

The GRADE approach, part 1: how to assess the certainty of the evidence

Matías Kirmayr^a , Carlos Quilodrán^a , Bárbara Valente^b , Cristóbal Loezar^{a,c} , Luis Garegnani^d , Juan Víctor Ariel Franco^{d,*} 

^a Escuela de Medicina, Universidad de Valparaíso, Valparaíso, Chile

^b Carrera de Medicina, Universidad Nacional de La Matanza, Buenos Aires, Argentina

^c Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

^d Centro Cochrane Asociado, Departamento de Investigación, Instituto Universitario Hospital Italiano, Buenos Aires, Argentina

* Corresponding author jvaf85@gmail.com

Citation Kirmayr M, Quilodrán C, Valente B, Loezar C, Garegnani L, Franco JVA. The GRADE approach, Part 1: how to assess the certainty of the evidence. *Medwave* 2021;21(01):e8109

Doi 10.5867/medwave.2021.01.8109

Submission date 22/9/2020

Acceptance date 28/12/2020

Publication date 09/02/2021

Origin Not commissioned.

Type of review Externally peer-reviewed by three reviewers, double-blind.

Keywords Evidence-Based Medicine, GRADE approach, Systematic reviews as topic

Abstract

The certainty of the evidence for interventions is the certainty or confidence that the true effect is within a particular range or relative to a threshold. In the new pyramid of evidence, systematic reviews represent the magnifying glass through which this certainty is evaluated. The GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach arises in response to the existence of multiple evidence classification systems, and it offers a transparent and structured process to develop and present summaries of evidence considering its certainty and, in a second step, the strength of the recommendations that they inform. The GRADE process begins with an explicit question that includes all important and critical outcomes explicitly. The main domains used to assess the certainty of the evidence are risk of bias, inconsistency, indirectness of evidence, imprecision, and publication bias. The factors that can increase the certainty of the evidence are dose-response gradient, large magnitude of an effect, and effect of plausible residual confounding. Finally, the Summary of Findings tables summarize the process in a simplified way and with controlled language. This narrative review's purpose is to address the GRADE approach's theoretical and practical underlying concepts in a simplified way

and with practical examples.

Main messages

- The GRADE approach establishes unified and transparent criteria to rate the certainty of the evidence and the strength of the recommendations.
- The GRADE methodology's scope includes assessing the certainty of the evidence and the strength of the recommendations on interventions, diagnostic tests, and prognostic models.
- Some of the specificities of GRADE for different types of evidence are in continuous development.
- This work addresses the GRADE methodology's main basic theoretical and practical concepts to assess the certainty of the evidence of interventions in friendly language. It provides examples aimed at the training of undergraduate and graduate students.

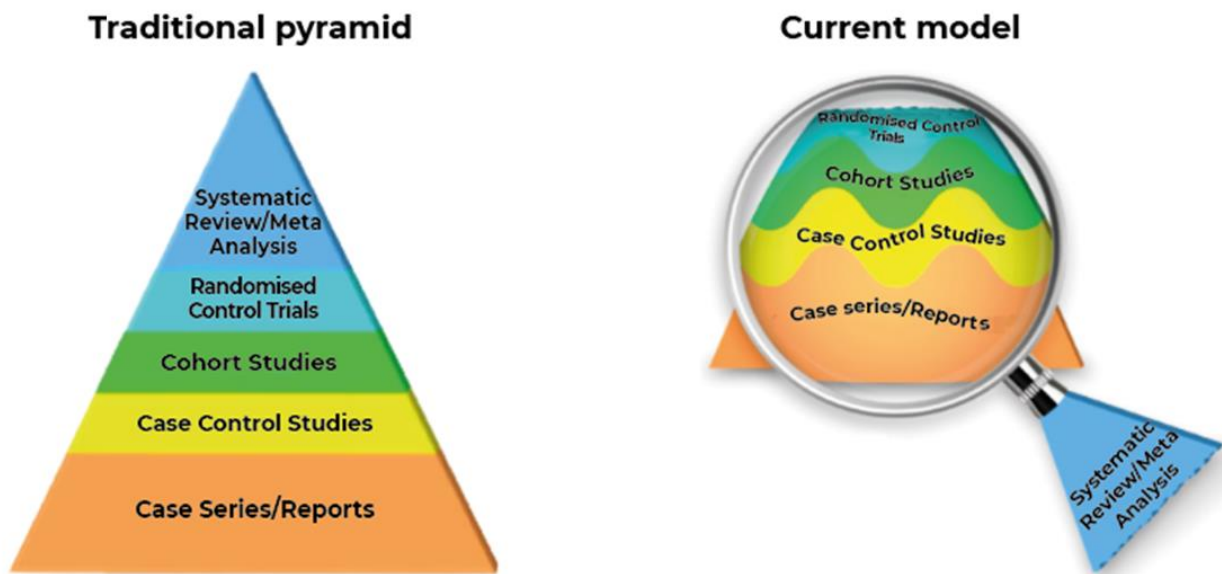
Introduction

Background of The GRADE approach

The concept of incorporating "evidence" into clinical practice was coined in Canada by a group of internists and clinical epidemiologists from McMaster University School of Medicine, who consolidated the concept of evidence-based medicine (EBM) given that the progressive and overwhelming increase in scientific publications was not associated with their certainty. EBM is based on pillars (patient preferences, clinical expertise, and the best available evidence) that contributed to the paradigm shift in medical science study and practice. To make the process of evaluating the best available evidence easier, it was necessary first to objectify its "certainty."

The Canadian Task Force formulated the first ranking of the evidence on the Periodic Health Examination in 1979, and five years later, this ranking was adapted by the U.S. Preventive Services Task Force (USPSTF). Another widely accepted ranking was established in 1995 by epidemiologist David L. Sackett. By this time, there are several coexisting classifications of the quality of the evidence, mainly based on the study design. They usually hierarchize "levels" of evidence (e.g., I, IIa, IIb, III). In turn, there are several classifications of the recommendations according to predefined criteria (e.g., grade A, B, C). The main problem with these systems is the lack of interoperability, so that a "level I" or a "grade A" recommendation does not always mean the same for each system¹. That is why GRADE emerges as a unified definition of the certainty of the evidence that can serve as input for the formulation of recommendations².

Figure 1. New evidence pyramid.



Source: Adapted from the article: Proposed new evidence-based medicine pyramid⁵.

Accordingly, the GRADE approach emerges as a framework to assess the certainty of evidence and to formulate recommendations. Therefore, for a given clinical therapeutic question, an observational study could eventually provide greater certainty of evidence than a randomized clinical trial, or vice versa. It should be considered that observational studies usually fall below clinical trials due to intrinsic

The certainty of the evidence for interventions is the confidence that the true effect is within a particular range or threshold^{3,4}. This definition has two critical contextual implications:

- 1) Systematic reviews should not make recommendations since, in these, the certainty of the evidence reflects the degree of confidence that an effect estimate is correct.
- 2) Panels that develop recommendations for clinical practice guidelines should make judgments about the certainty of the evidence related to the specific context in which the evidence is being used⁵.

GRADE in the new pyramid of evidence

The level of evidence represented by the old or traditional pyramid of evidence provided a static level of evidence for the different designs, positioning systematic reviews and meta-analysis at the highest level of the pyramid. The current model understands systematic reviews as a magnifying glass through which an analysis of the evidence is viewed and applied, the certainty of the evidence being variable and dynamic for different studies with the same design as shown in Figure 1⁵. As an example, to answer a therapeutic question, a clinical trial may not always be the best design: If it incorporates few patients with relevant biases, it may generate less reliable estimates than a cohort study with a large number of patients and a robust analysis of its variables. In this example, the magnifying glass that allows us to differentiate the studies in their internal structure would be the systematic review.

weaknesses in their design (e.g., they do not usually establish strong causality). It is also important to clarify that this approach does not eliminate the inevitable need for judgments with a certain degree of subjectivity but provides a framework for systematization and transparency⁶.

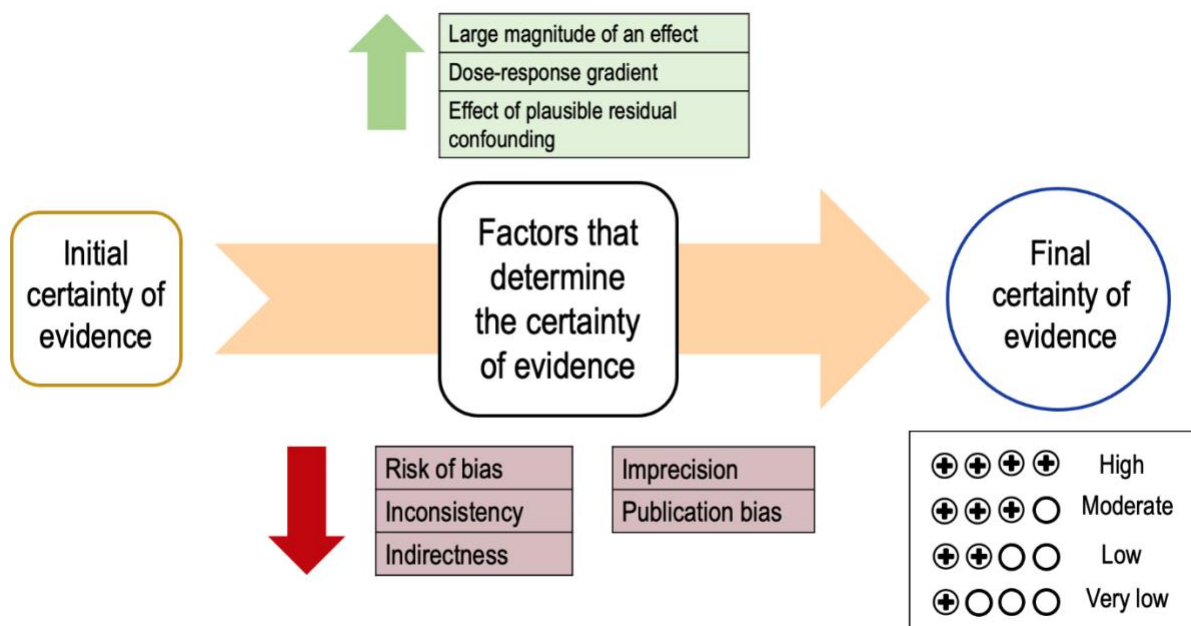
This article is the third of a methodological series of narrative reviews on general topics in biostatistics and clinical epidemiology, exploring and summarizing in a friendly language, published articles contained in the primary databases and specialized bibliography. The series is aimed at the training of undergraduate and graduate students. It is carried out by the Chair of Evidence-Based Medicine of the School of Medicine of the Universidad de Valparaíso, Chile, in collaboration with the Research Department of the Instituto Universitario del Hospital Italiano de Buenos Aires, Argentina, and the UC Evidence Center, at the Universidad Católica, Chile. This article aims to address the main basic theoretical and practical concepts of the GRADE methodology for assessing the certainty of the evidence for interventions.

The GRADE Approach

What is the GRADE methodology, and what is its scope?

This initiative arose as a response to the existence of multiple evidence classification systems (e.g., CTFPHC⁷, USPSTF⁸, CEBM, AHA / ACC⁹, and more). Based on a comprehensive analysis of all current recommendation systems, the GRADE methodology (Grading of Recommendations Assessment, Development and Evaluation) offers a transparent and structured process to develop and present summaries of evidence, including its certainty, for preparation of Summary of Findings tables in systematic reviews and make recommendations in medical care⁶. The scope of the GRADE methodology includes the assessment of the certainty of the evidence and the strength of the recommendations on interventions, diagnostic tests, and prognostic models. Some of the specificities of GRADE for different types of evidence are in continuous development.

Figure 2. Graphic summary of the GRADE approach.



Source: Figure prepared by the authors.

What do the categories awarded by GRADE mean?

Before evaluating the different steps of the process proposed by GRADE, it is convenient to know what each grade of certainty of evidence means¹⁰.

Table 1. Grading of the certainty of the evidence.

Grade	Definition
HIGH	We are very confident that the true effect is close to the estimate.
MODERATE	We have moderate confidence in the effect estimate: The true effect is likely to be close to the effect estimate, but there is a possibility that it is substantially different.
LOW	Our confidence in the estimate of the effect is limited: The true effect may be substantially different from the estimate.
VERY LOW	We have very little confidence in the estimate of the effect: The true effect is likely to be substantially different from the estimate.

Source: Adaptation of the article from the GRADE series: GRADE guidelines: 3. Rating the certainty of evidence¹⁰.

Steps proposed by GRADE to assess the certainty of the evidence

The GRADE process begins with an explicit question, including the specifications of all important and critical outcomes (a topic covered in an upcoming installment in this series). Based on this question and according to the methodological design used, an initial certainty is defined, which is then analyzed according to different domains, modifying the certainty of the evidence to, ultimately, reach a "final certainty." Figure 2 shows a summary of the different steps of the GRADE methodology.

As anticipated, the study design's initial level of evidence varies according to different factors or domains that influence it. These may lower the certainty of the evidence or increase it¹⁰. Generally, results derived from clinical trials start with a high level of initial evidence certainty and those derived from observational studies with a low level of certainty, although exceptionally, this may vary.

Among the domains that lower the certainty, there are:

1. Risk of bias.
2. Inconsistency.
3. Indirectness.
4. Imprecision
5. Publication bias.

These are registered with values (-1) or (-2) depending on whether the result of the domain analysis is considered "Serious" or "Very serious," respectively.

The factors that may increase the certainty of the evidence are:

1. Dose-response gradient.

Examples 1a and 1b. Levels of risk of bias.

1a. Low risk of bias	1b. High risk of bias
<p>A systematic review on the use of graduated compression stockings to prevent deep vein thrombosis (DVT) included 20 randomized controlled trials with 1681 participants. In general, the included studies had a low risk of bias when applying tools that evaluated it. Therefore, the high-certainty evidence was maintained on the claim that graduated compression stockings are effective in reducing the risk of DVT in hospitalized patients undergoing general and orthopedic surgery, with or without other background thromboprophylaxis methods, when clinically appropriate¹³.</p>	<p>A systematic review included three randomized clinical trials to compare surgery versus conservative treatment in lumbar disc herniation management. The benefit of surgery in reducing symptoms after one year or more is unclear, as the trials were considered to have inadequate allocation concealment, and outcomes were assessed by surgeons using a non-validated scoring tool without blinding. These characteristics constitute a high risk of bias that justifies a reduction by two levels of the certainty of the evidence: From high to low certainty of the evidence^{14,15}.</p>

2) Inconsistency (Heterogeneity): Inconsistency refers to the variation of the effect estimates (including the confidence interval), having excluded the leading causes of this variation¹⁶. For example, the effect of a pharmaceutical intervention may vary from one study to another due to the variability of the PICO question elements (patients, intervention, comparison, or outcome) in each of the studies. However, if this variability does not justify the differences in the estimate of the effect, the certainty of the evidence decreases.

Any inconsistency in the effect estimates can be assessed by the degree of overlap between study results should initially be assessed. If there is an overlap between the confidence intervals of each primary study's estimates, the effects may be deemed consistent, and conversely, if there is little overlap, these findings are inconsistent.

Example 2. Assessment of inconsistency.

<p>A clinical practice guideline for treating patients with chronic obstructive pulmonary disease assessed the efficacy of respiratory rehabilitation in the context of an exacerbation in terms of exercise tolerance with the 6-minute walk test. A systematic review that analyzed six studies for this test showed highly variable results across studies, with confidence intervals that did not overlap, represented with statistical heterogeneity ($I^2 = 93\%$), as shown in Figure 3. In these circumstances, confidence in the results decreases by at least one level. Therefore, the certainty of the evidence initially rated as high for a systematic review of randomized clinical trials, would become moderate.</p>
--

2. Large effect size.
3. Effect of plausible residual confounding.

These are also recorded as numbers (+1) or (+2) depending on their impact on the overall certainty.

A. Domains that lower the level of evidence

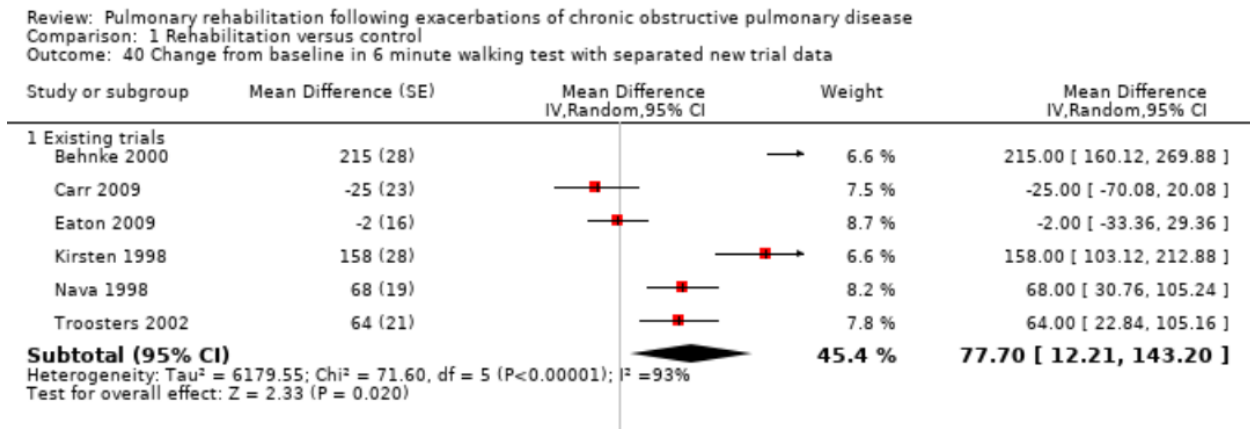
1) Risk of bias: Both randomized controlled trials (RCTs) and observational studies may incur an additional risk of misleading results if they have flaws in their design or conduct, which other publications call "validity" or "internal validity problems," and we call "study limitations" or "risk of bias"¹¹. The risk of bias can indicate a systematic deviation from the truth, reducing the certainty of the results, and, in a meta-analysis, it is more prudent to trust the one with the lowest risk of bias. It should be noted that risk of bias is not the same as bias, the latter being the systematic error, and the risk of bias the subjective assessment of the probability of this error. There are numerous tools to assess the risk of bias in randomized trials and observational studies (for example, the tools developed by Cochrane: RoB 2 and ROBINS-I respectively)^{11,12}.

Inconsistency is easier to understand through the visual inspection of a Forest plot (Example 2).

Furthermore, it is possible to resort to statistical tests, such as the heterogeneity test, in which a low p-value indicates inconsistency, and the I^2 statistic, which quantifies the proportion of variation in the point estimates due to differences between studies: the higher I^2 the greater the inconsistency (< 40% is considered low, 30 to 60% moderate, 50 to 90% substantial, and 75 to 100% significant)¹⁶.

It is important to clarify that this domain is not related to the distribution of effects regarding the line of no effect.

Figure 3. Example of Inconsistency - Meta-analysis of rehabilitation versus usual care in COPD exacerbations. Exercise tolerance test at 6 min.



Heterogeneity:
 TAU²: 6179.55
 Chi-square: 71.60
 df: 5 (p < 0,00001)
 I²: 93%
 Test for overall effect:
 Z: 2.33 (p = 0.02)

Source: Figure extracted from the article "Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease". Cochrane Database Syst Rev. 2011¹⁷.

3) Indirectness: Indirectness refers to the emerging problems in the concordance between the elements of the PICO question of the evidence and the PICO question of the systematic review or guideline that uses such evidence. The evidence may be indirect for the study question when there are differences in patients,

interventions, comparisons, or outcomes¹⁸. This means that the study does not directly answer the result raised in the clinical question, but instead, the study uses a different result (for instance, a surrogate outcome), but that could indirectly provide information on the outcome of interest (See Example 3).

Example 3. Evaluation of indirectness - Effect of alendronate on pathological fractures.

Clinical question

A systematic review evaluates the effect of alendronate in preventing pathological fractures at 10 years in women with osteoporosis. Suppose the primary outcome of an included study is an improvement in bone densitometry rather than radiographic alterations characteristic of fractures. In that case, it does not directly measure the effect on the presence or absence of fractures but an intermediate (surrogate) result that predicts the incidence of fractures (indirect outcome). Conversely, if a healthcare professional discusses the treatment options for osteoporosis with an 88-year-old woman considering the evidence from clinical trials in which the mean age was 61, he may consider that he is extrapolating the results (indirect population). In both cases, assuming they are randomized clinical trials, they would go from a high level of certainty to, possibly, a moderate level of certainty.

4) Imprecision: Results are imprecise when studies include relatively few patients and few events; consequently, they have wide confidence intervals around the effect, which does not allow a clear estimate of its magnitude¹⁹.

Precision can be evaluated in two ways:

A) examining the extremes of the confidence interval of the effect estimate in relation to a threshold.

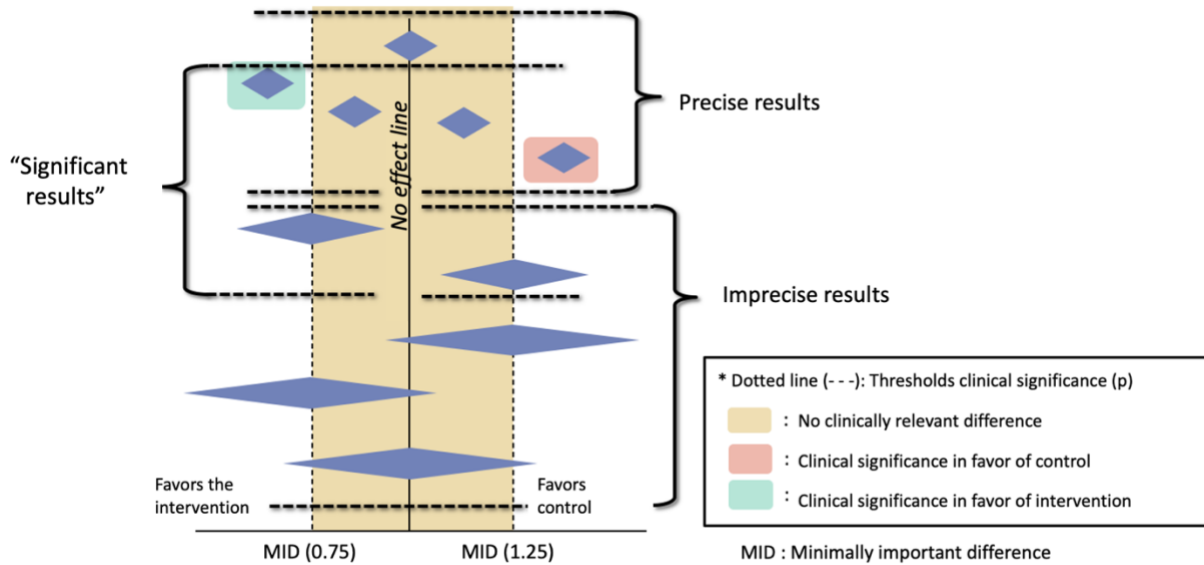
b) evaluating the amount of available data in relation to an optimal information size.

When using the first approach, it is necessary to define what threshold is used to estimate that an effect is clinically relevant and its magnitude (small or substantial). Usually, the minimally important

difference can be taken as a reference, for example, for continuous outcomes (Figure 4). For the second approach, it is necessary to calculate a clinical trial's sample size with adequate power for the outcome being evaluated. If the confidence interval crosses the established threshold or the amount of data does not reach the optimal information size, the certainty of the evidence is downgraded.

Graphically, what should be analyzed is the confidence interval of the meta-analysis (not of each individual study) represented by the diamond figure in the Forest plot shown as an example in Figure 4.

Figure 4. An explanatory figure of the concept of imprecision in relation to clinical significance.



Source: Figure made by the authors.

The minimally important difference (MID) corresponds to the smallest difference in score in the outcome of interest that patients consider important, indicating a change in treatment¹¹. This concept, which will be explored in a future article in the series, relates the magnitude of the change to treatment decisions in clinical practice, emphasizing that the patient's perception prevails over statistical significance. This concept applies to outcomes with continuous values such as those that use scales (for example, the visual analog scale to determine pain intensity), while in a dichotomous outcome (such as mortality), the effect estimate is through relative risks, odds ratio, or hazard ratio. There are no pre-established thresholds for

dichotomous outcomes as the minimally important difference; however, the relative risks of 0.75 and 1.25 are taken as a rule of thumb to assess the precision of the results. These relative threshold values translate into absolute risk differences, which are the most important for decision-making. For example, a relative risk of 0.75 could imply a reduction in the absolute risk of mortality from 40% to 30% (10% difference in mortality). If the event is rarer, a relative risk of 0.75 may be less relevant. For example, it could imply a reduction in cancer-specific mortality from 4 per 1000 to 3 per 1000 (an absolute difference of 1 per 1000).

Examples 4a and 4b. Continuous value precision and dichotomous outcome.

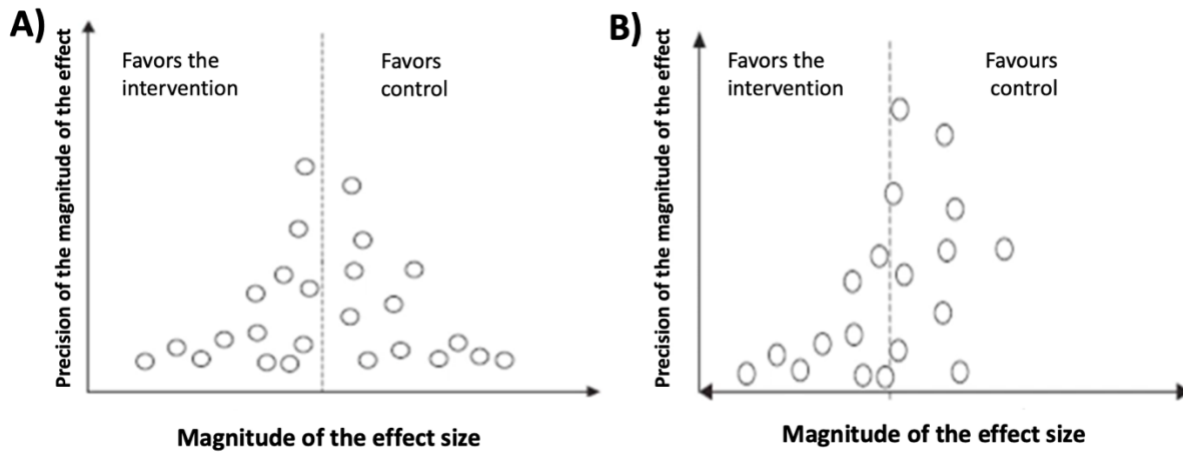
<p style="text-align: center;">4a. Imprecise continuous value</p> <p>In shoulder pain, the most widely used scale is the Constant Murley Functional Assessment graduated from 0 to 100. On this scale, a minimally important difference corresponds to a variation of 10.4 points²⁰. If a study determines that after a given surgery, patients' pain decreases, for instance, by 11 points (95% CI -9 to -12) on this scale, it could be considered a clinically significant intervention. However, when examining both extremes of the confidence interval, imprecision is observed when a scenario in which this difference may be clinically insignificant is included (percentage of patients who would decrease their pain between 9 to 10.4).</p>	<p style="text-align: center;">4b. Precise dichotomous outcome</p> <p>In preterm birth syndrome, the absolute risk of in-hospital mortality in newborn patients between 24 and 31 weeks is around 20%. The administration of antenatal corticosteroids demonstrated a relative risk of 0.5 (95% CI 0.4-0.6)²¹ in in-hospital mortality compared to no administration. When examining the confidence interval, both extremes lay below the threshold (rule of thumb of 0.75). Therefore, it is an intervention that shows a substantial effect over the entire range of possible effects. This also translates into a reduction in the absolute risk of neonatal mortality between 8% and 12%.</p>
--	---

5) Publication bias: It is possible that there are studies whose results contradict the results of the meta-analysis and that these have not been included, either because they were not published or because they are in non-indexed journals²². These studies typically have "negative" results (i.e., they do not show a significant effect in favor of the intervention). This phenomenon is denominated publication bias. However, this form of bias is hard to detect; therefore, it is difficult to lower the certainty of the evidence for this

cause confidently and, if there is sufficient evidence of this bias, the evidence is usually downgraded one level (-1).

Publication bias is usually assessed using a graph called Funnel Plot [Figure 5], representing how the studies are distributed around the study's real effect; when it is funnel-shaped, there would be no bias. If this bias exists, we will see a space between the studies giving this image of a truncated pyramid, as in Figure 5, with missing studies that contradict the intervention's central estimate.

Figure 5. Example of a funnel plot.



A) Uniform pyramidal distribution, without publication bias.

B) Truncated pyramid distribution, with publication bias.

Source: Figure extracted from GRADE guidelines: 5. Rating the quality of evidence publication bias²².

B. Factors that can increase the certainty of the evidence

Although the main role is taken by the factors that lower the level of certainty, it is important to know that factors can increase it. However, two conditions are required: that the study is observational and does not present any of the limitations (biases) previously mentioned. There are three reasons to increase the level of certainty by one point:

1) Large magnitude of effect: If the body of evidence gives large or very large estimates of the magnitude of an intervention's effect, it is possible to be more confident about the results. On such occasions, although observational studies are more likely to overestimate the true effect, the study design that is more prone to bias is unlikely to account for the full apparent benefit or risk. A meta-analysis of observational studies showed that bicycle helmets reduce head injury risk in cyclists by a wide margin (odds ratio 0.31, 95% CI 0.26 to 0, 37). In the absence of obvious bias that the association might create, this large effect suggests a moderate rather than low certainty evidence rating²³.

2) Dose-response gradient: The existence of a dose-response effect also increases the level of certainty. For example, there is a dose-response gradient between higher INR levels (an indicator of the degree of anticoagulation) and the increased risk of bleeding in

anticoagulant users. This finding would increase our certainty as to which supra-therapeutic anticoagulation levels might increase the risk of bleeding¹⁵.

3) Effect of potential residual confounding factors: When assessing the evidence, the possibility that there are unmeasured confounding factors that reduce or increase a given effect may be considered. When the result obtained is not affected by these factors, it is possible to increase the certainty by one level. For example, if an analysis considers that people who use condoms are likely to have more sexual partners, it would be expected that the risk of HIV infection is higher in this population. However, this factor did not reduce the demonstrated effect of the condom, making it such an effective measure for the prevention of HIV transmission that it outperforms this unmeasured confounding factor²⁴.

Summary of Findings Tables (SoF)

All this process proposed by GRADE is summarized quite simply in the Summary of Findings tables or SoF tables for their acronym.

These tables show the entire process for rating the certainty of the evidence for each outcome, with their relative and absolute effect, the certainty of evidence (with its justification), and the key message in a controlled language.

Figure 6. Summary of Findings table about Treatment with Sialogogues and non-pharmacological treatment for Xerostomia.

PHARMACOLOGICAL TREATMENT WITH SIALOGOGUE PLUS NON-PHARMACOLOGICAL TREATMENT COMPARED WITH ONLY NON-PHARMACOLOGICAL TREATMENT FOR XEROSTOMIA							
Components of the PICO question		Patients	Individuals with xerostomia				
		Intervention	Pharmacological treatment with sialogogue plus non-pharmacological treatment				
		Comparison	Only non-pharmacological treatment				
Outcomes	Outcomes	Relative effect (95% CI) -- Studies/ Patients	Estimated absolute effect*			Certainty of evidence (GRADE)	Key messages in simple terms
			NO sialogogue	WITH sialogogue	Difference (95% CI)		
	Reduction of dry mouth sensation (greater or equal to 25 mm on the VAS scale)**	RR 1,95 (1,62 to 2,36) -- 4 trials/ 616 patients [4, 9, 10, 12]	315 per 1000	614 per 1000	Difference: 299 more patients (199 to 460 more)	⊕⊕⊕⊕ High	The use of sialogogues in patients with xerostomia reduces the sensation of dry mouth.
	Unstimulated salivary flow (ml/min)	-- 2 trials/ 298 patients [9, 12]	Increase of 0,06 ml/min	Increase of 0,2 ml/min	MD: 0,14 ml/min more (0,01 to 0,27 more)	⊕⊕⊕⊕ ¹ Moderate	The use of sialogogues in patients with xerostomia probably increases salivary flow.
	Adverse effects	The dropout rate due to adverse effects was 30% compared to 16% in the placebo group [2]. The most frequent adverse effects were sweating (60% in higher doses and 10% in lower doses, versus 7% with placebo) and alteration of voiding frequency (10-15% vs 1-6%) [9, 12]***				⊕⊕⊕⊕ ¹ Moderate	The use of sialogogues in patients with xerostomia probably increases adverse effects.

95% CI: 95% confidence interval // RR: Relative risk // MD: Mean difference.
GRADE: Evidence grades, Grading of Recommendations Assessment, Development and Evaluation.
* The risk WITHOUT pharmacological treatment is based on the risk of the control group in the studies. The risk WITH pharmacological treatment (and its confidence interval) is calculated from the relative effect (and its confidence interval).
** The mean overall change in subjective dry mouth symptoms (xerostomia) was assessed by a change on a visual analog scale (VAS) of 0-100 mm or it was subjectively assessed using a categorical result of improvement compared to baseline.
*** Side effects seen when using Pilocarpine 5 mg 4 times a day for 12 weeks.
¹The quality of evidence was rated down by one level due to risk bias, as most studies are at high risk of selection bias.
Elaboration date: November, 2018.

Source: Schematic adaptation, extracted and translated from the Chilean Ministry of Health.

Example 5. Explanation of SoF Table - Treatment with sialogogues plus non-pharmacological treatment compared with only non-pharmacological treatment for Xerostomia.

In this example, we talk about non-pharmacological treatment associated or not with sialogogues. At the top of the table, there are the first three components of the PICO question, and the outcomes are developed in the table below. The first of these is:

Reduction in the sensation of dry mouth: In this case, the relative risk is 1.95 (1.62 to 2.36). This means that there is a "higher risk" that the dry mouth sensation will be reduced (which is good). The number of clinical trials and the number of patients is also shown; and separately the control and intervention group and the difference between them with their respective 95% CI (in this case 299 more patients per 1000 when managed with sialogogues, with dispersion from 199 to 460 patients per 1000).

Clarifications:

Regarding the certainty of the evidence, this is high, as expressed in the table by means of "+" figures (+ very low; ++ low; +++ moderate; and ++++ high).

In the second and third outcomes (unstimulated salivary flow and adverse effects), the certainty is moderate. A superscript (1) explains the reasons for this certainty referenced in the footnotes of the table. In this case, the risk of bias was argued, given that the majority of the studies presented selection bias, decreasing the certainty by one level.

Controlled language

Once the certainty of evidence has been obtained, it must be communicated through a controlled language, transmitting in simple words the entire process previously carried out. In the SoF table, it is specified in the box on the right titled "Key message in simple terms," where phrases such as "probably increases salivary flow/adverse effects" reflect moderate certainty; and statements such as "reduces the feeling of dry mouth" reflect high certainty.

Conclusions

Currently, the GRADE approach is one of the main methods used internationally in developing clinical guidelines and

recommendations since it provides a clear and transparent framework for assessing and presenting the certainty of the evidence. Although the GRADE working group addresses each of the concepts thoroughly in its extensive series on this methodology, at first, its understanding and application can be overwhelming for beginners. Through simple descriptions and examples of the GRADE methodology's fundamental elements, this article seeks to provide a simplified and summarized approach for its understanding and application by the general public, students, and new methodologists.

Notes

Authorship contributions

MK: Conceptualization, Methodology, Research, Original draft writing, Review and edition, Visualization, Project management. CQ - BV: Conceptualization, Methodology, Research, Original draft writing, Review and editing, Visualization. CL - LG: Conceptualization, Methodology, Review and edition, Visualization. JVAF: Conceptualization, Methodology, Original draft writing, Review and edition, Visualization.

Competing interest

All authors declare that they have no conflicts of interest involved in the preparation of this article.

Funding

We declare that we have no external sources of financing associated with the elaboration of this article.

Acknowledgments

We thank the professorship of Evidence-Based Medicine of the Medical School, Universidad de Valparaíso, Chile, for promoting this series and the collaboration of the Research Department of the Instituto Universitario Hospital Italiano, Buenos Aires, Argentina.

References

1. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* 2011 Jul;128(1):305-310. | CrossRef | PubMed |
2. Franco JVA, Arancibia M, Meza N, Madrid E, Kopitowski K. Clinical practice guidelines: Concepts, limitations and challenges. *Medwave.* 2020 Apr 30;20(3):e7887. | CrossRef | PubMed |
3. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ.* 2008 May 3;336(7651):995-8. | CrossRef | PubMed |
4. 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 Jul;87:4-13. | CrossRef | PubMed |
5. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med.* 2016 Aug;21(4):125-7. | CrossRef | PubMed |
6. 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 Apr;64(4):383-94. | CrossRef | PubMed |
7. Birtwhistle R, Pottie K, Shaw E, Dickinson JA, Brauer P, Fortin M, et al. Canadian Task Force on Preventive Health Care: we're back! *Can Fam Physician.* 2012 Jan;58(1):13-5. | PubMed |
8. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001 Apr;20(3 Suppl):21-35. | CrossRef | PubMed |
9. Correction to: 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation.* 2019 Sep 10;140(11):e649-e650. | CrossRef | PubMed |
10. 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 Apr;64(4):401-6. | CrossRef | PubMed |
11. 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 Apr;64(4):407-15. | CrossRef | PubMed |
12. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Assessing risk of bias in a randomized trial. *Cochrane Handbook for Systematic Reviews of Interventions.* 2019. 205–28. [Internet] | CrossRef |
13. Sachdeva A, Dalton M, Lees T. Graduated compression stockings for prevention of deep vein thrombosis. *Cochrane Database Syst Rev.* 2018 Nov 3;11(11):CD001484. | CrossRef | PubMed |
14. Gibson JN, Waddell G. Surgical interventions for lumbar disc prolapse. *Cochrane Database Syst Rev.* 2007 Jan 24;(1):CD001350. | CrossRef | PubMed |
15. Higgins JPT. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1.* The Cochrane Collaboration. 2008 [cited 2020 Sep 11]. [Internet]. | Link |
16. 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 Dec;64(12):1294-302. | CrossRef | PubMed |
17. Puhan MA, Gimeno-Santos E, Scharplatz M, Troosters T, Walters EH, Steurer J. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev.* 2011 Oct 5;(10):CD005305. | CrossRef | PubMed |
18. 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 Dec;64(12):1303-10. | CrossRef | PubMed |
19. 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 Dec;64(12):1283-93. | CrossRef | PubMed |
20. Donado-Gómez JH, Jiménez-Cotes EA, Cardona LM. Diferencia mínima clínicamente importante en algunas escalas de medición utilizadas en enfermedades musculoesqueléticas. *Revista Colombiana de Reumatología.* 2014 Sep 1;21(3):125–32.
21. Norman M, Piedvache A, Borch K, Huusom LD, Bonamy AE, Howell EA, et al. Association of Short Antenatal Corticosteroid Administration-to-Birth Intervals With Survival and Morbidity Among Very Preterm Infants: Results From the EPICE Cohort. *JAMA Pediatr.* 2017 Jul 1;171(7):678-686. | CrossRef | PubMed |
22. 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 Dec;64(12):1277-82. | CrossRef | PubMed |
23. Schünemann H. *The GRADE handbook.* Cochrane Collaboration. 2013.
24. Cáceres CF. Intervenciones para la prevención del VIH e ITS en América Latina y Caribe: una revisión de la experiencia regional. *Cadernos de Saúde Pública.* 2004;20(6):1468–85.
25. Ministerio de Salud de Chile. Informe de búsqueda y síntesis de evidencia de efectos deseables e indeseables. *Guía de Práctica Clínica Salud Oral Integral: Tratamiento y Rehabilitación Oral en Personas Adultas y Personas Mayores con Edentulismo Parcial o Total.* 2019. [Internet]. | Link |

Postal address
Potosí 4265, Buenos Aires,
Argentina



Esta obra de *Medwave* está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, *Medwave*.