Special Article

Quality and Strength: The GRADE System for Formulating Recommendations in Clinical Practice Guidelines☆

Calidad y fuerza: el sistema GRADE para la formulación de recomendaciones en las guías de práctica clínica

Pablo Alonso-Coello,[a,b,∗] David Rigau,[a,b] Andrea Juliana Sanabria,[a] Vicente Plaza,[c] Marc Miravitlles,[d] Laura Martinez[a,b]

[a] *Centro Cochrane Iberoamericano, Instituto de Investigación Biomédica (IIB-Sant Pau), Barcelona, Spain*
[b] *CIBER de Epidemiología Clínica y Salud Pública (CIBERESP), Barcelona, Spain*
[c] *Servicio de Neumología, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain*
[d] *Servicio de Neumología, Hospital Universitari Vall d'Hebron, Barcelona, Spain*

A R T I C L E   I N F O

Introduction

During the last decade, clinical practice guidelines (CPG) have experienced a huge surge and have become a fundamental tool in decision-making. These guidelines bring together the very best information available in the form of recommendations for clinical practice. In recent years, there have also been significant advances in the methodology for developing, updating and implementing CPGs.[1] These advances have led to greater attention being paid to the multidisciplinary composition of the groups charged with guideline development, including patients, conflict of interest management, exhaustive literature searches and the detailed evaluation of the quality and grading of the strength of the recommendations, among others.[2–5]

The Problem and a Potential Solution

Despite advances in the development of CPGs, there is still room for improvement in terms of quality.[6] One source of confusion is the availability of various systems for evaluating the quality of the evidence and the strength of the recommendations, with their greater or lesser limitations.[7,8] These systems are essential for allowing users to determine the confidence that they can place in the information provided by a CPG. For example, the first asthma CPGs guidelines used a system that did not grade the strength of the recommendations.[9] This is currently considered crucial, since there are other factors, in addition to the available evidence and its quality, which must be taken into account when developing recommendations and grading their strength (e.g. the risk-benefit ratio or costs).

In this context, an international group of epidemiologists, methodologists and clinicians from the major institutions responsible for developing CPGs have come up with a proposal, with the aim of agreeing on a common system which overcomes the limitations of previous systems.[10,11] This panel of professionals forms the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group. The GRADE system has been adopted by over 70 organisations throughout the world, including some as important as the WHO, the Cochrane Collaboration, National Institute of Clinical Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN) and publications such as *Clinical Evidence* and *Uptodate* (http://www.gradeworkinggroup.org/society/index.htm). In Spain, the National Programme for the Development of Clinical Practice Guidelines of the National Health System (http://www.guiasalud.es/web/guest/gpc-sns), and the GEMA, GesEPOC and semFYC guidelines and others have already adopted or used this system.[1,12–14]

What is the Difference Between GRADE and Other Systems?

The main differences between GRADE and other systems are the following:

- It evaluates the relative importance of the outcomes of interest for clinicians and patients.
- It differentiates clearly between quality of the evidence and strength of the recommendation.
- It provides explicit criteria for increasing or decreasing the quality of the evidence regardless of the study design (randomised clinical trial [RCT] or observational study).
- It considers values and preferences in the formulation of recommendations.
- It proposes a structured and specific process for developing recommendations.

These characteristics, along with a wide international consensus, make the GRADE system a systematic, explicit and transparent methodological framework for grading the quality of evidence and strength of recommendations.

The Importance of Delimiting the Clinical Question and Outcomes of Interest

One of the first steps in developing a CPG, regardless of the system used for evaluating the quality of evidence and strength of recommendations, is the definition of the clinical question by the developing group. This question must be well constructed, so a PICO (Population, Intervention, Comparison and Outcome) format is usually used.

The GRADE system is particularly important when considering outcomes of interest (e.g. exacerbation of symptoms or serious adverse effects), since these define the balance between the risks and benefits of the intervention under evaluation. The GRADE system specifies that not all outcomes of interest have the same importance and, accordingly, only the most relevant should influence our evaluation of the quality of evidence and the grading of recommendations.[15] Specifically, the outcomes are divided into the following categories: critical; important but not critical; and not important. The critical outcomes are those which should be given the most weight. This proposal implies that a group of guideline developers has to evaluate the relative importance of the outcomes included, as well as the perspective of the patient.

For a CPG for the management of patients with allergic rhinitis and asthma, for example, the developers considered that for the preventive treatment of asthma, the reduction of symptoms and reduction of exacerbations were critical outcomes for patients (Table 1).[16] Important but not critical outcomes included quality of life and adverse events. Spirometric or blood gas results were considered as unimportant outcomes. Only critical and important outcomes were taken into account throughout the development process.

Confidence in the Available Evidence

Guideline users need to know how much confidence they can place in the study results. This confidence constitutes the so-called quality of evidence. GRADE defines quality as the degree of confidence we have that the effect estimates are adequate to support a recommendation.[17] For example, in patients with stable chronic obstructive pulmonary disease (COPD), combined treatment with a long-acting beta-2 agonists and corticosteroids reduces the risk of exacerbations by 28% compared to placebo (RR 0.72; 95% CI 0.65–0.80).[14] This 28% reduction in the risk of exacerbations with the combined treatment is the estimated effect of the intervention. Confidence in this estimation depends on multiple factors, such as the limitations of the study design and conduct (risk of bias), consistency or accuracy of the results, and others.[17] The GRADE system evaluates quality for each of the outcomes considered critical within the same question of interest. In the example of the treatment of advanced COPD, in addition to the risk of exacerbations, the CPG developers could also take into consideration improvement in night-time symptoms. Thus, the quality of that outcome would also be evaluated, along with others, if necessary.

The GRADE system classifies quality of evidence as:

- High quality: high confidence that the estimate of the effect from the available literature is very close to the true effect.
- Moderate quality: the estimate of the effect is close to the true effect, but there may be substantial differences.
- Low quality: the estimate of the effect may be substantially different from the true effect.
- Very low quality: it is very likely that the estimate of the effect is substantially different from the true effect.

Which Factors Affect the Quality of the Evidence?

The different factors that can reduce confidence in the estimation of the observed effect are summarised in Table 2. RCTs, which initially provide high quality in evaluating the effect of interventions, are distinguished from observational studies, which initially are considered to provide low quality. Computation of these confidence-limiting factors will determine whether our confidence rises or falls. The GRADE system establishes that overall quality is equivalent to the lowest quality of all the critical outcomes considered.[18] Finally, it recognises that expert opinion influences the evaluation of the available evidence (regardless of the design) but it is not considered as a type of evidence in itself.

The factors that affect quality and, accordingly, the confidence that can be placed in the estimation of the effect, are described below.

Limitations in Design or Conduct

Limitations in design and conduct (risk of bias) differ between RCTs and observational studies. The following factors are taken into consideration in RCTs: lack of concealment of the randomisation

Table 1

PICO Question Components. Should Leukotriene Receptor Antagonists be Used for Treatment of Asthma in Patients With Allergic Rhinitis and Asthma?

Patients	Intervention	Comparator	Outcomes and Importance[a]
Patients with allergic rhinitis and asthma	Inhaled corticosteroids+leukotriene receptor antagonists	Inhaled corticosteroids	Reduction of daytime symptoms (7–9)[a] Reduction of night-time symptoms (7–9)[a] Reduction of exacerbations (7–9)[a] Quality of life (4–6)[a] Adverse events (4–6)[a] Spirometric results (1–3)

[a] GRADE suggests classifying the importance of outcomes on a 9-point scale: (1) 1–3: unimportant outcome; (2) 4–6: important but not critical outcome for decision-making; (3) 7–9: critical outcome for decision-making.

Table 2
Evaluation of Quality and Modifying Factors.

Study design	Quality of evidence	Lower if	Higher if
Randomised controlled trial→	High	*Risk of bias* −1 High −2 Very high	*Large effect* +1 Large +2 Very large
	Moderate	*Inconsistency* −1 Serious −2 Very serious	*Dose–response* +1 Evidence of a gradient
Observational study→	Low	*Indirectness* −1 Serious −2 Very serious	*All plausible confounding factors* 1+ Would reduce a demonstrated effect or 1+ Would suggest a spurious effect when results show no effect
	Very low	*Imprecision* −1 Serious −2 Very serious *Publication bias* −1 Likely −2 Very likely	

Adapted from: Guyatt G, 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(April (4)):383–94.

sequence, inadequate blinding, substantial loss to follow-up and the lack of intention-to-treat analyses, selective inclusion of outcomes of interest, and other less common factors, such as early termination of a study due to benefit, use of non-validated measurements, the carry-over effect in crossover clinical trials and bias in recruitment in cluster randomised trials.[19]

In observational studies, the following are taken into consideration: the presence of inappropriate population selection criteria, inappropriate measurements for exposure or the outcome of interest, insufficient control of confounding factors or incomplete follow-up.[19]

Inconsistent Results

The quality of evidence is reduced if the results are inconsistent or heterogeneous, i.e. if the results from various studies are very different. It should also be evaluated if the inconsistencies persist after the reasons which might explain any observed heterogeneity (e.g. differences in population, intervention, result outcomes or risk of bias) have been examined. If no reasons to explain the variability are identified, confidence is reduced, since there may be real differences between the effect estimations provided by the studies included.[20]

For example, a systematic review evaluating the efficacy of specific allergen immunotherapy vs placebo in adults with allergic rhinitis shows that the results for nasal symptoms are very different among the various studies, the confidence intervals do not overlap, the heterogeneity test is significant and the I^2 statistic is high.[21] In this kind of situation, there is less confidence in the results and for this reason, quality must be reduced (Fig. 1).

Lack of Direct Evidence

In situations where there is no direct comparison between the interventions under consideration or significant differences in the available studies and population, the interventions or outcomes proposed in the question of interest, it may be that only indirect information is available.[22]

For example, in the case of outcomes, the administration of 2 long-acting bronchodilators for the treatment of COPD was compared to administration of one long-acting beta-adrenergic bronchodilator combined with an inhaled corticosteroid. In this case, a single RCT provides spirometric results but no patient-relevant clinical outcomes (e.g. improved symptoms). Our confidence that an improvement in spirometric results reflects an improvement in the outcomes which mean more to patients is uncertain, and so confidence is lower.[23] Regarding the same question of interest, a meta-analysis was published providing results on the frequency of exacerbations based on studies that evaluated the administration of 2 bronchodilators, or that evaluated the administration of a long-acting beta-adrenergic bronchodilator plus an inhaled corticosteroid. However, no direct comparisons are available between these treatment strategies, so the estimation obtained from this meta-analysis is indirect and thus less trustworthy or of poorer quality.[24]

In the case of antihistamine treatment in patients with asthma and allergic rhinitis, the available evidence is indirect, due to differences in the population: the RCTs include up to 60% of patients without asthma at the start of the trial. Similarly, another example would be the evaluation of the efficacy of nasal decongestants as rescue therapy in patients with allergic rhinitis. The studies identified analyse the efficacy of the regular use of nasal
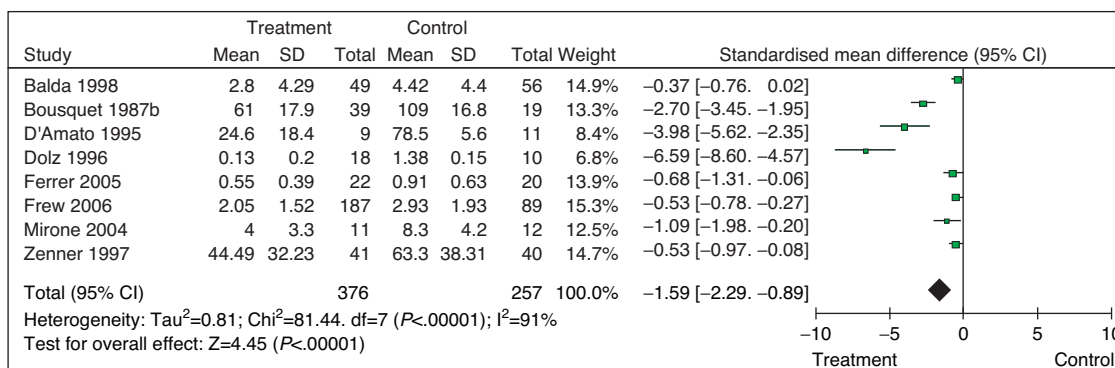

Study	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Standardised mean difference (95% CI)
Balda 1998	2.8	4.29	49	4.42	4.4	56	14.9%	−0.37 [−0.76. 0.02]
Bousquet 1987b	61	17.9	39	109	16.8	19	13.3%	−2.70 [−3.45. −1.95]
D'Amato 1995	24.6	18.4	9	78.5	5.6	11	8.4%	−3.98 [−5.62. −2.35]
Dolz 1996	0.13	0.2	18	1.38	0.15	10	6.8%	−6.59 [−8.60. −4.57]
Ferrer 2005	0.55	0.39	22	0.91	0.63	20	13.9%	−0.68 [−1.31. −0.06]
Frew 2006	2.05	1.52	187	2.93	1.93	89	15.3%	−0.53 [−0.78. −0.27]
Mirone 2004	4	3.3	11	8.3	4.2	12	12.5%	−1.09 [−1.98. −0.20]
Zenner 1997	44.49	32.23	41	63.3	38.31	40	14.7%	−0.53 [−0.97. −0.08]
Total (95% CI)			376			257	100.0%	−1.59 [−2.29. −0.89]

Heterogeneity: Tau^2=0.81; Chi^2=81.44. df=7 (*P*<.00001); I^2=91%
Test for overall effect: Z=4.45 (*P*<.00001)

Fig. 1. Efficacy of specific allergen immunotherapy compared to placebo in adults with allergic rhinitis. 95% CI: 95% confidence interval; SD: standard deviation. Adapted from: Calderon MA, et al. Allergen injection immunotherapy for seasonal allergic rhinitis. Cochrane Database of Systematic Reviews 2007, Issue 1.

decongestants (not in rescue situations), so the evidence available is also indirect. In both cases, the confidence that can be placed in the data from these studies for answering the questions raised is accordingly lower.[16]

Imprecise Results

For the effect of an intervention to be considered imprecise, the estimator of the effect must be evaluated, preferably in absolute terms, rather than relative terms, along with the corresponding confidence interval. If our recommendation were to change depending on which end of the confidence interval for an outcome is considered, taking into account the risks and disadvantages of the intervention, confidence in the effect estimator would fall due to its imprecision. Furthermore, even if the confidence interval is precise, if the number of events or number of subjects evaluated in the different studies is low, confidence may also have to be reduced.[25]

For example, in a recent guideline, the efficacy of H_1-antihistamines in reducing the development of asthma in children with various types of allergy was compared to placebo.[16] The results of the 3 RCTs show that H_1-antihistamines do not significantly reduce the risk of developing asthma. The absolute benefit shows that one end of the confidence interval provides a significant benefit (10 children fewer for each 100 treated will develop asthma, compared to placebo), which would generate a recommendation in favour of H_1-antihistamines, but the other end of the same interval showed harm (31 more children per 100 treated will develop asthma, compared to placebo), and this would generate a recommendation against this treatment. The panel of guideline authors decided to reduce confidence in this outcome due to imprecision.[16]

Suspected Publication Bias

Finally, in some situations, there may a suspicion that not all of the studies, primarily those with negative results, have been published, so there is a possibility that the effect may be overestimated.[26] This possibility must be examined if a set of small, positive, industry-funded trials is presented.[27] There are a number of statistical tests and plots for detecting this possible bias, the most popular being the funnel plot. In these cases, confidence in the estimation of an effect would be reduced.

What Factors Increase Quality of Evidence?

Situations in which increasing confidence in the results of a set of studies is justified are rarer and mainly apply to observational studies (cohorts and case–control), provided there are no other limitations in design and conduct (risk of bias).[28]

Strong Association

When the results of a study with no other limitations show an effect, whether protective or harmful, with a strong (relative risk or odds ratio >2 or <0.5) or very strong (relative risk or odds ratio >5 or <0.2) association, confidence in those results increases.[28] One example is the relationship observed between all-cause mortality and tobacco use, which was up to 3 times higher in smokers compared to non-smokers, in a prospective cohort of British doctors.[29] Confidence in this association is therefore at least moderate.

Dose–Response Gradient

A clear dose–response gradient can also be a reason for increasing confidence in the estimation of an effect, since it provides greater certainty about a potential cause–effect relationship. For example, it has been shown that the risk of developing COPD is proportional to cumulative tobacco use, being 2.6 times higher in smokers of 15–30 packets a year, and 5.1 times higher in smokers of more than 30 packets a year.[30] This gradient associating the factor

under study and the effect increases confidence in the relationship between tobacco use and COPD.

Potential Confounding Factors and Residual Bias

Occasionally situations can arise in which an effect associated with an intervention is observed, and, after the potential factors which could reduce the observed effect have been analysed, these factors, if they exist, can be considered to strengthen the conclusions obtained.[28] For example, a systematic review of observational studies showed a higher mortality rate in private for-profit hospitals, compared to private not-for-profit hospitals, even though the latter possibly had more serious patients.[31]

The GRADE system allows the evidence to be combined in a Summary of Findings (SoF) table, which gives a structured outline of the number of studies for each outcome of interest, quality of evidence and the results observed in relative and absolute terms. These SoF can be generated using a free download software programme called GRADEPro.[32]

One CPG on allergic rhinitis and asthma evaluated the use of single-agent oral leukotriene antagonists for the background treatment of asthma, compared to inhaled corticosteroids, in patients with allergic rhinitis and asthma.[16] A summary table of the available evidence for outcomes of interest is shown in Table 3. With regard to exacerbations requiring the use of systemic corticosteroids, it was observed that in absolute terms, these exacerbations are clearly reduced in patients using inhaled corticosteroids, compared to those using leukotriene antagonists. The group of patients receiving leukotriene receptor inhibitors had 30 more exacerbations per 1000 patients, compared to the group receiving inhaled corticosteroids (high quality). Conversely, the leukotriene receptor inhibitors produced fewer adverse effects (four fewer per 1000 patients) than inhaled corticosteroids (moderate quality).

Why Must the Strength of Recommendations be Graded?

Guideline users have to determine quickly how much they can trust that a recommendation will produce more desirable than undesirable consequences. The strength of the recommendation reflects a confidence gradient, with greater confidence in strong recommendations and lesser confidence in weak recommendations. In turn, the direction may be in favour of or against the recommendation (Fig. 2). Recommendations, whether strong or weak, have different implications for patients, healthcare professionals or management (Table 4).

According to GRADE, 4 basic factors influence the strength of recommendations: the risk-benefit balance, quality of evidence, patient values and preferences and finally, costs and resource utilisation.[33]

Risk-Benefit Balance

The balance between the effect of desirable and undesirable outcomes must be determined. To make this balance, a weight or a value must be assigned to the outcomes. This is done implicitly whenever the pros and cons of a decision are evaluated. However the guideline developers must specify these values as far as
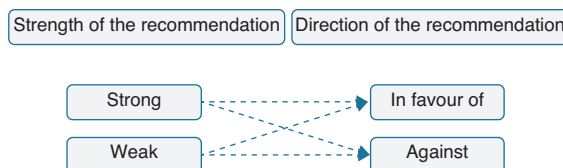

Fig. 2. Strength and direction of recommendations.

Table 3
Summary of Findings on Leukotriene Antagonists Compared to Inhaled Corticosteroids for the Treatment of Asthma in Patients With Allergic Rhinitis and Asthma.[a, b]

Outcomes	Participants (Studies) Follow-up[b]	Quality of Evidence (GRADE)	Relative Effect (95% CI)	Absolute Effect
Exacerbation requiring use of systemic steroids	1018 (2) 6–40 weeks	⊕⊕⊕⊕ High	RR 1.56 (1.36–2.00)	30 more per 1000 (between 19 and 53 more)
Hospital admission due to exacerbation	3189 (13) 6–40 weeks	⊕⊕⊕◯ Moderate[c]	RR 1.62 (0.64–4.15)	2 more per 1000 (between 1 fewer and 9 more)
Quality of life: change from baseline, measured using the asthma quality of life questionnaire, higher score signifies better quality of life	1027 (2) 8–16 weeks	⊕⊕⊕◯ Moderate[d]	–	MD −0.30 (−0.43 to −0.17)
Daytime symptoms (fewer=better)	2543 (6) 8–16 weeks	⊕⊕⊕⊕ High	–	SMD 0.29 (0.21 to 0.37)
Night-time symptoms (fewer=better)	1995 (6) 8–16 weeks	⊕⊕⊕⊕ High	–	SMD 0.21 (0.13 to 0.30)
Days without symptoms	1328 (5) 8–16 weeks	⊕⊕⊕⊕ High	–	MD −11.47 (−15.72 to −7.23)
Adverse effects	6277 (16) 6–40 weeks	⊕⊕⊕◯ Moderate[e]	0.99 (0.93–1.04)	4 fewer per 1000 (from 3 fewer to 13 more)

CI: confidence interval; RR: relative risk; MD: mean difference; SMD: standardised mean difference.

[a] Adapted from: Brozek JL, Bousquet J, Baena-Cagnani CE, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. J Allergy Clin Immunol. 2010;126(September (3)):466–76.

[b] Ducharme FM, Hicks GC. Anti-leukotriene agents compared to inhaled corticosteroids in the management of recurrent and/or chronic asthma in adults and children. Cochrane Database Syst Rev 2002;(3).

[c] Imprecision: 95% confidence interval includes no effect or 1% more patients requiring hospitalisation in absolute terms.

[d] Risk of bias, only 2 trials study this outcome.

[e] From 3% fewer to 2% more in absolute terms.

possible. When this balance shows a significant difference in the 2 types of outcome, it is more likely that a strong recommendation will be made. If the difference is more balanced, it is more appropriate to assign a weak recommendation. For example, in the case of inhaled corticosteroids for the maintenance treatment of persistent asthma, the benefits outweigh the risks and disadvantages. In this context, the GEMA guideline gave a strong recommendation in favour of the treatment.[12] However, in the case of severe asthma poorly controlled with inhaled corticosteroids and a long-acting beta-2 agonist, the formulated recommendation is weak, suggesting the use of oral corticosteroids, due to a more uncertain risk-benefit balance.[12]

Quality of Evidence

It is essential to know how far the estimation of the effect can be trusted for critical outcomes. When quality is high, it is more likely that a strong recommendation is formulated, and in contrast, if the quality is low, it is more likely that the recommendation will be weak. However, there are situations which justify a strong recommendation, even if only evidence of low or very low quality is available. For example, in pregnant women with asthma, the GEMA 2009 guidelines formulate a strong recommendation for not withdrawing maintenance treatment with corticosteroids plus

long-acting beta-2 adrenergic agonists due to the well-known risk of exacerbation after discontinuation, despite the availability of low quality evidence on the foetal toxicity of this combination.[12]

Values and Preferences

GRADE includes values and preferences as another of the factors to be evaluated when grading the strength of recommendations. Patients often have different opinions about what an outcome (and, as such, a treatment) involves, and the opinion of healthcare professionals often differs from that of patients.[34] Accordingly, the values and preferences of the patients must be taken into account in grading the strength of a recommendation. If confidence in these values and preferences is high and variability is low, it is more likely that the recommendation will be strong (and vice versa). Moreover, guideline developers should specify which values have been used for formulating the recommendations and their sources (e.g. taken from the literature or estimated from their interaction with patients in the decision-making process). In the case of the above-mentioned guideline, regarding the question about whether pre-schoolers with other allergic diseases should be treated with oral H_1-antihistamines to prevent wheezing or asthma, the developers specified that their recommendation assigns higher importance to avoiding the side effects of these drugs

Table 4
Implications of Strength of Recommendations.

	Strong recommendation	Weak recommendation
For patients	Most people in your situation would want the recommended course of action and only a small proportion would not	Most people in your situation would want the recommended course of action, but many would not
For clinicians	Most patients should receive the recommended course of action	You should recognise that different choices will be appropriate for different patients and that you must help each patient to arrive at a management decision consistent with her or his values and preferences
For policy makers	The recommendation can be adopted as a policy in most situations	There is a need for substantial debate and involvement of stakeholders

Adapted from: Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, Schünemann HJ; GRADE Working Group. Rating quality of evidence and strength of recommendations: going from evidence to recommendations. BMJ. 2008;336(May (7652)):1049–51.

than to a very uncertain reduction in the risk of developing wheezing and asthma.[16]

Costs and Resource Utilisation

Costs derived from a clinical decision are difficult to quantify, as the information is frequently out of date or applies to other healthcare settings. Economic analyses must be performed after evaluation of the risk-benefit balance, and it is important to specify the perspective of this economic analysis (i.e. whether it is from the patients' viewpoint or that of the healthcare system). Direct and indirect costs, or both, or the use of short-term or longer-term resources can be taken into consideration. A high cost reduces the probability of formulating a strong recommendation in favour of an intervention, and in contrast, a low cost will increase it.

Integration of Factors

When recommendations are formulated, all the factors mentioned above must be included for determining the strength of these recommendations. This process requires a weighted and explicit balancing of the factors, and accordingly, it is important that this process is reflected in detail in the CPG.

As mentioned above, an example from the CPG for rhinitis and asthma and the exclusive use of leukotriene receptors in the treatment of asthma[16] is given in Table 5. Regarding the risk-benefit balance, inhaled corticosteroids, compared to leukotriene antagonists, showed a reduction in exacerbations, improved daytime and night-time symptoms and an improvement in days without symptoms and quality of life. With regard to adverse effects, leukotriene antagonists presented fewer effects than inhaled corticosteroids, but the results are imprecise. The quality of evidence was evaluated as moderate, due to the imprecision of this and other results for the outcomes of interest evaluated (Table 3). The cost of the treatments is different, as the inhaled corticosteroids are cheaper than that of leukotriene receptor antagonists. Integration of these factors led to the formulation of a strong recommendation in favour of the use of inhaled corticosteroids compared to single-agent oral leukotriene receptor antagonists for the control of asthma.

Finally, another interesting aspect is the terminology used in producing the recommendations. The use of specific terms (words, numbers, letters, symbols, etc.) should optimally describe the

strength of the recommendations given. The use of expressions such as "it is recommended/it is not recommended" when talking of strong recommendations or "it is suggested/it is not suggested" for weak recommendations are some examples of wording. However the information currently available on this subject is very limited.[35] Future studies, some of which are initiatives of the GRADE group itself, will address this and other issues regarding the best presentation and dissemination of healthcare recommendations.[36]

Conclusion and Future Implications

The formulation of recommendations is a complex process involving multiple judgments and significant investment of resources. GRADE has demonstrated the complexity inherent in the process, while providing a systematic, structured tool to allow the formulation of explicit recommendations. Different groups can come to different conclusions with GRADE, but if they adhere to the process and publish their recommendations, users can determine if they are in agreement with the judgments shaping the final recommendations. GRADE is highly accepted among the international community and is being adopted by the principal institutions of guideline development, both internationally and in Spain. In the case of guidelines in the field of pneumology, guidelines such as GEMA and GesEPOC have already used the GRADE system. On the international front, organisations such as the American Thoracic Society or the Global Initiative for Asthma (GINA) already use it or are beginning to do so. GRADE is thus emerging as the methodology which should improve the quality of guidelines, and definitively, the quality of patient care.

Conflicts of Interest

PAC, DR, AJS and LM are members of the GRADE.

References

1. Grupo de trabajo sobre GPC. Elaboración de Guías de Práctica Clínica en el Sistema Nacional de Salud. Manual Metodológico. Madrid: Plan Nacional para el SNS del MSC. Instituto Aragonés de Ciencias de la Salud-I+CS; 2007. Guías de Práctica Clínica en el SNS: I+CS. N.° 2006/0I.
2. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van derWees P. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156:525–31.
3. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. Ann Intern Med. 2011;154:774–5.
4. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. Can Med Assoc J. 2010;182:E839–42.
5. Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010;152:738–41.
6. Alonso-Coello P, Irfan A, Sola I, Gich I, Delgado-Noguera M, Rigau D, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Qual Saf Health Care. 2010;19:e58.
7. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to rate the strength of scientific evidence: summary in: Rockville MD. AHRQ Evidence Report Summaries. Agency for Healthcare Research and Quality (US); 2002.
8. The GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. BMC Health Serv Res. 2004;4:38.
9. Plaza V, Bellido-Casado J, Alonso-Coello P, Rodrigo G. Asthma clinical practice guidelines: advantages and pitfalls. Arch Bronconeumol. 2009;45 Suppl. 1:25–9.
10. Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al., for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.
11. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knotterus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2010;64:380–2.
12. GEMA 2009. Guía española para el manejo del asma. Arch Bronconeumol. 2009;45 Suppl. 7:1–35.
13. Grupo de Trabajo de GesEPOC. Guía de Práctica Clínica para el diagnóstico y tratamiento de pacientes con enfermedad pulmonar obstructiva crónica (EPOC)—Guía Española de la EPOC (GesEPOC). Arch Bronconeumol. 2012;48 Suppl. 1:2–58.

Table 5
Integration of Factors for Grading the Strength of a Recommendation.

Risk-benefit balance
Single-agent oral leukotriene receptor antagonists for the treatment of asthma are less effective than inhaled corticoids in improving the symptoms of asthma and in the reduction of exacerbations requiring the use of systemic steroids (30 more per 1000 patients treated). These drugs have a lower rate of adverse effects than inhaled corticosteroids (4 fewer per 1000 patients treated)

Quality of evidence
The quality of evidence is moderate given the imprecision of the results of some studies regarding the critical outcomes considered

Patient values and preferences
Patient values and preferences are probably not different for the critical outcomes considered. It is very likely that the vast majority of the patients will be in favour of taking corticosteroids

Costs and use of resources
Corticosteroids cost less than leukotrienes

Recommendation
In patients with allergic rhinitis and asthma, the use of inhaled corticosteroids instead of single-agent leukotriene receptor antagonists is recommended for the treatment of asthma (*strong recommendation, moderate quality of evidence*)

14. Grupo de trabajo de la guía de práctica clínica sobre Atención Integral al paciente con Enfermedad Pulmonar Obstructiva Crónica (EPOC). Desde la Atención Primaria a la Especializada. Sociedad Española de Medicina de Familia (semFYC) y Sociedad Española de Neumología y Cirugía Torácica (SEPAR); 2010.

15. 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.

16. Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. J Allergy Clin Immunol. 2010;12:466–76.

17. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines 3: rating the quality of evidence. J Clin Epidemiol. 2011;64:401–6.

18. Guyatt G, 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.

19. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines 4: rating the quality of evidence—risk of bias. J Clin Epidemiol. 2011;64:407–15.

20. 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.

21. Calderon MA, Alves B, Jacobson M, Hurwitz B, Sheikh A, Durham S. Allergen injection immunotherapy for seasonal allergic rhinitis. Cochrane Database Syst Rev. 2007, http://dx.doi.org/10.1002/14651858.CD001936.pub2. Art. No.: CD001936.

22. 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.

23. Rabe KF, Timmer W, Sagkriotis A, Viel K. Comparison of a combination of tiotropium plus formoterol to salmeterol plus fluticasone in moderate COPD. Chest. 2008;134:255–62.

24. Mills EJ, Druyts E, Ghement I, Puhan MA. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. Clin Epidemiol. 2011;3:107–29.

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

26. Sterne JAC, Egger M, Moher D, editors. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of intervention. Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011 [accessed 12.06.12]. Available from: www.cochrane-handbook.org

27. 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.

28. 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. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ. 2004;328:1519.

30. Miravitlles M, Soriano JB, García-Río F, Muñoz L, Duran-Taulería E, Sánchez G, et al. Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. Thorax. 2009;64:863–8.

31. Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schunemann H, Haines T, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. CMAJ. 2002;166:1399–406.

32. GRADEpro [programa informático]. Version 3.2 para Windows. Jan Brozek, Andrew Oxman, Holger Schünemann; 2008 [accessed 11.02.13]. Available from: http://ims.cochrane.org/gradepro

33. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al., GRADE Working Group. Going from evidence to recommendations. BMJ. 2008;336:1049–51.

34. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ. 2001;323:1218–22.

35. Akl EA, Guyatt GH, Irani J, Feldstein D, Wasi P, Shaw E, et al. "Might" or "suggest"? No wording approach was clearly superior in conveying the strength of recommendation. J Clin Epidemiol. 2012;65:268–75.

36. Treweek S, Oxman AD, Alderson P, Bossuyt P, Brożek J, Davoli M, et al., the DECIDE Consortium. Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): protocol and preliminary results. Implementation Sci. 2013;8:6.