ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 158−172

Journal of Clinical Epidemiology

GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Nancy Santesso[a,b], Mark Helfand[d], Gunn Vist[c], Regina Kunz[e], Jan Brozek[a,b], Susan Norris[f], Joerg Meerpohl[g,h], Ben Djulbegovic[i,j,k], Pablo Alonso-Coello[l,m], Piet N. Post[n], Jason W. Busse[a,b,o], Paul Glasziou[p], Robin Christensen[q], Holger J. Schünemann[a,b]

[a]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada*
[b]*Department of Medicine, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada*
[c]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, Oslo 0130, Norway*
[d]*Oregon Evidence-based Practice Center, Oregon Health & Science University, Portland VA Medical Center, Portland, OR, USA*
[e]*Basel Institute of Clinical Epidemiology, University Hospital Basel, Hebelstrasse 10, Basel 4031, Switzerland*
[f]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*
[g]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg 79104, Germany*
[h]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, Freiburg 79106, Germany*
[i]*Center for Evidence-based Medicine and Health Outcomes Research, University of South Florida, Tampa, FL 33612, USA*
[j]*Department of Hematology, H. Lee Moffitt Cancer Center & Research Institute, 12901 Bruce B. Downs Boulevard, MDC02, Tampa, FL 33612, USA*
[k]*Department of Health Outcomes and Behavior, H. Lee Moffitt Cancer Center & Research Institute, 12901 Bruce B. Downs Boulevard, MDC02, Tampa, FL 33612, USA*
[l]*Iberoamerican Cochrane Center, CIBRESP - IIB Sant Pau, Barcelona 08041, Spain*
[m]*Epidemiology and Public Health CIBER (CIBERESP), Hospital de la Sant Pau, Creu i Sant Pau, Barcelona 08041, Spain*
[n]*Dutch Institute for Healthcare Improvement CBO, Utrecht, The Netherlands*
[o]*Institute for Work & Health, Toronto, Ontario M5G 2E9, Canada*
[p]*Centre for Research in Evidence-Based Practice, Faculty of Health Sciences, Bond University, Gold Coast, Queensland, 4229, Australia*
[q]*Musculoskeletal Statistics Unit, The Parker Institute, Copenhagen University Hospital, Frederiksberg, Denmark*

Accepted 30 January 2012; Published online 18 May 2012

Abstract

Summary of Findings (SoF) tables present, for each of the seven (or fewer) most important outcomes, the following: the number of studies and number of participants; the confidence in effect estimates (quality of evidence); and the best estimates of relative and absolute effects. Potentially challenging choices in preparing SoF table include using direct evidence (which may have very few events) or indirect evidence (from a surrogate) as the best evidence for a treatment effect. If a surrogate is chosen, it must be labeled as substituting for the corresponding patient-important outcome.

Another such choice is presenting evidence from low-quality randomized trials or high-quality observational studies. When in doubt, a reasonable approach is to present both sets of evidence; if the two bodies of evidence have similar quality but discrepant results, one would rate down further for inconsistency.

For binary outcomes, relative risks (RRs) are the preferred measure of relative effect and, in most instances, are applied to the baseline or control group risks to generate absolute risks. Ideally, the baseline risks come from observational studies including representative patients and identifying easily measured prognostic factors that define groups at differing risk. In the absence of such studies, relevant randomized trials provide estimates of baseline risk.

When confidence intervals (CIs) around the relative effect include no difference, one may simply state in the absolute risk column that results fail to show a difference, omit the point estimate and report only the CIs, or add a comment emphasizing the uncertainty associated with the point estimate.

Keywords: GRADE; Summary of findings; Evidence profile; Confidence in estimates; Quality of evidence; Systematic reviews

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 journal's Web site at www.elsevier.com.

* Corresponding author. CLARITY Research Group, Departments of Clinical Epidemiology & Biostatistics, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada. Tel.: 905-527-4322; fax: 905-523-8781.

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

0895-4356/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jclinepi.2012.01.012

1. Introduction

The first 11 articles in this series introduced the GRADE approach to systematic reviews and guideline development [1], discussed the framing of the question [2], and presented GRADE's concept of confidence in effect estimates [3] and how to apply it [4−9]. In this 12th article, we describe the final product of a systematic review using the GRADE process, Summary of Findings (SoF) tables that present, for each relevant comparison of alternative management strategies, the quality rating for each outcome, the best estimate of the magnitude of effect in relative terms, and the absolute effect that one might see across subgroups of patients with varying baseline or control group risks. The focus of this article is on binary outcomes. Box 1 presents the seven elements recommended for SoF tables. Tables 1−3, examples of SoF tables, highlight some of the issues in constructing such a table. Readers will find additional details in the Cochrane Handbook, Chapter 11 [10].

2. The seven elements of a SoF table

SoF tables include seven elements (Box 1). Uniformity of presentation is likely to facilitate readers' familiarity and comfort with SoF tables and is therefore desirable and facilitated by the use of GRADEpro software [11]. Initial user testing with consumers of guidelines (clinicians and researchers) guided the format of Table 1 [12,13]. In Table 1, putting what is most important first guided the order of the columns, and the presentation of absolute risks was guided by a finding that some respondents found presentation of risk differences confusing.

In addition, experimental evidence from a randomized trial of alternative formats suggests that some may find differing formats of SoF tables, such as that presented in Tables 2 and 3, preferable (Vandvik et al., unpublished data). In Table 2, the relative risk (RR) appears before the absolute risk on the basis that one uses the RR to calculate the absolute risk and, in both Tables 2 and 3, a column presents the absolute difference between groups. GRADEpro has been programmed to be responsive to these issues and has become increasingly flexible in accommodating alternative formats.

Uncertainty also exists regarding optimal terminology. Table 1 uses the term ''illustrative comparative risks'' and the designation ''assumed risk'' because uncertainty in the estimate of baseline risk is ignored in making the calculations. Some GRADE members believe that ''illustrative comparative risks'' might confuse, and other tables substitute ''absolute risk.'' The other tables use alternative designations for the control group and intervention group risks. Further study may provide additional information about the optimal wording choices.

Table 4 presents the full evidence profile associated with Table 1 addressing the desirable and undesirable consequences of wearing compression stockings on long plane rides. The table is atypical in that for some cells, which are shaded, it includes two sets of judgments, based on the same evidence—one of which is in regular type, the other in italics. The first is the judgment of Cochrane review authors [14]; the second (italicized) is the judgment of thrombosis experts in a guideline sponsored by the American College of Chest Physicians [15].

This example demonstrates that the great merit of GRADE is not that it eliminates judgments—and thus disagreements—but rather that it makes the judgments transparent. For the many close-call judgments that are required in evaluating evidence, disagreement between reasonable individuals will be common. GRADE allows readers to readily discern the nature of the disagreement. Decision makers are then in a position to make their own judgments about the relevant issues. The SoF table (Table 1) uses the judgments of the Cochrane reviewers.

3. Choosing which outcomes to present

SoF tables should ideally present results of all patient-important outcomes—possibly noting which ones are critical—without, however, overwhelming the reader. GRADE suggests inclusion of no more than seven outcomes, including both benefits and harms. If there are more than seven outcomes that are judged important, reviewers should choose the seven most important. This number is based on our intuition about the amount of information users can grasp, and an informal survey of attendees at a Cochrane Colloquium, and is therefore largely arbitrary. Limiting to seven may require combining related but different outcomes of approximately equal importance (e.g., calculating and presenting the number of patients who experienced either vomiting or diarrhea, considering these two as relatively equal minor gastrointestinal effects of temporary duration).

4. Presentation of direct vs. indirect evidence

Sometimes, direct measures of the patient-important outcomes are unavailable or, as in Table 1, no events have occurred (for symptomatic venous thrombosis and pulmonary embolism). In such instances, reviewers should present their inferences regarding treatment effects on patient-important outcomes on the basis of the results of surrogate measures. That the inferences are coming from surrogates should be clearly labeled, and will almost certainly result in rating down the confidence in effect estimates for indirectness.

What are the mechanics of making inferences regarding patient-important outcomes from surrogates? The simplest approach is to find a best estimate of the baseline risk for the patient-important outcome, and apply the relative effect from the surrogate (see Box 2 for an example of the arithmetic of applying an RR estimate to a baseline risk). For instance, in Table 1, to estimate the absolute reduction in

What is new?

Key points

Summary of Findings (SoF) tables provide succinct, easily digestible presentations of confidence in effect estimates (quality of evidence) and magnitude of effects.

SoF table should present the seven (or fewer) most important outcomes—these outcomes must always be patient-important outcomes and never be surrogates, although surrogates can be used to estimate effects on patient-important outcomes.

SoF table should present the highest quality evidence. When quality of two bodies of evidence (e.g., randomized trials and observational studies) is similar, SoF table may include summaries from both.

SoF table should include both relative and absolute effect measures, and separate estimates of absolute effect for identifiable patient groups with substantially different baseline or control group risks.

risk with stockings, we used an estimate of baseline risk from a meta-analysis and applied the RR from the surrogate, asymptomatic thrombosis.

Whenever the direct measure of a patient-important outcome is suboptimal (such as low-quality evidence) and a surrogate measure exists, reviewers have the option of focusing on whichever measure (the direct or surrogate measure) they feel yields higher-quality evidence or, as in Table 1, presenting both. As in Table 1, however, if they choose to focus partly or completely on surrogate results, reviewers must label the surrogate (in this case, asymptomatic venous thrombosis) for what it is, and include in its presentation the patient-important outcome for which it is a substitute (symptomatic thrombosis).

Another reason to present both direct and indirect measures is that the target audience for the review or guideline will want to see both. Table 3 presents an example of such a situation. Here, the review authors address the effect of low-intensity, pulsed, ultrasound on fracture healing [16]. Although one could argue that the single trial that directly addresses function provides the higher-quality evidence, the clinical community of relevance is likely to be (misguidedly perhaps) more interested in radiographic fracture healing (the surrogate outcome for function). Thus, for nonoperatively managed fractures and for operatively managed fractures, the investigators chose to present both direct evidence of functional improvement from one trial and indirect evidence from radiographic healing, despite the fact that the direct evidence was of higher quality because it did not suffer from indirectness (Table 3).

5. Presentation of randomized controlled trials or observational studies

Randomized controlled trials (RCTs) usually provide higher-quality evidence than observational studies and, if RCTs are available, SoF tables should generally restrict themselves to reporting RCT results. On occasion, however, limitations of RCTs or particular strengths of observational studies may lead to conclusions that their confidence in effect estimates is similar, or that observational studies provide higher-quality evidence.

For instance, consider the use of octreotide to prevent recurrent hypoglycaemia in patients with sulfonylurea overdose. Neither observational studies nor RCTs have addressed issues of mortality or long-term sequelae; thus, decisions must be based on the frequency of repeated hypoglycaemic episodes in the face of intravenous glucose administration.

The only RCT that addressed this issue administered a single dose of octreotide (the drug is ordinarily given as a continuous drip) [17]. Of those randomized to octreotide, 10 (45%) of 22 suffered recurrent hypoglycaemic episodes as did 6 (33%) of 18 control patients (RR 1.36, 95% confidence interval [CI] = 0.61−3.0). Three control, but no actively treated patients, suffered more than one recurrent hypoglycaemic episode. One would rate down confidence in estimates from this study for imprecision, and for indirectness of the intervention, suggesting an overall rating of low confidence in estimates.

At least 27 case reports have documented a marked decrease in hypoglycaemic episodes following octreotide administration [18,19]. Without untreated controls, these reports would be classified as very low-quality evidence but for

Box 1 Seven elements of a Summary of Findings table

1. A list of all important outcomes, both desirable and undesirable;
2. A measure of the typical burden of these outcomes (e.g. control group, estimated risk);
3. A measure of the risk in the intervention group or, alternatively or in addition, a measure of the difference between the risks with and without intervention;
4. The relative magnitude of effect;
5. Numbers of participants and studies addressing these outcomes;
6. A rating of the overall confidence in effect estimates for each outcome (which may vary by outcome); and possibly;
7. Comments.

Table 1. Summary of Findings table: Compression stockings compared with no compression stockings for people taking long flights

Patients or population: Anyone taking a long flight (lasting more than 6 hr)
Settings: International air travel
Intervention: Compression stockings[a]
Comparison: Without stockings

	Illustrative comparative risks[b] (95% CI)					
	Assumed risk	Corresponding risk				
Outcomes	**Without stockings**	**With stockings (95% CI)**	**Relative effect (95% CI)**	**Number of participants (studies)**	**Quality of evidence (GRADE)**	**Comments**
Symptomatic DVT	0 per 1,000	0 per 1,000 (−1.5 to 1.5)	Not estimable	2,637 (Nine studies)	⊕⊕⊕O *Moderate* due to imprecision[c]	0 participants developed symptomatic DVT in these studies.
Symptomatic DVT—inferred from surrogate, symptomless DVT	Low-risk population[d] 5 per 10,000	0.5 per 10,000 (0–1.25)	RR 0.10 (0.04–0.25)	2,637 (Nine studies)	⊕⊕⊕O *Moderate* due to indirectness[e]	
	High-risk population[d] 18 per 10,000	1.8 per 10,000 (1–8)				
Superficial vein thrombosis	13 per 1,000	6 per 1,000 (2–15)	RR 0.45 (0.18–1.13)	1,804 (Eight studies)	⊕⊕⊕O *Moderate* due to imprecision[f]	CI includes both benefit and harm
Edema, postflight values measured on a scale from 0, no edema, to 10, maximum edema	The mean edema score ranged across control groups from 6.4 to 8.9	The mean edema score in the intervention groups was on average 4.72 lower (4.91–4.52).		1,246 (Six studies)	⊕⊕OO *Low* due to risk of bias (unblinded, unvalidated measure)[g]	All these studies conducted by the same investigators. Extent of edema seems too great to be credible
Pulmonary embolus	0 per 1,000	0 per 1,000 −1.5 to 1.5	Not estimable	2,637 (Nine studies)	⊕⊕⊕O *Moderate* due to imprecision[c]	0 participants developed pulmonary embolus in these studies
Pulmonary embolus—inferred from surrogate, symptomless DVT	Low-risk population[d] 27 per million	3 per million (1–7)	RR 0.10 (0.04–0.25)	2,637 (Nine studies)	⊕⊕⊕O *Moderate* due to indirectness[e]	
	High-risk population 97 per million	10 per million (4–95)				
Death	Estimates not available, but risk extremely low		Not estimable	2,637 (Nine studies)	See comment	0 participants died in these studies, small proportion of pulmonary emboli would result in death

(*Continued*)

Table 1. Continued

Patients or population: Anyone taking a long flight (lasting more than 6 hr)

Settings: International air travel

Intervention: Compression stockings[a]

Comparison: Without stockings

	Illustrative comparative risks[b] (95% CI)					
	Assumed risk	Corresponding risk				
Outcomes	**Without stockings**	**With stockings (95% CI)**	**Relative effect (95% CI)**	**Number of participants (studies)**	**Quality of evidence (GRADE)**	**Comments**
Adverse effects	See comment	See comment	Not estimable	1,182 (Four studies)	⊕⊕OO *Low due to risk of bias (unblinded, unvalidated measure)*	The tolerability of the stockings was described as very good with no complaints of side effects in four studies[h]

Abbreviations: DVT, deep vein thrombosis; CI, confidence interval; RR, risk ratio; GRADE, GRADE Working Group grades of evidence (see explanations).

[a] All the stockings in the nine trials included in this review were below-knee compression stockings. In four trials, the compression strength was 20–30 mmHg at the ankle. It was 10–20 mmHg in the other four trials. Stockings come in different sizes. If a stocking is too tight around the knee, it can prevent essential venous return causing the blood to pool around the knee. Compression stockings should be fitted properly. A stocking that is too tight could cut into the skin on a long flight and potentially cause ulceration and increased risk of DVT. Some stockings can be slightly thicker than normal leg covering and can be potentially restrictive with tight footwear. It is a good idea to wear stockings around the house before travel to ensure a good, comfortable fitting. Stockings were put on 2–3 hr before the flight in most of the trials. The availability and cost of stockings can vary.

[b] The basis for the assumed risk is provided in footnotes. The corresponding risk (and its 95% CI) is based on the assumed risk in the intervention group and the relative effect of the intervention (and its 95% CI).

[c] The imprecision refers to absolute measures, not the relative. The decision to rate down presumes that people would value a very small reduction in venous thrombosis or pulmonary embolism. For the relative, it is not possible to make an estimate.

[d] Estimates for control event rates for venous thrombosis and for pulmonary embolism come from Philbrick JT, Shumate R, Siadaty MS, et al. Air travel and venous thromboembolism: a systematic review. J Gen Intern Med 2007;22:107–114. Definition of high risk includes previous episodes of DVT, coagulation disorders, severe obesity, limited mobility because of bone or joint problems, neoplastic disease within the previous 2 yr, large varicose veins.

[e] Here, there are two reasons for *indirectness*. One is that estimates of relative risk reduction come from the surrogate. The second is that there is uncertainty regarding the baseline risk.

[f] The CI includes both an increase and a small but possibly important decrease.

[g] The measurement of edema was not validated or blinded to the intervention. All of these studies were conducted by the same investigators.

[h] None of the other trials reported adverse effects, apart from four cases of superficial vein thrombosis in varicose veins in the knee region that were compressed by the upper edge of the stocking in one trial.

Table 2. Summary of Findings table—Should LMWH rather than VKAs be used for long-term treatment of VTE?[a,*]

Bibliography: Low molecular weight heparin compared with vitamin K antagonists for the long treatment of venous thromboembolism: a systematic review. Clive Kearon (unpublished)[b]

				Anticipated absolute effects *Time frame is 6 mo for all outcomes except PTS, which is 2 yr*	
Outcomes	**Participants (studies) follow-up**	**Quality of evidence (GRADE)**	**Relative effect (95% CI)**	**Risk with VKA**	**Risk difference with LMWH (95% CI)**
Overall mortality	2,496 (7 RCTs) 6 mo	⊕⊕O *Moderate* due to imprecision[c]	RR 0.96 (0.81–1.13)	164 deaths per 1,000[d]	No significant difference Seven fewer deaths per 1,000 (from 31 fewer to 21 more)
Recurrent VTE Symptomatic deep venous thrombosis and pulmonary embolism	2,727 (8 RCTs) 6 mo	⊕⊕⊕O *Moderate* due to risk of bias	RR 0.62 (0.46–0.84)	Low risk (no cancer) 30 VTEs per 1,000[d]	11 fewer VTE per 1,000 (from 5 fewer to 16 fewer)
				Moderate risk (nonmetastatic cancer) 80 VTEs per 1,000[d]	30 fewer VTE per 1,000 (from 13 fewer to 43 fewer)
				High risk (metastatic cancer) 200 VTEs per 1,000[d]	76 fewer VTE per 1,000 (from 32 fewer to 108 fewer)
Major bleeding	2,737 (8 RCTs) 6 mo	⊕⊕⊕O *Moderate* due to imprecision[e]	RR 0.81 (0.55–1.2)	Low to moderate risk (without or with cancer) 20 bleeds per 1,000[f]	No significant difference Four fewer bleeds per 1,000 (from nine fewer to four more)
				High risk (metastatic cancer) 80 bleeds per 1,000[f]	No significant difference 15 fewer bleeds per 1,000 (from 36 fewer to 16 more)
PTS Self-reported leg symptoms and signs	100 (1 RCT) median 3 mo	⊕⊕OO *Low* due to risk of bias and imprecision	RR 0.85 (0.77–0.94)	200 PTS per 1,000[g]	30 fewer per 1,000 (from 12 fewer to 46 fewer)

Abbreviations: CI, confidence interval; RR, risk ratio; PTS, Post-Thrombotic Syndrome; RCT, randomized controlled trial; LMWH, low molecular weight heparin; VKA, vitamin k antagonist; VTE, venous thromboembolism.

* The basis for the baseline risk (e.g., the median control group risk across studies) is provided in footnotes. The anticipated absolute effect is expressed as risk difference (and its 95% CI) and is based on the baseline risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[a] Limited to LMWH regimens that used 50% or more of the acute treatment dose during the extended phase of treatment.

[b] Meta-analysis is based on RCTs as referenced in the text of Kearon et al., [Chest 2012;Suppl:e419S-94]. The control group risk estimate for mortality comes from this meta-analysis.

[c] We did not rate down for risk of bias: borderline decision due to possible selective outcome reporting with one study not reporting deaths.

[d] Control group risk estimates come from cohort study by Prandoni 2002, adjusted to 6-mo time frame.

[e] We did not rate down for risk of bias although lack of blinded outcome assessment for major bleeds: borderline decision (we considered this outcome as not subjective).

[f] Control event rates from cohort studies by Prandoni 2002 and Beth 1995, adjusted to 6-mo time frame.

[g] Control event rate comes from observational studies in review by Kahn 2004, adjusted to 2-yr time frame. All patients wore pressure stockings.

Table 3. Summary of Findings table—RCTs of low-intensity pulsed ultrasound (LIPUS) for more rapid return to function (measured by direct measure and a surrogate—radiographic fracture healing)

Outcomes	No. of studies/ patients	Absolute effect		Relative effect (95% CI)	Quality
		Baseline risk	Mean difference (95% CI)		
Nonoperatively managed fresh fractures					
Return to function	1 RCT; 101 patients	15.1 days	Clavicle 1.95 days (−6.33 to 2.42)	13% Increase in time to return to work (6.0% decrease to 37.0% increase)	⊕⊕⊕O *Moderate* due to imprecision
Return to function inferred from surrogate—radiographic healing	3 RCTs; 158 patients	190 days	Tibia −88 days (−50.4 to −125.6)	36.9% Reduction in healing time (25.6% to 46.0%)	⊕⊕OO *Low* due to indirectness due to surrogate and risk of bias
		77 days	Radius −26 days (−6.4 to −38.6)		
		62 days	Scaphoid −18.8 days (−7.6 to −30.0)		
Operatively managed fresh fractures					
Return to function	2 RCTs[a]; 61 patients	79.1 days	Tibia −24.0 days (+14.3 to −62.3)	27.5% Reduction in time to full weight bearing (9.5% increase to 52.0% decrease)	⊕⊕OO *Low* due to risk of bias and imprecision
Return to function inferred from surrogate—radiographic healing	2 RCTs; 61 patients	132.5 days	Tibia −17.7 days (+69.8 to −105.2)	16.6% Reduction in healing time (76.8% increase to 60.7% decrease)	⊕OOO *Very low* due to risk of bias and imprecision and indirectness due to surrogate

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

[a] A third, negative, trial by Handolin et al. (2005c) reported on a functional outcome, mean Olerud-Molander score, but did not provide the associated measure of variance to allow for statistical pooling.

the apparently large and rapid effects (repeated hypoglycaemic episodes that markedly decreased or ceased after the administration of octreotide). Considering the magnitude and rapidity of effect, one might classify these case reports, in aggregate, as providing low-quality evidence.

Given similar quality evidence, it would be inappropriate to rely exclusively on either the RCT or the case reports in constructing a SoF table regarding the administration of octreotide for hypoglycaemia associated with sulfonylurea overdose. The results of case reports and the RCT appear inconsistent; the overall confidence in effect estimates could therefore be classified as low or very low.

There may be instances in which the confidence in estimates from the observational studies is clearly superior to that of RCTs; under these circumstances, one would restrict the SoF table to observational studies. When randomized trials clearly provide greater confidence in estimates, one would restrict the SoF table to randomized trials. In general, in situations in which both sets of studies provide important evidence with more or less equal confidence in estimates, we encourage review and guideline authors to summarize both types of studies in separate rows in their SoF tables as in Table 5.

6. Dealing with analytic approaches that yield different results

Systematic reviews, in exploring sources of heterogeneity, may sometimes find that alternative analyses (''sensitivity analyses'') yield appreciably different results. For example, a systematic review of glucosamine for treating osteoarthritis found differences in pain reduction when including only trials with concealed allocation vs. all trials [20]. Presenting two rows, one summarizing each analytic approach, would have left the inevitably less-equipped readers with the decision about which analysis is more credible. Rather, the authors focused on the analysis in which they had more confidence (in this case, restricted to trials with concealed allocation).

The authors did, however, note the alternative result in the ''comments'' column of the row in which they presented the pain results. This implies that they themselves had some uncertainty regarding which analysis was most credible, and wanted to alert readers to the alternative. Judgments of the credibility of alternative analyses require similar considerations to those of subgroup analyses, a topic we dealt with in a previous article in this series [6].

Table 4. Evidence profile: Compression stockings vs. no compression stockings for people taking long flights[a]

Quality assessment						Summary of findings					
						Number of patients			Absolute risk		
No of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Without compression stockings	With compression stockings	Relative risk (95% CI)	Control risk	Risk difference (95% CI)	Quality
Symptomatic DVT Direct evidence											
9 (RCT)	No serious limitations *Very serious limitations*[b]	No serious inconsistency	No serious indirectness	Serious imprecision	Undetected	0/1,323	0/1,314	Not estimable (no events)	0 per 1,000	0 per 1,000 (−1.5 to 1.5)	⊕⊕⊕O Moderate ⊕OOO *Very low*
Indirect evidence (based on symptomless DVT as a surrogate outcome for symptomatic DVT)											
9 (RCT)	No serious limitations *Very serious limitations*[b]	No serious inconsistency	Serious indirectness	No serious imprecision	Undetected	Surrogate Symptomless DVT 47/1,323	Surrogate Symptomless DVT 3/1,314	RR 0.10 (0.04–0.25)	5 per 10,000 18 per 10,000	Low risk 4.5 per 10,000 (4–5) High risk 16.2 per 10,000 (14–17.5)	⊕⊕⊕O Moderate ⊕⊕OO *Very low*
Superficial vein thrombosis											
8 (RCT)	No serious limitations *Serious limitations*[c]	No serious inconsistency	No serious indirectness	Serious imprecision	Undetected	12/901	4/903	RR 0.45 (0.18–1.13)	13 per 1,000	Results failed to show a difference between stockings and no stockings	⊕⊕⊕ Moderate ⊕⊕OO *Low*
Edema (postflight values measured on a scale from 0, no edema, to 10, maximum edema)											
6 (RCT)	Very serious limitations[d]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	7- or 8-hr flight Mean 6.4–6.9; 349 participants	Mean 2.2–2.4; 348 participants	—	Weighted mean difference: −4.72 (−4.91 to −4.52) Favors stockings		⊕⊕OO *Low*
						12-hr flight Mean 7.9–8.9; 272 participants	Mean 2.6–3.3; 277 participants				
Pulmonary embolus Direct evidence											
9 (RCT)	No serious limitations *Very serious limitations*[b]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	0/1,323	0/1,314	Not estimable (no events)	0 per 1,000	0 per 1,000 (−1.5 to 1.5)	⊕⊕⊕⊕ High ⊕⊕OO *Low*

(*Continued*)

Table 4. Continued

Quality assessment						Summary of findings					
						Number of patients			Absolute risk		
No of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Without compression stockings	With compression stockings	Relative risk (95% CI)	Control risk	Risk difference (95% CI)	Quality
Indirect evidence (based on symptomless DVT as a surrogate outcome for symptomatic DVT)											
9 (RCT)	No serious limitations *Very serious limitations*[b]	No serious inconsistency	Serious indirectness	No serious imprecision	Undetected	Surrogate Symptomless DVT 47/1,323	Surrogate Symptomless DVT 3/1,314	RR 0.10 (0.04–0.25)	27 per 1,000,000 97 per 1,000,000	Low risk; 24 per 1,000,000 (20–26) High risk; 87 per 1,000,000 (76–94)	⊕⊕⊕O Moderate ⊕OOO *Very low*
Adverse effects											
4 (RCT)	Very serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	0/1,182	0/1,182	Not available	The tolerability of the stockings was described as very good with no complaints of side effects.		⊕⊕OO *Low*

Abbreviations: DVTT, deep vein thrombosis; RCT, randomized controlled trial; RR, risk ratio.

[a] The footnotes from Table 1 apply here as well, but we have not repeated them.

[b] The thrombosis experts felt that lack of concealment, lack of blinding, and use of a technically inferior approach to ascertaining venous thrombosis constituted very serious limitations. The Cochrane group did not think these constituted serious limitations.

[c] Serious limitations included lack of concealment and lack of blinding.

[d] The Cochrane group felt that lack of concealment and blinding constituted very serious limitations in the context of an unvalidated edema rating and description of tolerability of stockings. The thrombosis experts did not address these outcomes.

Box 2 Calculations in Summary of Findings tables and evidence profiles

The RR of symptomatic deep vein thrombosis from nine RCTs is 0.10 (95% CI = 0.04–0.26).

The risk in the control group (estimated or assumed risk) from observational studies is 5 per 10 000.

Risk with intervention (*corresponding risk*) = Risk with control × RR

$= 5 \times 0.10$

= 0.5 per 10,000

Risk difference = Risk with control − risk with intervention

$= 5 - 0.5$

= 4.5 per 10,000

One uses exactly the same process to calculate the CIs around the risk difference, substituting the extremes of the CI (in this case 0.04 and 0.26) for the point estimate (in this case, 0.10). For instance, for the upper boundary of the CI:

Risk with intervention = 5 × 0.26 = 1.3 per 10,000

Risk with control − risk with intervention = 5 − 1.3 = 3.7 per 10,000

Risk difference = (0.74 × 5)/10,000 = 3.7/10,000.

7. Measures of relative effect

Options for expressing relative measures of effect include the RR (synonym: risk ratio), odds ratio (OR), rate ratio, and hazard ratio [21–23]. ORs have advantageous statistical properties [24]. RRs, however, are more understandable intuitively, and easier to use for estimating absolute measures of effect in individual patients [21]. We find these advantages of RRs compelling (for more details, see Box 3). Meta-analysis can generate RRs or ORs from 2 × 2 tables using appropriate statistical techniques [22,23].

Using hazard ratios requires time-to-event data and relatively complex analytic approaches [25,26]. Time-to-event data will—at least outside of cancer studies—seldom be available for an entire group of studies that inform a particular clinical question. Moreover, hazard ratios are less familiar to clinicians (again, with the exception of clinicians focused on cancer), and are always farther from 1.0 than are RRs. Thus, clinicians familiar with RRs for a wide variety of interventions may overestimate the magnitude of effect when presented with a hazard ratio for a particular intervention.

A special case of reporting data that, in theory, can be considered continuous are counts of events per patients (e.g., the number of disease exacerbations per patient or the number of new polyps per patient in one group compared

Table 5. Summary of Findings table: Use of octreotide in patients with sulfonylurea overdose

Outcomes	Participants (studies) follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no octeotride	Risk difference with octeotride
Recurrent hypoglycemia from randomized trials	1 RCT, 40 patients in emergency room	⊕⊕OO *Low* due to imprecision and indirectness	RR 1.36 95% CI 0.61–3.0	33%	No significant difference 7 fewer deaths per 1,000 (from 31 fewer to 21 more)
Persistent hypoglycemia from observational studies	27 case reports	⊕⊕OO *Low* from observational studies. Would be very low with no control, but effects large and rapid in some reports	All reported decrease in hypoglycemia following octeotride administration	All patients had persistent hypoglycemia	All reported decrease in hypoglycemia following octeotride administration

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

with another). When events are rare, the analyses often focus on rates. Rates relate the counts to the amount of time during which they could have happened. For example, the result of one arm of a clinical trial could be that investigators counted 20 exacerbations of chronic obstructive pulmonary disease in 100 patients during a period of 300 person-years of follow-up. The rate associated with this result would be 0.067 per person-year or 6.7 per 100 person-years. To summarize such findings, investigators use the rate ratio in meta-analyses that compare the rates of events in the two groups by dividing one by the other. Table 6 provides an example of such a situation. When events become more frequent, investigators may treat the data as a continuous outcome.

8. Measures of absolute effect

As we have pointed out, relative measures tend to be consistent across risk groups, whereas absolute measures do not [22,27–29]. Making management choices, however, focuses on trading off absolute effects on patient-important outcomes, therefore requiring both relative and absolute measures to appear in SoF tables.

The unrepresentativeness of patients in randomized trials, and the lack of consistency of absolute measures across risk groups and across individual trials argue against direct calculation of pooled risk differences from the data in randomized trials. The alternative process begins with selection of a baseline risk (control group risk): ideally this would come from well-designed observational studies. For instance, the baseline risk for symptomatic deep venous thrombosis and for pulmonary embolism in Tables 1 and 4 come from a systematic review summarizing the results of observational studies [30]. Box 2 shows the calculations involved in generating absolute differences from baseline risks and RRs using the outcome of venous thrombosis from Table 1.

Using ORs provides an alternative with advantages and disadvantages (Box 3). As a guideline developer, one may have only the OR available, or as a systematic review author, one may choose to use the OR. In either circumstance, using the OR to generate an estimate of risk difference involves converting baseline risk to odds, multiplying by the OR, and converting the resulting odds back to risks. Alternatively, one can use the following formula (where RC is the risk in the control group):

$$\text{Risk difference per } 1,000 = 1,000 \times \text{RC} - \left(\frac{\text{OR} \times \text{RC}}{1 - \text{RC} + (\text{OR} \times \text{RC})}\right)$$

Unfortunately, high-quality observational studies are often unavailable. Typical limitations include suboptimal surveillance for outcomes, and potentially biased ascertainment of outcomes. If high-quality observational studies are not available, we suggest using the median risk (rather than the weighted average) among the control groups in the included studies or, if it is available, the control group risk from a single trial with far larger sample size than other available trials. If there is important variation in control group risks, authors should consider presenting a range of risks within that observed in the included studies (that is, present a range of baseline risks). One then applies the RR to two or more baseline risks to generate possible intervention group risks.

Absolute effects are likely to differ across patient groups. Data from observational studies (and occasionally from randomized trials) may allow reliable identification of subgroups at substantially different risk of adverse outcomes. If such data allows clinicians to readily identify these subgroups by their presenting clinical features, review authors should present absolute risks for intervention and control groups (and/or differences in risk between intervention and control groups) for each of these prognostic subgroups. Therefore, if authors find moderate- or high-quality evidence regarding clinical features that reliably distinguish between patients at substantially different risk of the outcomes of interest, they should use the baseline risk in these patient groups, along with the RR, to generate

Box 3 Should review authors use RRs or ORs?

RRs and ORs tend (in contrast to risk differences) to be similar across risk groups. ORs have statistical properties that are superior to those of RRs, which become particularly apparent when one uses these relative measures to generate absolute effects (risk differences—see Box 2). One is that the OR leads to the same risk difference whether one counts events in a negative or positive way, whereas RR do not. For example, RRs will yield different results in translating to risk differences if one considers mortality (e.g., 20% die) or survival (e.g., 80% survive). A second is that use of RR can generate impossible values of risk (i.e., outside of the range of 0−1.0). For instance, if one applies a RR of 1.2 from a meta-analysis to a baseline risk of 90% the result is an impossible intervention group risk of 1.08. ORs always generate risks of 0−1.0.

Conversely, as baseline risk of undesirable outcomes increases above 50% with the use of RR, the risk difference increases (as it should intuitively), whereas the risk difference using the OR decreases (counterintuitively). This is the price we pay for having the same risk difference whether one frames the issue using the desirable (e.g., survival) or undesirable (e.g., death) outcome.

Choosing either OR or RR is easily defensible. The authors of this article prefer RR because of ease of interpretation, and ease of use for generating risk differences (see Box 2). RRs may, however, be problematic when RRs are greater than 1 and high baseline risks may occur (e.g., a baseline risk of 67% or more with a RR $>$ 1.5) resulting in intervention group probabilities greater than 1.0. RRs may also be problematic when positive or negative framing may be considered reasonable (e.g., death or survival when mortality over 50%; symptoms as improved or unimproved). Under these circumstances, ORs may be preferable.

expected risks with the intervention. Box 4 describes considerations that arise when risks differ across patient groups.

9. Presentation of absolute effects

We suggest presenting the absolute effect—both benefits and harms—as natural frequencies (events per 10,000 patients in Table 1, although more frequent events can be presented as events per 1,000 or even per 100 patients) because this facilitates decision making [31−34]. When events are sufficiently frequent, percentages may be as well, or marginally better, understood [35]. Although many clinicians prefer numbers needed to treat (NNTs), they may be more difficult to interpret when it is necessary to consider multiple outcomes. Reporting NNTs may be particularly appropriate in abstracts, or in summary tables with only two to three outcomes; natural frequencies or percentages are likely to be more easily interpretable in other contexts. Review and guideline authors may want to tailor their presentations to the specific audiences they are addressing; differing formats may be optimal for differing audiences. Whatever choice is made, the presentation should be consistent across all outcomes in a single SoF table. This need for consistency also applies with regard to dealing with presentation of absolute effects when relative effects are very imprecise (Box 5).

10. Absolute effects—confidence intervals

We further suggest reporting the CIs around the absolute risk in the intervention group (as in Tables 1 and 6) or around the difference between intervention and control groups (as in Tables 2−5). Just as one calculates the absolute risk in the intervention group on the basis of the absolute risk in the comparison group and the point estimate of the RR, the calculation of the CIs around the absolute risks in the intervention group is based on the absolute risk in the comparison group and the CIs around the RR. When the baseline risk is very low, however, CIs calculated on the basis of RRs may be misleading. Under these circumstances, direct calculations based on absolute risks are preferable [36].

RevMan provides options for calculations of RR or OR (from which one can estimate risk differences—see Box 2 and, for ORs, text in ''Measures of absolute effect'') or, for situations when baseline risk is very low, direct calculation of risk differences.

11. Absolute effects—choice of time frame

In Table 1, the time frame for measurement of outcome is both obvious and short—symptomatic thrombosis, if it exists, will occur within days of a long flight. For conditions such as primary and secondary prevention of cardiovascular events, or cancer recurrence, there are options for choice of the duration of follow-up. Reviewers should therefore always indicate the length of follow-up to which the estimates of absolute effect refer. Note, this length of follow-up may not be the length of follow-up in the RCTs that generated the estimates of relative effect, or the observational studies or RCTs that led to estimates of baseline risk. Rather, it will be some time frame judged appropriate for balancing the desirable and undesirable consequences of alternative management strategies.

Longer follow-up periods are associated with higher absolute risks and higher risk differences between intervention and control. This can lead to potentially important

Table 6. Summary of Findings table—Presenting less common outcome measures: rate ratios and quality-of-life data

Combined corticosteroid and long-acting beta-agonist in one inhaler for chronic obstructive pulmonary disease

Patient or population: patients with moderate and severe chronic obstructive pulmonary disease
Settings: outpatient
Intervention: corticosteroid and long-acting beta-agonist in one inhaler[a]
Comparison: no treatment

Outcomes	Absolute risks[b] (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Estimated control risk; no treatment	Corresponding risk; combined inhaler				
Exacerbation rate (follow-up: 3 yr)	The mean exacerbation rate in the control groups was 3 exacerbations in 3 yr[c]	The mean exacerbation rate in the intervention groups was 2 exacerbations in 3 yr[c]		4,226 (5)	⊕⊕⊕O Moderate[d]	Rate ratio 0.74 (0.69, 0.79)
Hospitalizations	See comment	See comment	Not estimable	0 (0)	See comment	Limited data for hospitalizations was presented in the trials.
Mortality (follow-up: 3 yr)	Medium risk population[c] 15 per 100	12 per 100 (10–14)	OR 0.79 (0.65–0.96)	5,752 (7)	⊕⊕⊕⊕ High	
Quality of life St. George's Respiratory Questionnaire Scale from: 0 to 100 (follow-up: 3 yr)	The mean quality of life in the control groups was 48 points[c]	The mean quality of life in the intervention groups was 2.90 lower (3.61 to 2.18 lower)		3,346 (4)	⊕⊕⊕⊕ High	Mean difference did not reach a patient important improvement of 4 points.
Pneumonia (follow-up: 3 yr)	Medium risk population[c] 12 per 100	20 per 100 (17–23)	OR 1.83 (1.51–2.21)	5,739 (8)	⊕⊕⊕⊕ High	
Any adverse events (follow-up: 3 yr)	Medium risk population[c] 90 per 100	91 per 100 (90–92)	OR 1.10 (0.96–1.27)	5,493 (8)	⊕⊕⊕⊕ High	Data from fluticasone/salmeterol studies.

Abbreviations: CI, confidence interval; OR, odds ratio.

[a] Both long-acting beta-agonists and inhaled corticosteroids can be used in combination for the treatment of chronic obstructive pulmonary disease. Of the 11 included studies, two evaluated fluticasone/salmeterol at 250 mcg/50 mcg twice daily and seven at 500 mcg/50 mcg twice daily; and two evaluated budesonide/formoterol at 320 mcg/9 mcg twice daily. All studies permitted the use of inhaled short-acting beta-agonists on demand.

[b] The basis for the risk in untreated patients (e.g., the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% CI) is based on the risk in the untreated patients and the relative effect of the intervention (and its 95% CI).

[c] Risk in untreated patients based on TORCH trial.

[d] Withdrawal of participants with severe frequent exacerbations may have biased results.

differences in readers' perceptions of the apparent magnitude of effect (Box 6). Often, extending the time frame involves the assumption that event rates will stay constant over time.

12. Dealing with no events in either group

When no participant in any trial has suffered the outcome of interest, the trials provide no information about relative effects (and one can thus argue that there is no point in rating the quality of the evidence). However, particularly if there are large numbers of patients, the data may provide high-quality evidence that the absolute difference between alternative management strategies is small or very small. If reviewers believe this is the appropriate inference for an important or crucial outcome, they can rate the confidence in effect estimates, and base the estimate of precision on the CI around the absolute effect (as in Tables 1 and 4). A program to make the calculation based on the available statistical methods [37] is available from the corresponding author.

13. Uncertainty around estimates of baseline risk

Note that Table 1 provides estimates of risk in the intervention group based on the CIs around the RR. We do not,

Box 4 Differing risks across different patient groups

In Table 1, reviewers identified risk factors for asymptomatic DVT (previous episodes of DVT, coagulation disorders, severe obesity, limited mobility because of bone or joint problems, cancer, and large varicose veins) that, when considered together, more than tripled the risk of thrombosis. Applying the RR of 10% allowed calculation of expected event rates for the high- and low-risk populations using prophylactic stockings. In the low-risk population, applying the RR of 10% to the risk without the intervention of 5 per 10,000 generates a risk of 0.5 per 10,000 with the intervention. In the higher-risk population, the corresponding numbers are 18 and 1.8 per 10,000. Table 3 presents another such example for the outcomes of venous thrombosis (three risk strata) and bleeding (two risk strata).

Reference

[1] Philbrick JT, Shumate R, Siadaty MS, Becker DM. Air travel and venous thromboembolism: a systematic review. J Gen Intern Med. 2007;22: 107−114.

however, provide estimates of uncertainty regarding the estimates of baseline risk in high- and low-risk control groups. Not presenting such estimates reflects a high priority on simple presentations that clinicians and patients will find easily digestible.

Box 5 Presenting absolute effects when estimates of relative effect are imprecise

When CIs around the relative risks are wide (including both benefit and large harm) providing a point estimate for the intervention that differs from that of the comparator, or a CI around a risk difference, may give the impression of an effect that does not exist. If reviewers or guideline developers share this concern, in the absolute risk difference column (or the intervention group risk column, depending on the format chosen) they may choose to state only that the result failed to show a difference between intervention and control; omit the point estimate and report only the CIs; or add a comment emphasizing the uncertainty associated with the point estimate (or some combination of the three strategies). Note that in Table 1 for superficial vein thrombosis, we present estimates of absolute effect and include a comment that notes that the CI includes both benefit and harm. In Table 4, which uses the same data, we do not provide absolute estimates, but merely note that the result fails to show a difference.

Box 6 The impact of choice of time frame on readers' perceptions of effect

Consider primary prophylaxis with aspirin for the prevention of myocardial infarction (MI) in asymptomatic individuals with risk factors for development of coronary disease (so-called high risk). Estimates of risk of MI in such individuals—despite the high-risk label—is very low, approximately 6 per 1,000 per year [1]. The benefits of regular use of aspirin are correspondingly low—between one and two MIs—prevented per 1,000 patients taking aspirin over the course of a year [1]. Given that aspirin is associated with an increased risk of gastrointestinal bleeding, few would be enthusiastic about this magnitude of benefit. If one considers a time frame of a decade, however, aspirin use will prevent approximately 14 MIs per 1,000 patients (an absolute benefit of 1.4%). This latter framing potentially makes the intervention appear more attractive.

Reference

[1] Baigent C, Blackwell L, Collins R, Emberson J, Godwin J, Peto R, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomized trials. Lancet. 2009;373(9678):1849−1860.

Potentially, all the issues that raise uncertainty about estimates of absolute effects could raise uncertainty about estimates of baseline risks: risk of bias, indirectness if surrogate measures are used, imprecision, inconsistency, and publication bias. GRADE has chosen to thus far more or less ignore uncertainty in estimates of baseline risk in its criteria for rating confidence in effect estimates. This is a pragmatic decision that avoids overwhelming complexity and keeps the systematic review manageable.

Nevertheless, guideline developers should be aware of this neglected source of uncertainty, and in certain circumstances may wish to include it in considerations about confidence in effect estimates for individual outcomes. When such considerations arise, we suggest classifying them under ''indirectness.'' Presenting a plausible range of baseline risks may, to some extent, ameliorate the problem.

14. What to do when there is no published evidence regarding an important outcome

We encourage systematic review authors and guideline developers to specify all important outcomes before

commencing their reviews. If they do so, it is possible that they may find no published evidence regarding one or more outcomes (quality of life and rare side effects are two outcomes that may be subject to this problem). We suggest that if sufficiently important, such an outcome would warrant a row in the SoF table, with the confidence in effect estimates rating (and other cells aside from the comments) being either left blank or classified as very low-quality evidence.

15. Conclusion

The SoF table provides all the key information necessary for making decisions between competing health care management strategies [38]. Therefore, although not an absolute requirement for use of the GRADE approach, the SoF table is an invaluable tool for providing a succinct, accessible, transparent evidence summary for patients, health care providers, and policy makers.

References

[1] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94.

[2] 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.

[3] 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.

[4] 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;6:407−15.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Schünemann H, Oxman A, Higgins J, Vist G, Glasziou P, Guyatt G. Presenting results and 'Summary of findings' tables. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 500. Chichester, UK: Wiley; 2008.

[11] Brozek J, Oxman A, Schünemann H. GRADEpro. [Computer program]. Version 3.2 for Windows. Available at http://www.cc-ims.net/gradepro or http://mcmaster.flintbox.com/technology.asp?page=3993. 2008. Accessed February 29, 2012.

[12] Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. J Clin Epidemiol 2010;63:620−6.

[13] Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. J Clin Epidemiol 2010;63:607−19.

[14] Clarke M, Hopewell S, Juszczak E, Eisinga A, Kjeldstrom M. Compression stockings for preventing deep vein thrombosis in airline passengers. Cochrane Database Syst Rev 2007;3.

[15] Geerts WH, Bergqvist D, Pineo GF, Heit JA, Samama CM, Lassen MR, et al. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). Chest 2008;133(6 Suppl):381S−453S.

[16] Busse JW, Kaur J, Mollon B, Bhandari M, Tornetta P 3rd, Schunemann HJ, et al. Low intensity pulsed ultrasonography for fractures: systematic review of randomised controlled trials. BMJ 2009;338:b351.

[17] Fasano CJ, O'Malley G, Dominici P, Aguilera E, Latta DR. Comparison of octreotide and standard therapy versus standard therapy alone for the treatment of sulfonylurea-induced hypoglycemia. Ann Emerg Med 2008;51:400−6.

[18] Carr R, Zed PJ. Octreotide for sulfonylurea-induced hypoglycemia following overdose. Ann Pharmacother 2002;36:1727−32.

[19] Crawford BA, Perera C. Octreotide treatment for sulfonylurea-induced hypoglycaemia. Med J Aust 2004;180(10):540−1. author reply 1.

[20] Towheed T, Maxwell I, Abastassuades T, Shea B, Houpt J, Welch V, et al. Glucosamine therapy for treating osteoarthritis. Cochrane Database Syst Rev 2009;4.

[21] Walter SD. Choice of effect measure for epidemiological data. J Clin Epidemiol 2000;53:931−9.

[22] Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 2002;21:1575−600.

[23] Deeks J, Higgins J, Altman D. Analyzing data and undertaking meta-analyses. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 500. Chichester, UK: Wiley; 2008.

[24] Eckermann S, Coory M, Willan AR. Consistently estimating absolute risk difference when translating evidence to jurisdictions of interest. Pharmacoeconomics 2011;29:87−96.

[25] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815−34.

[26] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16.

[27] 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.

[28] 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(1):72−6.

[29] 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(13):1707−28.

[30] Philbrick JT, Shumate R, Siadaty MS, Becker DM. Air travel and venous thromboembolism: a systematic review. J Gen Intern Med 2007;22:107−14.

[31] Gigerenzer G. The psychology of good judgment: frequency formats and simple algorithms. Med Decis Making 1996;16(3):273−80.

[32] Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ 2003;327(7417):741−4.

[33] Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. Acad Med 1998;73:538−40.

[34] Galesic M, Gigerenzer G, Straubinger N. Natural frequencies help older adults and people with low numeracy to evaluate medical screening tests. Med Decis Making 2009;29:368−71.

[35] Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. Ann Intern Med 2011;155(2):87−96.

[36] Murad MH, Montori VM, Walter SD, Guyatt GH. Estimating risk difference from relative association measures in meta-analysis can infrequently pose interpretational challenges. J Clin Epidemiol 2009;62: 865−7.

[37] Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998;17:873−90.

[38] Djulbegovic B, Soares H, Kumar A. What kind of evidence do patients and practitioners need: evidence profiles based on 5 key evidence-based principles to summarize data on benefits and harms. Cancer Treat Rev 2006.