EDITORIAL

The ophthalmologist's guide to evaluating the certainty of evidence using the GRADE approach

Evidence-based medicine demands knowledge of the best available research evidence to guide clinical decisions, which most often comes from rigorous systematic reviews and meta-analyses that identify, summarize, pool, and appraise all research evidence addressing a particular research question. While results from individual studies may appear impressive, our confidence in the body of evidence may be undermined by issues such as limitations in study designs or differences between the questions addressed in studies and the question of interest. To address such challenges, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group developed a system to evaluate the certainty (quality) of evidence—called GRADE. The GRADE approach intends to improve interpretation of evidence for decision-making [1].

The GRADE approach is applied to a body of evidence. A body of evidence refers to all research studies addressing a particular clinical question summarized in a systematic review and meta-analysis. According to the GRADE approach, the certainty of a body of evidence may be rated either as high, moderate, low, or very low (Fig. 1). High certainty evidence suggests that the estimated effect (the results from a rigorous systematic review and meta-analysis) is likely close to the true effect. Conversely, low or very low certainty evidence suggests that the estimated effect may in fact be substantially different from the true effect.

Original GRADE guidance proposed that for questions of causal inference, a body of evidence comprised of randomized trials starts at high certainty and non-randomized studies start at low certainty [2]. This is because, randomized trials, by virtue of randomization, achieve balance (more or less) in prognostic factors such that any observed differences in outcomes between randomized arms can be confidently attributed to the intervention under investigation. In non-randomized studies, however, participants with and without an exposure may differ with regard to prognostic factors such that any differences in outcomes between participants may be an artifact of differences in other prognostic factors. Even if investigators use sophisticated design and analytic methods to adjust for a comprehensive list of prognostic factors, important factors that are unknown or unmeasured may still influence results. This phenomenon is called residual confounding and is the reason why non-randomized studies are initially rated at low certainty.

Newer GRADE guidance now suggests that a body of evidence comprised of non-randomized studies can also start at high certainty and the certainty of evidence may be downgraded by considering limitations of the evidence in comparison to a "target trial"—a hypothetical trial, without any limitations, that may or may not be feasible, addressing the question of interest [3]. Using this approach however, a body of evidence comprised of non-randomized studies will almost always still land at low or very low certainty due to concerns with residual confounding.

The certainty of a body of evidence may be rated down by one or more levels due to concerns related to five factors: risk of bias (i.e., study limitations that may lead to systematic under- or over-estimation) [4], inconsistency (i.e., unexplained heterogeneity in results across studies) [5], indirectness (i.e., differences between the questions addressed in studies and the question of interest) [6], imprecision (i.e., the magnitude of confidence intervals around an estimate in relation to the minimum difference in the outcome that patients find important) [7], and publication bias (i.e., the tendency for studies with statistically significant results or positive results to be published, published faster, or published in journals with higher visibility) [8]. The certainty of evidence may also be rated up in select rare scenarios: when there is a dose-response relationship, a large effect, or when all plausible confounders act in the opposite direction than the observed effect [9].

GRADE is sometimes misinterpreted as evaluating risk of bias or misused as a risk of bias tool. GRADE, however, considers many factors beyond risk of bias and is intended to be applied alongside, rather than instead of, a risk of bias tool [4]. Systematic reviewers are expected to first assess the risk of bias of individual studies using an established risk of bias tool. Then, when applying GRADE, they should consider the risk of bias ratings of individual studies to make judgments about the risk of bias of the entire body of evidence. Of course, to apply GRADE, reviewers will also need to consider factors beyond risk of bias as described above.

Box 1 presents an application of the GRADE approach to a body of evidence addressing a question related to ophthalmology.

Box 1: Application of GRADE to a systematic review in ophthalmology

A systematic review and meta-analysis of randomized trials investigated the effects of face-down positioning following full-thickness macular hole surgery [10]. The systematic review investigated the proportion of patients with macular hole closure following surgery.

The review identified eight trials that reported on macular hole closure. It used the RoB 2.0 tool to assess risk of bias [11]. For this outcome, the majority of evidence came from trials rated at low risk of bias. Figure 2 summarizes the results of the meta-analysis. The meta-analysis yields a relative risk of 1.05 (95% CI: 0.99 to 1.12).

The application of the GRADE approach to binary outcomes requires consideration of absolute effects (e.g., risk differences) [12]. Fortunately, relative effects obtained from meta-analyses like relative risks and odds ratios can be transformed to absolute effects [13]. 86.9% of patients without face-down positioning experience macular hole closure. Using the relative risk obtained from the meta-analysis, this translates to an absolute risk of 912 in 1000 patients with face-down positioning (95% CI: 860 to 973) or a risk difference of 43 more (95% CI: 9 fewer to 104 more).

The systematic review authors rated the certainty of the body of evidence as low. The body of evidence is comprised of randomized trials. The investigators downgraded the certainty of evidence by one level for imprecision because the lower and upper bounds of the confidence intervals suggest opposing clinical actions. The lower bound of the confidence interval (9 fewer cases of macular hole closure in 1000 patients) suggests face-down positioning has little or no effect on macular hole closures. Conversely, the upper bound of the confidence interval (104 more cases of macular hole closure in 1000 patients) suggests that face-down positioning may importantly increase the proportion of patients with macular hole closure.

The investigators also downgraded the certainty of evidence by one level for inconsistency. Whether the effect estimates were sufficiently inconsistent to warrant rating down for inconsistency is subjective. Some may argue that the confidence intervals and point estimates are sufficiently similar to forego rating down for inconsistency. If the investigators were interested in whether the effect estimate is within a narrow range, then rating down for inconsistency would be justifiable. The investigators suggested that rating down for inconsistency was a borderline decision and that they were also concerned about potential indirectness because it was unclear whether the effect of face-down positioning is consistent across other surgical factors, such as peeling of the internal limiting membrane, type of gas, size of the internal limiting membrane peel, and duration of face-down positioning.

The final rating of the certainty of evidence is thus low, suggesting that face-down positioning may increase the risk of macular hole closures. The review also judged other outcomes (visual acuity and complications) as low certainty. Low certainty evidence suggests that the true effect may be different from the estimated effect and that new evidence may importantly change the effect estimate.

UPDATES

The GRADE Working Group has refined and updated its guidance since its inception. Official guidance from the GRADE Working Group is summarized in GRADE Guidance Papers, while GRADE Concept Papers and GRADE Notes discuss concepts relevant to GRADE and case studies [14]. While theoretically anyone can use the GRADE approach to evaluate the certainty of a body of evidence, the confident application of the GRADE approach will require familiarity with the prodigious volume of GRADE guidance papers.

Initially, GRADE guidance focused on assessing the certainty of evidence for causal questions. Since then, GRADE guidance has been extended to apply to questions of prognosis [15] and diagnosis [16]. GRADE now also offers guidance for evaluating the certainty of evidence from network meta-analyses [17, 18]. GRADE has recently clarified judgments related to imprecision [19]. Newer GRADE guidance emphasizes that the certainty of evidence represents the certainty that the true effect lies on one side of a specified threshold or within a chosen range [19].

Most recently, GRADE has been expanded to offer guidance on moving from evidence to recommendations, by way of GRADE Evidence-to-Decision frameworks [20]. These frameworks provide a structured process for moving from evidence to guideline recommendations, systematically considering all factors that may bear on the direction and strength of the recommendations, including the balance between benefits and harms, certainty of evidence, values and preferences, and sometimes, cost-effectiveness, acceptability, feasibility, and equity.

GRADE also offers guidance on summarizing results, developed based on feedback from evidence users and other stakeholders [21]. According to this guidance, high certainty evidence is described with declarative statements, moderate certainty evidence with 'probably', low certainty evidence with 'may' and very low indicated with 'very uncertain'.

Advantages of the GRADE approach

GRADE has now been adopted by over 100 organizations worldwide, including the World Health Organization, the National Institute for Health and Care Excellence, and the Cochrane Collaboration, and across a diverse range of health fields [22].

The GRADE approach offers a systematic and transparent process for assessing the certainty of evidence and moving from evidence to recommendations. Its application ensures that all factors that may bear on the certainty of evidence or on the direction and strength of recommendations are systematically and transparently considered. Nevertheless, its application invariably involves subjective judgments and such judgments may vary between reviewers, guideline developers, and other decision-makers. Disagreements between reasonable individuals are expected. The transparency facilitated by GRADE, however, allows parties to identify specific sources of disagreement.

The application of GRADE also promotes the consideration of patient values and preferences. According to the GRADE approach, patient perspectives inform the selection of outcomes for consideration for guideline recommendations and other decisions, whether the benefits of a particular course of action outweigh harms, and the minimum important difference in outcomes that patients find important [23]. Considering values and preferences respects the rights of citizens to participate in health decision-making, aligns guidelines with the needs and priorities of the communities they are intended to serve, ensures recommendations are logistically feasible and acceptable, and improves support for the recommendations [24, 25].
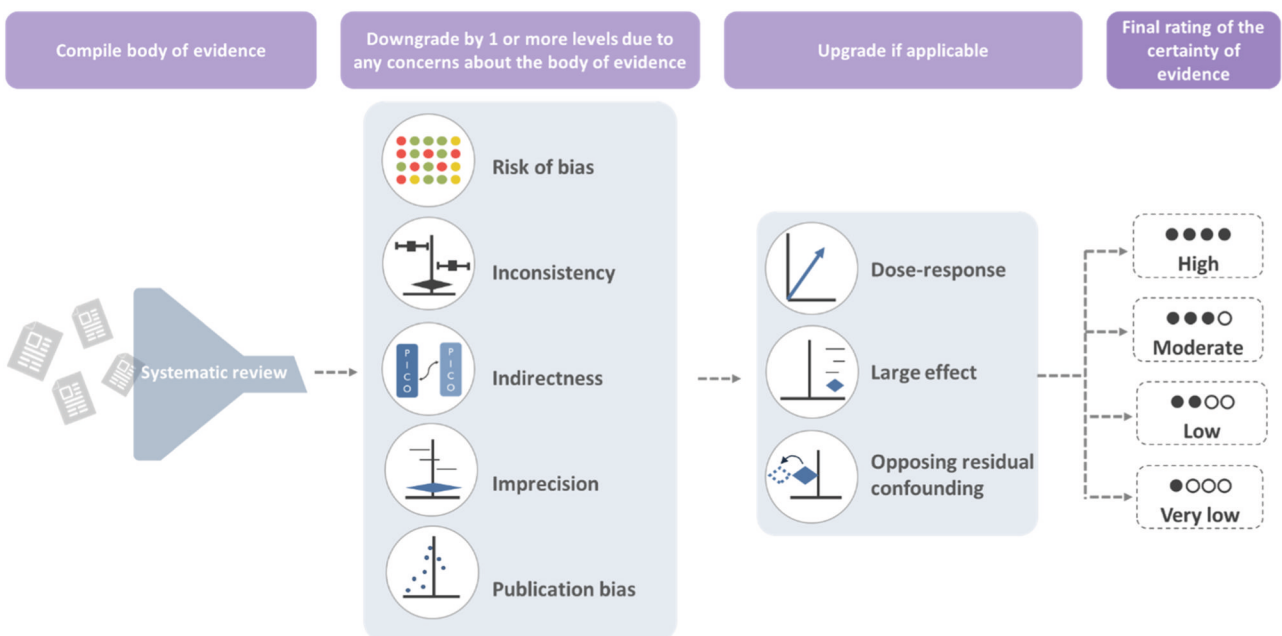

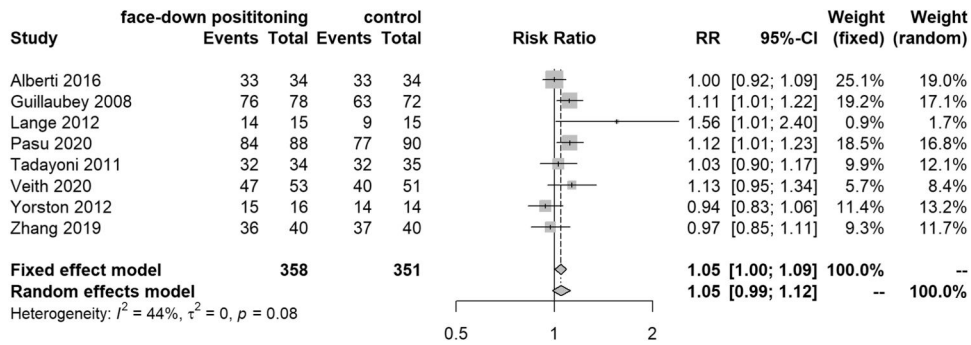**Fig. 1 GRADE Approach.** Overview of the GRADE approach to assessing the certainty of evidence.

Study	face-down positioning		control		Risk Ratio	RR	95%-CI	Weight (fixed)	Weight (random)
	Events	Total	Events	Total					
Alberti 2016	33	34	33	34		1.00	[0.92; 1.09]	25.1%	19.0%
Guillaubey 2008	76	78	63	72		1.11	[1.01; 1.22]	19.2%	17.1%
Lange 2012	14	15	9	15		1.56	[1.01; 2.40]	0.9%	1.7%
Pasu 2020	84	88	77	90		1.12	[1.01; 1.23]	18.5%	16.8%
Tadayoni 2011	32	34	32	35		1.03	[0.90; 1.17]	9.9%	12.1%
Veith 2020	47	53	40	51		1.13	[0.95; 1.34]	5.7%	8.4%
Yorston 2012	15	16	14	14		0.94	[0.83; 1.06]	11.4%	13.2%
Zhang 2019	36	40	37	40		0.97	[0.85; 1.11]	9.3%	11.7%
Fixed effect model		358		351		1.05	[1.00; 1.09]	100.0%	--
Random effects model						1.05	[0.99; 1.12]	--	100.0%

Heterogeneity: $I^2 = 44\%$, $\tau^2 = 0$, $p = 0.08$

0.5 1 2

Fig. 2 Macular Hole Closure Meta-Analysis Results. Meta-analysis comparing macular hole closure between face-down positioning and control.

CONCLUSION

GRADE presents a systematic and transparent approach to evaluating the certainty of evidence and for moving from evidence to recommendations. It has been adopted by multiple authoritative international organizations. Since its inception, GRADE has been refined and updated and will likely continue to evolve based on the experiences of reviewers and guideline developers. Its application offers several strong advantages, including the promotion of consistency in standards to evaluate evidence across health fields and explicit consideration of patient values and preferences.

Dena Zeraatkar[1,2], Tyler Pitre[3], Mark Phillips [ID][2], David H. Steel [ID][4,5], Charles C. Wykoff[6,7], Tien Yin Wong [ID][8,9,10], Mohit Bhandari[2,11], Varun Chaudhary [ID][2,11 ✉] and On behalf of the Retina Evidence and Trials INternational Alliance (R.E.T.I.N.A.) Study Group*

[1]Department of Anesthesia, McMaster University, Hamilton, ON, Canada. [2]Department Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada. [3]Division of Respirology, Department of Medicine, University of Toronto, Toronto, ON, Canada. [4]Sunderland Eye Infirmary, Sunderland, UK. [5]Biosciences Institute, Newcastle University, Newcastle Upon Tyne, UK. [6]Retina Consultants of Texas (Retina Consultants of America), Houston, TX, USA. [7]Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA. [8]Tsinghua Medicine, Tsinghua University, Beijing, China. [9]National Eye Centre, Singapore, Singapore. [10]Singapore Eye Research Institute, Singapore, Singapore. [11]Department of Surgery, McMaster University, Hamilton, ON, Canada. A list of authors and their affiliations appears at the end of the paper. ✉email: vchaudh@mcmaster.ca

REFERENCES

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.
2. Balshem H, Helfand M, Schünemann 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. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105–14.
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;64:407–15.
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, 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.
8. 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.
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. Chaudhary V, Sarohia GS, Phillips MR, Zeraatkar D, Xie JS, Nanji K, et al. Role of positioning after full-thickness macular hole surgery: a systematic review and meta-analysis. Ophthalmol Retin. 2023;7:33–43.
11. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
12. Pitre T, Jassal T, Park L, Coello PA, de Souza R, Zeraatkar D. Reporting and interpretation of effects in non-randomized nutritional and environmental epidemiology: a methods study. Ann Epidemiol. 2023;77:37–43.
13. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol. 2013;66:158–72.
14. McGowan J, Akl EA, Coello PA, Brennan S, Dahm P, Davoli M, et al. Update on the JCE GRADE series and other GRADE article types. J Clin Epidemiol. 2021;140:163–4.
15. Alfonso I, Frederick AS, Maicon F, Carolina A, Eddie L, Bernard B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350:h870.
16. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy. 2009;64:1109–16.
17. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
18. Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol. 2018;93:36–44.
19. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017;87:4–13.
20. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. BMJ. 2016;353:i2016.
21. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2020;119:126–35.
22. Lima JP, Chu X, Guyatt GH, Tangamornsuksan W. Certainty of evidence, why? J Bras Pneumol. 2023;49:e20230167.
23. Zeng L, Helsingen LM, Bretthauer M, Agoritsas T, Vandvik PO, Mustafa RA, et al. A novel framework for incorporating patient values and preferences in making guideline recommendations: guideline panel surveys. J Clin Epidemiol. 2023;161:164–72.
24. Sacristán JA, Aguarón A, Avendaño-Solá C, Garrido P, Carrión J, Gutiérrez A, et al. Patient involvement in clinical research: why, when, and how. Patient Prefer Adherence. 2016;10:631–40.

25. Armstrong MJ, Mullins CD, Gronseth GS, Gagliardi AR. Impact of patient involvement on clinical practice guideline development: a parallel group study. Implement Sci. 2018;13:55.

AUTHOR CONTRIBUTIONS

COMPETING INTERESTS

ON BEHALF OF THE RETINA EVIDENCE AND TRIALS INTERNATIONAL ALLIANCE (R.E.T.I.N.A.) STUDY GROUP

Varun Chaudhary[2,11], Mohit Bhandari[2,11], Charles C. Wykoff[6,7], Sobha Sivaprasad[12], Lehana Thabane[2], Peter Kaiser[13], David Sarraf[14], Sophie J. Bakri[15], Sunir J. Garg[16], Rishi P. Singh[17,18], Frank G. Holz[19], Tien Y. Wong[10,20], Robyn H. Guymer[21,22], David H. Steel [iD][4,5] and SriniVas R. Sadda[23,24]

[12]NIHR Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, UK. [13]Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. [14]Retinal Disorders and Ophthalmic Genetics, Stein Eye Institute, University of California, Los Angeles, CA, USA. [15]Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA. [16]The Retina Service at Wills Eye Hospital, Philadelphia, PA, USA. [17]Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. [18]Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. [19]Department of Ophthalmology, University of Bonn, Bonn, Germany. [20]Singapore National Eye Centre, Duke-NUD Medical School, Singapore, Singapore. [21]Department of Surgery, (Ophthalmology), The University of Melbourne, Melbourne, VIC, Australia. [22]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia. [23]Department of Ophthalmology, David Geffen School of Medicine at University of California, Los Angeles, CA, USA. [24]Doheny Eye Institute, Los Angeles, CA, USA.