of effect rather than whether there is an effect. Final decision not to rate down for inconsistency.

[c] Not only are all studies industry funded, but they are all of small or moderate size. Furthermore, the funnel plot (Fig. 2) could be interpreted as suggesting the possibility of publication bias. Final decision: rate down for likelihood of publication bias.

[d] We rated down for publication bias. Although there also was concern about a high risk of bias and inconsistency, we did not further rate down the quality of evidence because not every criterion appeared to justify rating down by one level.

[e] The lower boundary of the CI would suggest no treatment-induced adverse effects, whereas the upper boundary suggests more than a doubling of adverse effects relative to placebo.

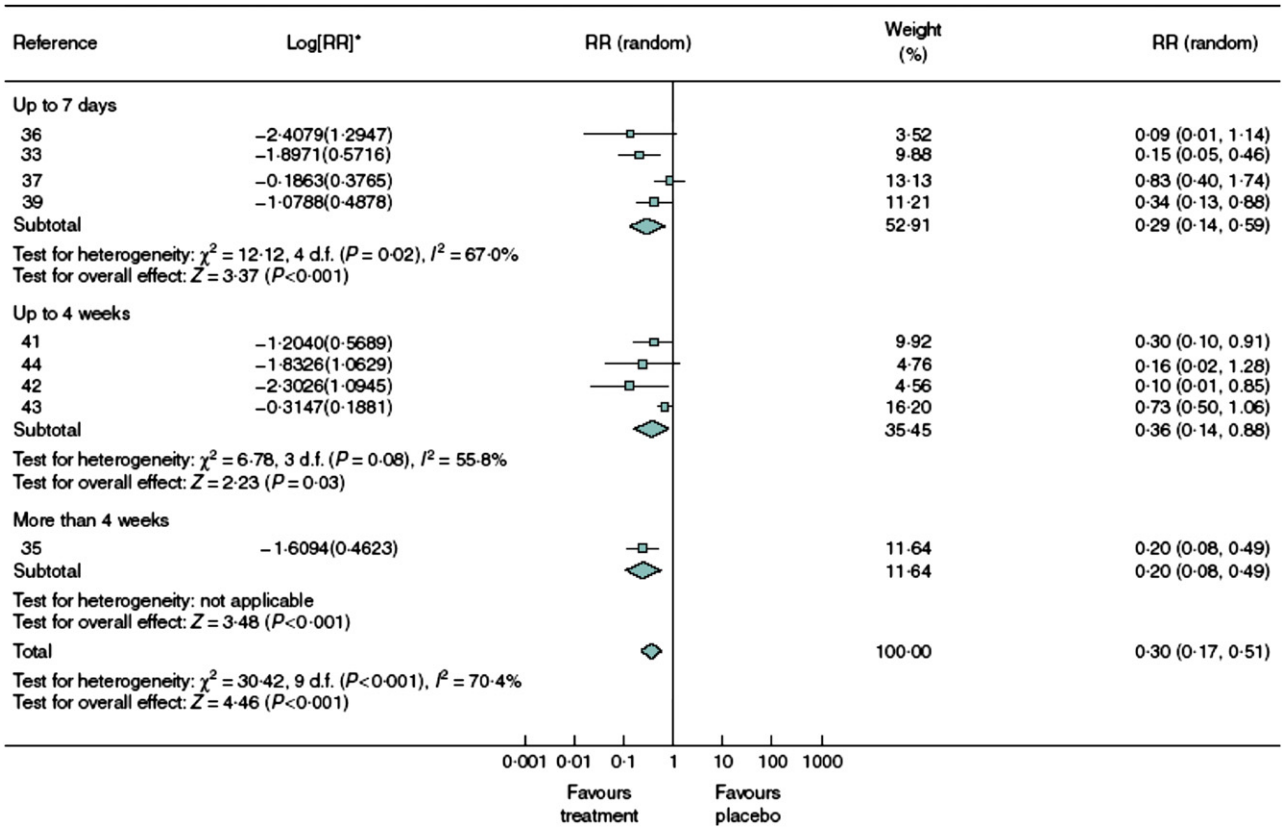

Fig. 1. Forest plot of the results of a systematic review of flavonoids for the treatment of hemorrhoids for the outcome of persisting symptoms or lack of improvement.

evidence in each case and rejects rating up for magnitude of effect, would judge the evidence warranting very low confidence. Both reviewers, having made judgments for individual criteria, might be dismayed that the overall rating (high or very low) does not really capture their confidence in effect estimates.

This example highlights the fact that each criterion for rating quality of evidence up or down reflects not discrete categories but a continuum from minimal limitations to very serious limitations. When the body of evidence is intermediate with respect to a particular criterion, the decision whether a study falls above or below the threshold for rating confidence up or down (by one or two levels) may be arbitrary. In such instances, it is particularly desirable to describe the rationale for the final decisions.

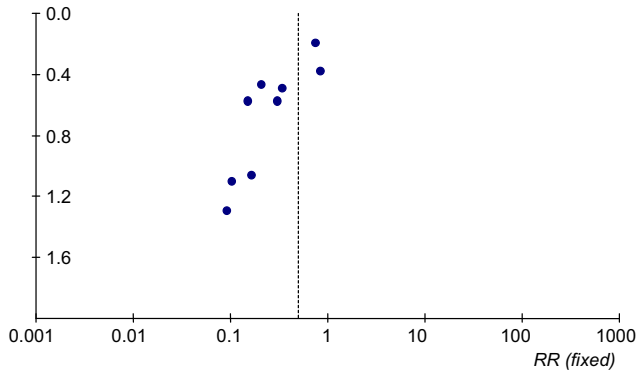

Fig. 2. Funnel plot of studies of flavonoids for ameliorating symptoms in patients with hemorrhoids.

In the case of flavonoids for hemorrhoids, both reviewers—charitable and harsh with respect to individual domains—may, taking a broad look at the evidence, agree that overall it lies on the border of moderate to low quality evidence (which was the conclusion of the authors of the review) [9]. In that case, reviewers may pick one or two domains (risk of bias, inconsistency, or publication bias) of limitations that would explain their reasoning. For example, the associated explanation could read: "We rated down for publication bias. Although there was also concern about a high risk of bias and inconsistency, we did not further rate down confidence in effect estimates because not every criterion appeared to justify rating down by one level." This reflects the necessity to take an overall or gestalt view of the body of evidence. In the evidence profile presentation (Table 1), the final decision is that the body of evidence warrants of moderate confidence, and the chosen reason for rating down confidence is likely publication bias.

Having difficulties about placing the evidence in either the moderate or low confidence category emphasizes that

the overall confidence rating also is a continuum, and contextual decisions are necessary when confidence is near the threshold between categories. The authors of the review acknowledged this by suggesting that ratings of either moderate or low confidence would be reasonable.

We encourage review and guideline authors to be explicit when they encounter similar situations, acknowledging borderline decisions in one or more domains. The evidence profile (Table 1) demonstrates such a presentation (note in particular footnote d).

Despite the limitations of breaking continua into discrete categories, treating each domain for rating confidence up or down as a discrete category enhances transparency. Indeed, the example highlights once again that the great merit of GRADE is not that it necessarily ensures reproducible judgments (observers will inevitably differ in close-call situations when rating up or down for individual domains or for the overall confidence per outcome) but that it achieves explicit and transparent judgment. In such close-call situations, apparent disagreement about whether to rate confidence up or down may represent very little disagreement on a continuum if that disagreement occurs near a threshold between categories (i.e., the threshold between rating down and not rating down). Furthermore, when the overall confidence is near a threshold (e.g., moderate or low confidence), systematic reviewers and guideline developers using GRADE may reduce their angst by recognizing that the disagreement, when the confidence rating is viewed as a continuum, is small.

3. Determining the confidence in effect estimates across outcomes

GRADE is the first formal system of rating quality of evidence to acknowledge that quality may differ across outcomes and to explicitly address this issue. For systematic reviews that are not associated with recommendations, and therefore do not require an overall confidence rating across outcomes, we suggest presenting confidence ratings for each important outcome and not determining the confidence in effect estimates across outcomes.

Such systematic reviews may, however, subsequently inform guidelines that do require implicit or explicit judgments about the overall confidence in effect estimates. It is better to be explicit, and it is logical that the overall confidence in effect estimates cannot be higher than the lowest confidence in effect estimates for any outcome that is critical for a decision. We therefore suggest applying the lowest confidence rating of the critical outcomes as the overall confidence associated with a recommendation. This requires distinguishing between outcomes that are critical and ones that are important but not critical.

Consider a systematic review of alternative strategies for Whipple resection for pancreatic cancer, one of which preserves the pylorus and the other, the standard approach, which does not [11]. The evidence in this review for different outcomes varied from moderate to very low confidence in effect estimates (Table 2). In cases such as this, guideline developers must consider whether undesirable consequences of therapy are important but not critical to the decision regarding the optimal management strategy or whether they are critical. If an outcome for which evidence is of lower quality is a critical outcome for decision making, then the rating of overall quality of the evidence must reflect this lower quality evidence. If the outcome for which confidence is lower is an important but not critical outcome, the overall rating will reflect the higher confidence in estimates from the critical outcomes.

Thus, for this example, if those making recommendations felt that gastric emptying problems were critical, the overall rating of the confidence in effect estimates would be very low. If gastric emptying were important but not critical, the overall confidence would be low (on the basis of results from the clearly critical perioperative mortality) despite the presence of moderate confidence regarding 5-year survival.

4. Which outcomes are critical may depend on the evidence

The overall confidence in effect estimates may not come from the outcomes judged critical at the beginning of the guideline development process—that is, judgments about what is critical may change when considering the results. For instance, a particular adverse event (e.g., severe nausea and vomiting) may be considered critical at the outset. If it turns out, however, that the event occurs very infrequently—say, less than 3% of patients—the final decision may be that the adverse effect is important but not critical.

Consider, once again, the flavonoids for hemorrhoids review (Table 1) [9]. In addition to the risk of bias (concealment not explicit, questionnaires not validated) and publication bias problems associated with the primary outcome of persisting symptoms, the adverse effect outcome suffers from imprecision. Therefore, whatever judgment of confidence one might make about persisting symptoms, adverse effects would warrant lower confidence. However, even assuming the boundary of the CI associated with the largest increase in adverse effects (an approximate doubling in comparison to placebo) represented the true impact of treatment, the adverse effects would still be relatively infrequent (approximately 6.3%) and minor in nature. Despite these considerations, some might consider the adverse effects critical and thus rate the overall confidence in effect estimates low. Others would not and may therefore rate the overall confidence in effect as moderate.

Consider the choice facing individuals without documented coronary heart disease (CHD) but at high risk (e.g., male smokers over 60 with hypertension, elevated

Table 2. GRADE Evidence Profile: different resection strategies for pancreatic carcinoma associated with different evidence quality of different outcomes (question: pylorus-preserving pancreaticoduodenectomy vs. standard Whipple pancreaticoduodenectomy in pancreatic or periampullary cancer?; setting: inpatients)

Quality assessment						Summary of findings					
						No. of patients			Absolute effect		
No of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	SWPD	PPPD	RR[a] (95% CI)	Control rate	Risk difference (95% CI)	Quality
Mortality at 5 years											
Three (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	94/114	93/115	RR 0.98 (0.87–1.11)	825/1,000	20 fewer per 1,000 (−120 to +80)	⊕⊕⊕○ Moderate
In-hospital mortality											
Six (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	12/244	4/246	RR 0.40 (0.14–1.13)	49/1,000	20 fewer per 1,000 (−50 to +10)	⊕⊕○○ Low
Biliary leaks											
Three (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	0/133	2/135	RR 4.77 (0.23–97.96)	0/1,000	20 more per 1,000 (−20 to +50)	⊕⊕○○ Low
Delayed gastric emptying											
Five (RCT)	Serious limitations[b]	Serious inconsistency[d]	No serious indirectness	Serious imprecision[c]	Undetected	56/220	66/222	RR 1.52 (0.74–3.14)	255/1,000	110 more per 1,000 (−80 to +290)	⊕○○○ Very low
Blood transfusions (units)[e]									Best estimate SWPD group	WMD (95% CI)	
Five (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	320		—[e]	2.45	WMD −0.66 (−1.16 to −0.25)	⊕⊕⊕○ Moderate
Hospital stay (days)[e]											
Five(RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	446		—[e]	19.17	WMD −1.45 (−3.28 to +0.38)	⊕⊕○○ Low

Abbreviations: SWPD, standard Whipple pancreaticoduodenectomy; PPPD, Pylorus-preserving pancreaticoduodenectomy; WMD, weighted mean difference.

[a] All data based on random effect models.

[b] Unclear allocation concealment in all studies, patients blinded in only one study, outcome assessors not blinded in any study, >20% loss to follow-up in three studies, not analyzed using intention to treat in one study.

[c] CI includes possible benefit from both surgical approaches.

[d] Unexplained heterogeneity; $I^2 = 72.6\%$, $P = 0.006$.

[e] Continuous outcome, therefore no relative effect is given.

cholesterol despite attempts at reduction with diet, diabetes, and a family history of CHD): should they use statins to lower their risk of cardiovascular events? A meta-analysis of rigorous randomized trials in such individuals demonstrated consistent, statistically significant reductions in major CHD events and stroke but nonsignificant reductions in CHD deaths [12]. Serious adverse effects were unusual, and all adverse effects were readily reversible with drug discontinuation [13].

Guideline developers considering a recommendation for or against statins in high-risk individuals are likely to start the process of arriving at a recommendation considering all four outcomes (i.e., death from cardiovascular causes, myocardial infarction, stroke, and adverse effects) as critical. In reviewing the evidence, they find that for three of the four outcomes (myocardial infarction, stroke, and toxicity) the evidence warrants high confidence. For CHD deaths, however, because of imprecision, evidence warrants moderate confidence. Should the overall confidence rating across outcomes be high or moderate?

The judgments made at the beginning of the review process suggest that the answer is ''moderate.'' Most patients, however, once it is established that their risk of stroke and major coronary events decreases with statins, would find compelling reason to use the medication. Whether CHD mortality decreases is (as long as it is very unlikely it increases) no longer relevant to the decision. Considering this, the overall confidence rating is most appropriately designated as high confidence.

The principle is that if there is higher confidence in some critical outcomes to support a decision in favor of an intervention (i.e., benefits on critical outcomes clearly outweigh undesirable effects of the intervention, for which there also is high-quality evidence) one need not rate down confidence because of lower confidence in other critical outcomes that support the same recommendation. To put it another way: an outcome is no longer critical if, across the range of possible effect of the intervention on that outcome, the recommendation or its strength would remain unchanged. Such judgments require careful consideration and are probably rare.

5. Conclusions

GRADE defines criteria for rating the confidence in effect estimates for a given outcome, thereby allowing systematic review authors and guideline developers to arrive at an outcome-specific confidence in effect estimates rating. Although this rating system represents discrete steps on an ordinal scale, it is helpful to view confidence in effect estimates as a continuum. An overall confidence in effect estimates rating across outcomes is only relevant in settings when recommendations are being made. In general, it is based on the critical outcome that provides the lowest confidence in effect estimates.

References

[1] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395−400.

[2] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401−6.

[3] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence−study limitations (risk of bias). J Clin Epidemiol 2011;64:407−15.

[4] Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence-imprecision. J Clin Epidemiol 2011;64:1283−93.

[5] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence-inconsistency. J Clin Epidemiol 2011;64:1294−302.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence-indirectness. J Clin Epidemiol 2011;64:1303−10.

[7] Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence-publication bias. J Clin Epidemiol 2011;64:1277−82.

[8] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011;64:1311−6.

[9] Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansdell D, Johanson JF, et al. Meta-analysis of flavonoids for the treatment of haemorrhoids. Br J Surg 2006;93:909−20.

[10] Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006;333:597−600.

[11] Karanicolas PJ, Davies E, Kunz R, Briel M, Koka HP, Payne DM, et al. The pylorus: take it or leave it? Systematic review and meta-analysis of pylorus-preserving versus standard whipple pancreaticoduodenectomy for pancreatic or periampullary cancer. Ann Surg Oncol 2007;14:1825−34.

[12] Thavendiranathan P, Bagai A, Brookhart MA, Choudhry NK. Primary prevention of cardiovascular diseases with statin therapy: a meta-analysis of randomized controlled trials. Arch Intern Med 2006;166:2307−13.

[13] Armitage J. The safety of statins in clinical practice. Lancet 2007;370:1781−90.