



CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia


Special article

GRADE System: Classification of Quality of Evidence and Strength of Recommendation[☆]



José Luis Aguayo-Albasini, Benito Flores-Pastor,* Víctor Soria-Aledo

Servicio de Cirugía General, Hospital General Universitario JM Morales Meseguer, Campus de Excelencia Internacional Mare Nostrum, Universidad de Murcia, Murcia, Spain

ARTICLE INFO

Article history:

Received 16 April 2013

Accepted 19 August 2013

Available online 12 March 2014

Keywords:

Evidence quality

Strength of recommendation

GRADE system

Clinical practice guidelines

A B S T R A C T

The acquisition and classification of scientific evidence and subsequent formulation of recommendations constitute the basis for the development of clinical practice guidelines. There are several systems for the classification of evidence and strength of recommendations; the most commonly used nowadays is the Grading of Recommendations, Assessment, Development and Evaluation system (GRADE). The GRADE system initially classifies the evidence into high or low, coming from experimental or observational studies; subsequently and following a series of considerations, the evidence is classified into high, moderate, low or very low. The strength of recommendations is based not only on the quality of the evidence, but also on a series of factors such as the risk/benefit balance, values and preferences of the patients and professionals, and the use of resources or costs.

© 2013 AEC. Published by Elsevier España, S.L. All rights reserved.

Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación

R E S U M E N

La adquisición y jerarquización de la evidencia, así como la posterior formulación de recomendaciones, constituyen la base del desarrollo de las guías de práctica clínica. Sistemas de graduación de la calidad de la evidencia y de la fuerza de las recomendaciones han existido muchos y actualmente se va imponiendo el modelo *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE). En el sistema GRADE la calidad de la evidencia se clasifica, inicialmente, en alta o baja, según provenga de estudios experimentales u observacionales; posteriormente, según una serie de consideraciones, la evidencia queda en alta, moderada, baja y muy baja. La fuerza de las recomendaciones se apoya no solo en la calidad de la evidencia, sino en una serie de factores como son el balance entre riesgos y beneficios, los valores y preferencias de pacientes y profesionales, y el consumo de recursos o costes.

© 2013 AEC. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Palabras clave:

Calidad de la evidencia

Fuerza de la recomendación

Sistema GRADE

Guías de práctica clínica

[☆] Please cite this article as: Aguayo-Albasini JL, Flores-Pastor B, Soria-Aledo V. Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación. Cir Esp. 2014;92:82-88.

* Corresponding author.

E-mail addresses: florespastorbenito2@gmail.com, benitom.flores@carm.es (B. Flores-Pastor).
2173-5077/\$ – see front matter © 2013 AEC. Published by Elsevier España, S.L. All rights reserved.

Introduction

Evidence-based medicine (EBM) requires medical practitioners to combine their medical knowledge and judgement with the best existing scientific knowledge. Determining the best evidence requires skills of identification, critical analysis and prioritising published evidence. The former stage is essential, as any recommendation or grade of recommendation proposed in terms of preventive or therapeutic surgery or concerning a diagnostic procedure must be directly related to the quality (and other factors) of the existing evidence.

EBM is chiefly of interest to groups of experts who develop clinical practice guidelines (CPG) for research on a disease or health problem and for diagnosis, treatment and prevention. Up to 8 stages are described in the development of a guideline (Table 1), but only stages 3–8 concern us in this article (formulating questions, acquiring evidence, assigning quality and drawing up recommendations). Obtaining useful CPG is not an easy task due to the varied nature of the individuals making up the groups or committees of experts who create these guidelines, their different points of view and methods, and the similar variability of scientific information available on a particular topic.^{1–3} Until a few years ago these groups of experts used an informal methodology to reach a consensus, but recently procedures for prioritising evidence and establishing appropriate recommendations have improved. Here the system for the *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE) comes into play.

There are a great many sophisticated systems for categorising scientific evidence, including the English model, the Oxford Centre for Evidence-Based Medicine (OCEBM), the Scottish Intercollegiate Guidelines Network (SIGN) or the American College of Chest Physicians (ACCP) used by the ACCP itself in their guidelines on venous thrombosis up until their seventh revision.^{4–6} All of them attributed different quality levels to studies on a particular problem, which then enabled different degrees of recommendation to be made. However, some disadvantages soon emerged, such as the fact that these systems were developed principally as a result of a consensus of expert opinion and were not validated.⁷ Therefore, occasionally, different systems were not categorising the same studies in terms of similar levels of evidence. Indeed, sometimes no agreement was reached on the same model. Moreover, some systems were better at estimating the quality of evidence than establishing the grade of recommendation,

and vice versa. All the above-mentioned meant that occasionally the CPG were not completely reliable.

The GRADE working group's proposal was communicated in 2004. It was created by an international and multidisciplinary group of methodologists, experts in CPG and clinical doctors, in an attempt to deal with the problems mentioned above.^{8,9} The advantage of the system is that it is a thorough and transparent method for classifying quality of evidence and for allocating a grade or strength of recommendation. We shall develop these points as the GRADE system does, but first we shall outline the steps to be followed in the formulation of clinical questions.

Formulation of Clinical Questions in PICO Format and Search for Answers

Once the scope of a CPG has been established, a series of clinical questions need to be defined which are grouped into sections of organisation, prevention, diagnosis, treatment, prognosis, etc. PICO (acronym for Patient–Intervention–Comparison–Outcome) is the preferred method used to move from a generic clinical question to a specifically formulated one to facilitate a bibliographic search and preparation of recommendations for each question. Thus:

- a. Patient: or population, disease statuses, age groups, comorbidities, etc.
- b. Intervention: treatment, diagnostic test, aetiological agent, etc.
- c. Comparison: possible alternative to intervention under research as a regular treatment or placebo, gold reference standard of a diagnostic test, lack of aetiological agent, etc.
- d. Outcomes: relevant outcome variables in the case of studies on efficacy, prognosis or aetiology, and validity estimators in the case of diagnostic tests (sensitivity, specificity, probability coefficients, etc.).

When clinical questions are formulated in PICO format they are defined in a specific manner and there is no ambiguity as to what is being probed and moreover, as each type of question corresponds to a type of study with the appropriate design for its answer, the format helps towards conducting a literature search. During the formulation of clinical questions all the possible outcome variables must be defined. This is an even more relevant issue when used in preparing the GRADE system recommendations, where the variables are qualified as to their importance for clinicians and patients and are weighted on a scale from 1 to 9. Only variables with a score from 7 to 9 are considered key in affecting a GRADE system decision and the clinical questions need to be specified to these key variables. The answers to these questions on key outcomes shall be those which are used to grade the recommendations. Variables with a score of 4–6 are classified as important but not crucial for decision making. Those given a score from 1 to 3 shall be considered unimportant and will not be included in the evaluation or influence the recommendations. The strict and accurate selection of key outcome variables means that the studies are selected equally and thus it is possible that the

Table 1 – Stages in the Preparation of a Clinical Practice Guideline.

1. Definition of scope and objectives
2. Creation of CPG preparation group
3. Formulation of the clinical questions (PICO)
4. Search for evidence
5. Assessment and synthesis of literature
6. Formulation of recommendations
7. External review
8. Edition

CPG: clinical practice guideline; PICO.

findings that are going to be used to infer recommendations and therefore their strength, can vary from one CPG to another.¹⁰

A documentalist is sometimes needed to collaborate in finding the answers that we are looking for to PICO questions. We might need to consult previous GPC, updated systematic revisions, or original studies. When the scientific evidence has been found, it has to be categorised according to its methodological quality (internal validity), the importance of its outcomes and their applicability.

Finally, the strength of recommendations is graded according to a set system. Until now the Scottish SIGN system has been one of the systems used for questions in relation to treatment or prognosis, and the English Oxford CMBE system for questions on diagnosis. At present the GRADE system is starting to be used which we shall mention later. The GRADE Working Group proposed a different approach based on previous systems, which boasts a better structure and greater transparency and information.⁸⁻¹⁴ The advantages of this approach are (a) it weighs up the relative importance of the outcome variables and chooses the ones which are key; (b) it offers detailed descriptions of the evidence quality criteria with respect to specific outcomes and uses explicit definitions and sequential judgements during the categorisation process; (c) it separates the quality of the evidence and the strength of recommendations; (d) and it also considers the balance between benefits and risks, the patient's values and the consumption of resources or costs. It also provides the so-called evidence profile tables and summary of findings; these are unique and essential tables which we shall discuss later.

Levels of Evidence

GRADE defines the quality of evidence as the extent to which one can be confident that an estimate of effect is correct in order for a recommendation to be made. An assessment is made of each key outcome; therefore the same comparison of a therapeutic or preventive intervention can receive different allocations of quality of evidence. The GRADE system sets four categories for rating quality of evidence: high, moderate, low and very low. Table 2 shows what each of the 4 categories represents in terms of their initial and current conception.

The first stage of the GRADE system considers experimental studies as high quality (randomised clinical trials) and

observational studies as low quality (case-control, cohorts). In the second stage, for refining the level of quality, the system sets a series of items to be considered and which can either lower or raise the initially allocated level of quality.

a. Items which lower quality:

- 1) *Limitations with the design and execution of the study (risk of bias)*: Insufficient or incorrect randomisation, lack of blinding, major losses to follow-up, analysis without intention to treat and trials ending prematurely.
- 2) *Inconsistency of outcomes*: when outcomes display a great deal of unexplained variability or heterogeneity. Particularly if some studies show substantial benefits and others no effect or even harm.
- 3) *Uncertainty as to whether the evidence is direct (indirectness)*: Following the PICO method, this can occur with the patients studied (age, gender or clinical status differences), with the intervention, if it is similar but not identical; with the comparison made; or with the outcomes, if some are compared short term and others long term, etc.
- 4) *Imprecision*: this occurs if the confidence intervals (CI) are broad, the samples are small or there are few events.
- 5) *Publication bias*: when there is a high probability of unreported studies, mainly due to a lack of impact, or when all the relevant outcome variables have not been included.

b. Items which raise quality:

- 1) Strong association: findings of relative effects $RR > 2$, or < 0.5 in observational studies with no confounding factors.
- 2) Very strong association: findings of relative effects $RR > 5$, or < 0.2 based on studies where there are no problems with bias or precision.
- 3) Where there is a dose-response gradient.
- 4) Evidence that all possible confusion or bias factors might have reduced the effect observed.

Situations which can determine increased confidence in the results of observational studies are uncommon. In such cases this increase should only be considered if there are no design or execution limitations (which could diminish quality) and there is also a very major and immediate effect or radical change in the prognosis after a particular intervention.

All these items determine, according to the scores shown in Table 3, whether the level of quality of the evidence is lowered

Table 2 – GRADE System: Meaning of the 4 Levels of Evidence.

Quality levels	Current definition	Previous concept
High	High confidence in the correlation between true and estimated effect	Confidence that the estimation of effect will not vary in subsequent studies
Moderate	Moderate confidence in the estimated effect. It is possible that the true effect is very different from the estimated effect	Subsequent studies may have a significant impact on our confidence in the estimate of effect
Low	Limited confidence in the estimated effect. The true effect may be very different from the estimated effect	It is very likely that subsequent studies change our confidence in the estimate of effect
Very low	Very little confidence in the estimated effect. The true effect is very probably different from the estimated effect.	Any estimate is very uncertain

Table 3 – Classification of the Level of Evidence According to the GRADE System.

Type of study	A priori quality level	Decreases if	Increases if	A posteriori quality level
Randomised studies	High	Risk of bias –1 significant –2 very significant	Effect +1 large +2 very large	High
		Inconsistency –1 significant –2 very significant	Dose-response +1 obvious gradient	Moderate
Observational studies	Low	No direct evidence –1 significant	All confounding factors: +1 would reduce observed effect	Low
		–2 very significant	+1 would suggest a spurious effect if there is no observed effect	
		Imprecision –1 important –2 very important		Very low
		Publication bias –1 likely –2 very likely		

or raised. Once the analysis has been completed, the GRADE experts summarise all the evidence with regard to the specific questions and the outcome variables chosen beforehand, in summary tables which they term evidence profiles [GRADE EP] and summary of findings [GRADE SoF]. EP and SoF tables have different purposes and are aimed at different collectives.^{15,16}

EP tables are more wide-ranging, they present the relevant findings for each key outcome, providing them in lines in the different columns, where the number of studies and the number of patients are expressed, the design (randomised or observational), the comparisons made, observed effect estimates in terms of relative effect RR (with its 95% CI) and absolute effect, and they also include an explicit assessment of the factors which weigh the quality of the studies (design limitations, inconsistency, indirectness, publication bias, etc.). Lastly they categorise the quality of the evidence for each outcome in plus signs (+) from 4 to 1, in other words, high, moderate, low or very low quality, respectively, with the meaning shown in Table 3. For tables which cover the evidence levels for questions relating to diagnostic tests, the format may be different. Evidence profiles are aimed at a small collective of CPG reviewers and creators and to anybody who questions or wants to check the goodness-of-fit of an assessment.

SoF tables are more concise and only offer the relevant findings of each outcome, i.e., as we mentioned earlier, the number of studies and the number of patients, comparisons made, observed effect estimates in terms of relative effect RR (with its 95% IC) and absolute effect. They also express the quality awarded. The summaries of findings are aimed at a wider collective, principally users of CPG and readers of systematic reviews. There is software (GRADEpro) for creating EP and SoF tables.¹⁴ For more information on EP and SoF consult Guyatt et al.¹⁶

In this context, we highlight a table in the ACCP guidelines for *antithrombotic therapy and prevention of thrombosis* (9th ed.) which summarises the evidence for starting pharmacological thromboembolism prophylaxis in surgical patients (Table 4).

We observe a decrease in the quality of evidence in 2 of the outcomes of interest (fatal pulmonary embolism and non-fatal symptomatic venous thromboembolic disease); in one case, due to the imprecision associated with possibly not achieving an effect, and in the other, due to limitations in the design of one of the studies. If we make a comparison with the 8th edition of these guidelines, we observe that the level of evidence and the degree of recommendation for pharmacological thromboembolic prophylaxis in moderate risk patients undergoing surgery has decreased in the new guidelines.^{17,18}

Degree of Recommendation

The GRADE system sets out recommendations based on a series of considerations.^{8,9,11,12} These are as follows: (1) risk-benefit balance: this is based on the knowledge that the majority of clinicians will offer patients therapeutic or preventive measures as long as the advantages of the intervention exceed its risks and collateral damage. The certainty or uncertainty of the clinician in considering the risk-benefit balance will greatly determine the strength of the recommendation. (2) Quality of evidence: the second factor is the methodological quality of the studies for each outcome variable, this factor is weighted by the details we set out above and which could raise or lower the level of evidence. In general, the degree of recommendation follows the level of evidence but not always. (3) The values and preferences of the patients have also to be considered. To that end a value judgement needs to be made and the values and preferences of the population in our area need to be established with any possible individual differences. (4) An estimate of resource consumption and costs.

There are still no appropriate studies which analyse patients' values and preferences in specific situations. In any event, values and preferences strengthen the degree of the recommendation when there is high concordance and weaken it when there is variability. Cost analysis usually

Table 4 – Effect of Prophylaxis Using Low Molecular Weight Heparin Compared to no Prophylaxis for Thromboembolytic Disease in Surgical Patients.

Outcome of interest	No. of participants (studies)	Quality of evidence (GRADE)	Relative effect (95% CI)	Comparative risk (95% CI)	
				No prophylaxis group	LMWH group
Fatal PE (follow-up: 7–270 d)	5142 (5 studies)	Moderate ^a	RR 0.54 (0.27–1.1)	Low risk population	
				3‰	2‰ (1–3)
				Moderate risk population	
				6‰	3‰ (2–7)
Fatal haemorrhage (follow-up: 21–270 d)	5078 (4 studies)	Moderate		High risk population	
				12‰	6‰ (3–13)
				Low risk population	
				1‰	0‰ (0–0)
Non-fatal symptomatic TED (follow-up: 21–270 d)	4890 (3 studies)	Moderate ^b	RR 0.31 (0.12–0.81)	High risk population	
				2‰	0‰ (0–0)
				Low risk population	
				15‰	5‰ (2–12)
Non-lethal haemorrhage (follow-up: 7–270 d)	5457 (7 studies)	High	RR 2.03 (1.37–3.01)	Moderate risk population	
				30‰	9‰ (4–24)
				High risk population	
				60‰	19‰ (7–49)
				Low risk population	
				12‰	24‰ (16–36)
				High risk population	
				22‰	45‰ (30–66)

Source: adapted from Mismetti et al.¹⁷
PE: pulmonary embolism; TED: thromboembolic disease; LMWH: low molecular weight heparin.
^a 95% CI includes the possibility of no effect (>1).
^b There were limitations with the design of one study.

requires the services of health economy experts. In general it is considered that an intervention can be classified as very cost-effective if it costs < the average per capita income of a country or region per quality adjusted life year (QALY) gained. Up to 3 times the average per capita income per QALY gained may be tolerable. Threshold tables have been developed on this subject.^{19,20}

Finally, the recommendations are simply graded in binary form as: strong (grade 1) or weak (grade 2), either for or against. A strong recommendation implies that the great majority of patients would agree (or disagree) with the recommended action. Clinicians should implement the action for most patients and the health authorities would have to adopt the recommendation as a health policy in the majority of situations. A weak recommendation implies that the majority of patients would accept (or reject) the recommended action, but a significant number of them would not. Clinicians must recognise that there are different options that are appropriate for different patients and, in this case, each patient requires help to reach the decision which is most consistent with their values and preferences. The health authorities would have to debate with the interest groups whether this recommendation should be implemented.

In the example we gave earlier, we observed that pharmacological pulmonary thromboembolism prophylaxis significantly reduces the risk of non-fatal venous thromboembolic disease (TED), not of fatal pulmonary embolism (PE), with an increased risk of major non-fatal haemorrhage (risk-benefit balance). On the other hand, we can see how the quality of evidence reduces as a result of the imprecision noted in the outcomes of fatal PE and the limitations in the design of a particular study. In short, a weak recommendation is established in favour of the use of pharmacological prophylaxis for patients with moderate thromboembolic risk.¹⁸

Limitations and Future of the GRADE System and its Use in Spain

Certain limitations should be highlighted. Firstly, the method was developed to deal above all with questions related to alternative interventions, treatment or prevention, not risk or prognosis and it entails difficulties in terms of diagnostic tests, public health and health system issues. Secondly, it only covers steps 3–6 (Table 1) in the elaboration of a CPG. And

thirdly, although the system makes highly systematic, transparent and reproducible judgements, it does not completely eliminate any disagreements which might exist when assessing a piece of evidence or when deciding alternative courses of action, given that there is always a subjective element in any judgement.

For those wishing to go into the GRADE method in more depth, such as authors of systematic reviews or health technology assessment studies, CPG panellists and methodologists, there is a wide-ranging and thorough series of sequential articles in this regard which have been published in the *Journal of Clinical Epidemiology* between 2011 and 2013 and is yet to be completed.^{15,21-32}

In Spain, several prestigious scientific journals of significant impact have covered the GRADE phenomenon—the *Revista de Atención Primaria*,³³ *Medicina Clínica*,³⁴ *Archivos de Bronconeumología*³⁵ and *Revista Española de Cardiología*.¹³ Its use has also been reported in health technology assessment³⁶ and in the development of CPG.³⁷⁻³⁹

Conflict of Interests

There is no conflict of interest.

REFERENCES

- Romero Simó M, Soria Aledo V, Ruiz López P, Rodríguez Cuéllar E, Aguayo Albasini JL. Guidelines and clinical pathways. Is there really a difference? *Cir Esp*. 2010;88:81-4.
- Jaeschke R, Guyatt GH, Dellinger P, Schünemann H, Levy MM, Kunz R, Norris S, Bion J. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a774.
- Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301:831-41.
- Phillips B, Ball C, Sackett DL. Levels of evidence and grades of recommendations. Oxford: Centre for Evidence-Based Medicine; 1998.
- Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ*. 2001;323:334-6.
- Guyatt GH, Schünemann HJ, Cook D. Applying the grades of recommendation for antithrombotic and thrombolytic therapy: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126:1795-875.
- Upshur RE. Are all evidence-based practices alike. Problems in the ranking of evidence. *CMAJ*. 2003;169:672-3.
- Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I. Critical appraisal of existing approaches. The GRADE Working Group. *BMC Health Serv Res*. 2004;4:38 [accessed 01.03.13]. Available from: <http://www.biomedcentral.com/1472-6963/4/38>
- Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, et al. Systems for grading the quality of evidence and the strength of recommendations II. Pilot study of a new system. *BMC Health Serv Res*. 2005;5:25 [accessed 01.03.13]. Available from: <http://www.biomedcentral.com/1472-6963/5/25>
- Arceles Martínez JI. Las nuevas guías de prevención y terapia antitrombótica del American College of Chest Physicians (ACCP). *Rev Esp Anestesiología Reanim*. 2012. <http://dx.doi.org/10.1016/j.redar.2012.04.024>.
- Grupo de trabajo sobre GPC. Elaboración de guías de práctica clínica en el Sistema Nacional de Salud. Manual metodológico. Guías de Práctica Clínica en el SNS: ICS N° 2006/01. Madrid: Plan Nacional para el SNS del MSC. Instituto Aragonés de Ciencias de la Salud-ICS; 2007.
- Grondin SC, Schiemann C. Evidence-based medicine levels of evidence and evaluation systems. In: Ferguson MK, editor. *Difficult decisions in thoracic surgery* London: Springer-Verlag London Limited; 2011.
- Alonso-Coello P, Solà I, Ferreira-González I. La formulación de recomendaciones con GRADE: cuestión de confianza. *Rev Esp Cardiol*. 2013;66:163-7.
- Brozek J, Oxman A, Schünemann HJ. GRADE pro (computer program) version 3.2 for Windows [accessed 22.03.13]. Available from: <http://mcmaster.flintbox.com/technology.asp?Page=3993>.
- Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol*. 2011;64:380-2.
- Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383-94.
- Mismetti P, Laporte S, Darmon JY, Buchmüller A, Decousus H. Meta-analysis of low molecular weight heparin in the prevention of venous thromboembolism in general surgery. *Br J Surg*. 2001;88:913-30.
- Douketis JD, Spyropoulos AC, Spencer FA, Mayr M, Jaffer AK, Eckman MH, et al. Perioperative management of antithrombotic therapy: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians. Evidence based clinical practice guidelines. *Chest*. 2012;141(2 Suppl.):e326S-50S.
- World Health Organization. Threshold values for intervention cost-effectiveness by region. Geneva, Switzerland: World Health Organization; 2010.
- Guyatt HG, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, et al. Methodology for the development of antithrombotic therapy and prevention of thrombosis guidelines. *Chest*. 2012;141:53S-70S.
- 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.
- Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz J, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64:401-6.
- Guyatt HG, 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.
- Guyatt HG, 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.
- 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.
- 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.
- 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.
- 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.

29. Brunotti M, Shelmit I, Pregno S, Vale L, Oxmann AD, Lord J, et al. GRADE guidelines: 10. Rating the quality of evidence for resource use. *J Clin Epidemiol*. 2013;66:140-50.
30. Guyatt GH, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66:151-7.
31. 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.
32. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. 13 GRADE guidelines: 13. Preparing summary of findings tables-continuous outcomes. *J Clin Epidemiol*. 2013;66:173-83.
33. Marzo Castillejo M, Montaña Barrientos A. The GRADE system in taking clinical decisions and elaboration of recommendations and clinical practice guidelines. *Aten Primaria*. 2007;39:457-60.
34. Alonso-Coello P, Rigau D, Solà I, Martínez García L. Formulating health care recommendations: the GRADE system. *Med Clin (Barc)*. 2012. <http://dx.doi.org/10.1016/j.medcli.2012.10.012>.
35. Alonso-Coello P, Rigau D, Sanabria AJ, Plaza V, Miravittles M, Martínez L. Quality and strength. The GRADE system for formulating recommendations in clinical practice guidelines. *Arch Bronconeumol*. 2013. <http://dx.doi.org/10.1016/arbres.2012.12.001>.
36. Ibarгойen-Roteta N, Gutiérrez Ibarluzea I, Rico-Iturrioz R, López-Argumedo M, Reviriego-Rodrigo E, Cabriada Nuño JL, et al. The GRADE approach for assessing new technologies as applied to apheresis devices in ulcerative colitis. *Implement Sci*. 2010. <http://dx.doi.org/10.1186/1748-5908-5-48>.
37. Marzo Castillejo M, Rotaecche del Campo R, Basora Gallifa J. SemFYC also adopts the GRADE system. *Aten Primaria*. 2010;42:191-3.
38. Gisbert JP, Calvet Calvo X, Ferrándiz Santos J, Mascort Roca JJ, Alonso-Coello P, Marzo Castillejo M. Managing of the patient with dyspepsia. Clinical practice guideline. Update 2012. *Aten Primaria*. 2012;44:728-33.
39. Gomollón F, García-López S, Sicilia B, Gisbert JP, Hinojosa J. Therapeutic guidelines on ulcerative colitis: a GRADE methodology based effort of GETECCU. *Gastroenterol Hepatol*. 2013;36:104-14.