

## Original article

## Methodology for development of the Allergic Rhinitis and its Impact on Asthma Guideline 2008 update

**Background:** We describe the methodology for the 2008 update of the Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines. The methodology differs from the 2001 edition in several respects. The most prominent change is the application of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to compiling evidence, assessing the quality of evidence and grading of recommendations.

**Methods and results:** Representatives of the GRADE working group joined the ARIA guideline panel to achieve these tasks. While most recommendations result from existing systematic reviews, systematic reviews were not always available and the panel compiled the best available evidence in evidence profiles without conducting actual reviews. The panel conducted two meetings and used the GRADE criteria to assess the quality of evidence (four categories of high, moderate, low and very low) and the strength of recommendation (strong and weak) based on weighing up the desirable and undesirable effects of management strategies, considering values and preferences influencing recommendations, and resource implications. The guideline panel has chosen the words ‘we recommend’ – for strong recommendations and ‘we suggest’ – for weak recommendations. Both categories indicate the best course of action for a given patient population, but their implementation, requires different considerations as we describe subsequently in this article.

**Conclusions:** The 2008 update of the ARIA guidelines has become more evidence-based. Future iterations of the guidelines will further be improved by following the described processes even closer, such as ensuring availability of updated high quality systematic reviews for each question.

**J. L. Brozek<sup>1</sup>, C. E. Baena-Cagnani<sup>2,3</sup>,  
S. Bonini<sup>4,5,\*</sup>, G. W. Canonica<sup>3,\*</sup>,  
G. Rasi<sup>4,6</sup>, R. G. van Wijk<sup>7,\*</sup>,  
T. Zuberbier<sup>8,\*</sup>, G. Guyatt<sup>9,10</sup>,  
J. Bousquet<sup>11,\*</sup>, H. J. Schünemann<sup>1,8</sup>**

<sup>1</sup>S.C. Epidemiologia, Italian National Cancer Institute ‘Regina Elena’, Rome, Italy; <sup>2</sup>Faculty of Medicine, Catholic University of Córdoba, Córdoba, Argentina; <sup>3</sup>Allergy & Respiratory Diseases, DIMI, Dept. of Internal Medicine, University of Genoa, Genoa, Italy; <sup>4</sup>Institute of Neurobiology and Molecular Medicine - CNR, Rome, Italy; <sup>5</sup>Department of Medicine, Second University of Naples, Naples, Italy; <sup>6</sup>Agenzia Italiana del Farmaco (ALFA), Rome, Italy; <sup>7</sup>Section of Allergology, Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands; <sup>8</sup>Department of Dermatology and Allergy, Charité Universitätsmedizin Berlin, Berlin, Germany; <sup>9</sup>CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada; <sup>10</sup>Department of Medicine, McMaster University, Hamilton, Canada; <sup>11</sup>Clinique des Maladies Respiratoires, CHU Montpellier and INSERM, France

Key words: allergic rhinitis; ARIA; asthma; evidence; GRADE; guidelines; recommendations.

Holger Schünemann MD, PhD  
Department of Epidemiology  
Italian National Cancer Institute ‘Regina Elena’  
Via Elio Chianesi 53  
00144 Rome  
Italy

\*Members of GA<sup>2</sup>LEN.

Accepted for publication 29 August 2007

A variety of approaches exist to grade levels of evidence and the strength of recommendations (1, 2). The large number of systems for measuring the quality of evidence and recommendations is confusing (3) and the approaches often have important shortcomings (1). In particular, the vast majority of guidelines such as the first Allergic Rhinitis and its Impact on Asthma (ARIA)

report (4), when moving from evidence to recommendations, focused solely on the quality of evidence for efficacy without considering other factors that influence the strength of recommendations explicitly (5).

An evidence-based approach to make health care decisions acknowledges that evidence alone is insufficient, and that values and preferences, clinical circumstances as well as clinical expertise inevitably influence decisions (6). The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to developing clinical practice guidelines suggests that panels follow the

*Abbreviations:* ARIA, Allergic Rhinitis and its Impact on Asthma; GRADE, Grades of Recommendation, Assessment, Development and Evaluation; WHO, World Health Organization.

sequential steps including explicitly defining the question that the recommendation is addressing, determining all important outcomes and their relative weight deciding on the eligibility criteria for the evidence to be considered and conducting a comprehensive search for evidence, evaluating the quality of evidence, summarizing the findings of the studies, balancing the desirable and undesirable consequences of the alternative management strategies, and finally, acknowledging the values and preferences underlying the recommendations. This process ends with formulating a recommendation for or against an action, and a grading of the strength of that recommendation based on the balance of desirable (benefits) and undesirable (harms and burden) consequences, and the methodological quality of the supporting evidence (7, 8).

The 'Guidelines for WHO guidelines' recommend using a specific, uniform grading system developed by the GRADE Working Group (9). The GRADE approach has been evaluated favourably by respected organizations (2) and is being used increasingly by many organizations, including the World Health Organization (WHO), American Thoracic Society, American College of Chest Physicians and the UK National Institute of Health and Clinical Excellence and other organizations around the globe (8).

In this article, we describe the methodology for the 2008 update of the ARIA guidelines using the GRADE approach. Factors that ideally would not, but often influence recommendations are resource utilization and conflicts of interest (10). To maintain transparency of this update of the ARIA guidelines we also considered resource utilization implicitly and we explicitly declared the potential conflicts of interest of the authors, listing them in the guideline document following the WHO criteria for declaring potential conflicts of interest (11).

### Preparatory steps

The ARIA guideline panel prepared a draft manuscript containing specific recommendations and statements about the management of allergic rhinitis and its impact on asthma. These recommendations resulted from the previous edition of ARIA guidelines (4) and the subsequent updates of the selected topics that the ARIA guideline panel published (12–16). This work was supplemented by efforts of an independent team of GRADE Working Group representatives (JLB and HJS). The panel ended with an assessment of the quality of underlying evidence and the judgement on its strength according to the GRADE system for each recommendation (7, 8). To achieve this, the panel applied the following detailed steps (8).

### Defining the clinical question

An evidence-based guideline should include clear health care questions followed by an evaluation of the evidence,

transparent judgements about the quality of evidence, and an unambiguous recommendation that answers each clinical question. Any clinical management question has four components that include the population of patients it refers to, the management strategy, alternative management strategies, and the outcomes of interest (17). For the ARIA guideline update, we asked a specific, structured clinical question for each recommendation about the prevention and treatment of allergic rhinitis as well as for the treatment of rhinitis and asthma in the same patient.

When two recommendations concerned the same population and the same intervention, but the outcomes of interest were different, we combined these recommendations into one that explicitly advises the clinician or patient what is the preferred course of action in certain population considering all the important outcomes (e.g. use of subcutaneous allergen-specific immunotherapy for the treatment of rhinitis symptoms and for the prevention of asthma in children with allergic rhinitis (weak recommendation based on very low quality evidence).

On occasion, when the same recommended course of action concerned different populations (e.g. children and adults), or there were specific circumstances in which the recommendation might not apply, the recommendation was composed of more than one statement. For example, a clinical question 'should immediate and total cessation of exposure or occupational allergen control be used in patients sensitive to occupational allergens?' is answered by the corresponding recommendation, that in patients sensitive to occupational allergens, we recommend immediate and total cessation of exposure and when total cessation of exposure is not possible, we suggest specific strategies aimed on minimizing occupational allergen exposure. In the following paragraphs we will describe the intermediate steps that were required to link a question like the one described to the corresponding recommendation.

### Identifying and summarizing evidence

The GRADE approach to developing guidelines postulates that before grading the quality of evidence and strength of each recommendation, guideline developers should first identify a recently well-done systematic review of the appropriate evidence answering the relevant clinical questions, or conduct one when there is none available. This should be followed by a transparent evidence summary, such as creation of GRADE evidence profiles, on which to base judgements (18).

For the ARIA guideline 2008 update, this postulate is partially fulfilled. Members of the ARIA guideline panel, including the GRADE representatives, performed extensive literature searches (12–14, 16) addressing the clinical questions covered by the guideline. In addition, we identified up-to-date valid systematic reviews by searching

Quality assessment							Summary of findings					Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	No of patients		Effect		Quality	
							antigen avoidance diet	control	Relative (95% CI)	Absolute		
Atopic eczema in first 12-18 months (follow-up 12-18 months)												
2	randomized trial	serious <sup>8</sup>	no serious inconsistency	no serious indirectness	serious <sup>9</sup>	none	35/157	39/177	RR 1.01 (0.57 to 1.79)	2 more per 1000 <sup>11</sup>	⊕⊕OO LOW	CRITICAL
Asthma in first 18 months (follow-up 18 months)												
2	randomized trial	serious <sup>8</sup>	no serious inconsistency	no serious indirectness	serious <sup>10</sup>	none	4/157	2/177	RR 2.22 (0.39 to 12.67)	13 more per 1000 <sup>11</sup>	⊕⊕OO LOW	CRITICAL
Allergic rhinitis/conjunctivitis in first 18 months (follow-up 18 months)												
1	randomized trial	serious <sup>8</sup>	no serious inconsistency	no serious indirectness	very serious <sup>2</sup>	none	0/81	0/82	RR 0 (0 to 0)	0 fewer per 1000	⊕OOO VERY LOW	CRITICAL
Allergic urticaria in first 18 months (follow-up 18 months)												
1	randomized trial	serious <sup>8</sup>	no serious inconsistency	no serious indirectness	serious <sup>3,10</sup>	none	3/81	3/82	RR 1.01 (0.21 to 4.87)	0 more per 1000	⊕⊕OO LOW	CRITICAL
Any atopic condition in first 18 months (follow-up 18 months)												
1	randomized trial	serious <sup>8</sup>	no serious inconsistency	no serious indirectness	serious <sup>4</sup>	none	15/81	20/82	RR 0.76 (0.42 to 1.38)	59 fewer per 1000 <sup>11</sup>	⊕⊕OO LOW	CRITICAL
preterm birth												
2	randomized trial	serious	no serious inconsistency	no serious indirectness	very serious <sup>5</sup>	none	3/98	0/138	RR 10.06 (0.53 to 192.26)	0 more per 1000	⊕OOO VERY LOW	CRITICAL
birth weight (measured with: grams <sup>11</sup> ; Better indicated by more)												
2	randomized trial	serious <sup>6</sup>	no serious inconsistency	serious <sup>7</sup>	no serious imprecision	none	98	138	-	WMD -83 (-221 to 55)	⊕⊕OO LOW	IMPORTANT

<sup>1</sup> no relative effect  
<sup>2</sup> 1 trial with no events  
<sup>3</sup> one trial with 163 patients and 3 events  
<sup>4</sup> only one trial with 163 patients 45 events  
<sup>5</sup> two small trials, one with no events and another with 3 events in experimental group  
<sup>6</sup> method of randomization not described, no information on concealment, serious uncertainty about adherence to the intention-to-treat principle.  
<sup>7</sup> uncertainty if birth weight is a reliable indicator of malnutrition  
<sup>8</sup> in one trial concealment not adequate, randomization not described, number of randomized women not reported  
<sup>9</sup> Wide confidence intervals with conceivable considerable benefits and harms  
<sup>10</sup> Wide confidence intervals with conceivable considerable benefits and harms and few events in both groups  
<sup>11</sup> Not significant

Figure 1. Example of an evidence profile.

the MEDLINE database, Cochrane Library, and in selected cases also reference lists of the most recent narrative reviews, related systematic reviews, or studies on the topic. We based our judgements on these systematic reviews and, if applicable, on additional randomized trials published afterwards. We developed GRADE evidence profiles (Fig. 1) for most of the clinical questions based on the systematic reviews. These concise evidence profiles allowed panel members to base their judgements on the same and concisely summarized evidence (19).

When there was no recent systematic review available, we did not perform our own systematic reviews, but followed a pragmatic approach and searched MEDLINE, Cochrane Central Register of Controlled Trials (CENTRAL), and Google Scholar for relevant randomized trials.

For many questions randomized trials were not available and we relied on available observational studies that were identified for a prior ARIA guideline edition (4), its subsequent updates (12–14, 16), or additional nonstructured searches for observational studies. We evaluated the original observational studies to inform judgements about the quality of the underlying evidence. In such situations, we based the judgements about the quality of evidence on a crude examination of the most important studies.

### Judgements about the quality of the evidence

The decision to follow or not to follow a certain management approach depends not only on the best estimates of its anticipated desirable and undesirable consequences, but also on the degree of confidence in these estimates. In the GRADE system the quality of evidence reflects the extent to which a guideline panel’s confidence in an estimate of the effect is adequate to support a particular recommendation (7).

The GRADE system classifies the quality of evidence into four categories: high, moderate, low, and very low (Table 1) (7, 8). As in most other systems, in the GRADE system basic study design remains crucial in determining the confidence in the estimates of the effects of an intervention, but certain limitations or merits can change the quality of the evidence. For recommendations regarding therapeutic strategies, randomized trials generally provide higher quality evidence than observational studies. Rigorously designed and executed observational studies, correspondingly, provide higher quality evidence than case series and individual observations. In the GRADE system randomized trials without important limitations constitute high quality evidence, observational studies without important limitations or special merits make up low quality evidence, and unsystematic

Table 1. Key quality criteria and the grades of the quality of evidence

---

The quality of evidence for each important or critical outcome is based on  
The overall study design

Factors that decrease quality of evidence

- Limitations of the studies (including the detailed study design and their execution)
- Inconsistency of the evidence across studies
- Indirectness (uncertainty about generalizability) of the evidence because of differing population, intervention, comparison, and/or outcomes
- Imprecision of the estimate
- High likelihood of reporting bias

Factors that increase quality of evidence

- Strong or very strong associations
- Dose-response effects
- Likelihood that all plausible residual confounding would decrease the observed effect

---

Evidence was classified as 'high', 'moderate', 'low' or 'very low' based on the above criteria and on the following definitions:

High: further research is very unlikely to change the confidence in the estimate of effect.

Moderate: further research is likely to have an important impact on the confidence in the estimate of effect and may change the estimate.

Low: further research is very likely to have an important impact on the confidence in the estimate of effect and is likely to change the estimate.

Very low: any estimate of effect is very uncertain.

clinical observations constitute evidence of a very low quality.

Separation of the strength of a recommendation from the quality of supporting evidence is critical when making recommendations. The GRADE system permits strong recommendations supported by low-quality evidence from observational studies, at the same time allowing weak recommendations based on high-quality evidence. While the former is a rare occurrence in the unusual case where factors determining the strength of a recommendation (see below) are suggesting this as the best course of action, weak recommendations in the face high quality evidence are frequent. For example, if the ARIA panel was uncertain about the actual magnitude of the desirable and undesirable effects of an intervention it was more likely to make a weak recommendation (eg. use of sublingual allergen-specific immunotherapy in children). When there was an apparent large difference between advantages and disadvantages, the panel was appropriately reluctant to offer a strong recommendation in the face of low-quality evidence.

**Decisions about the strength of recommendation**

The strength of a recommendation reflects the extent to which a guideline panel can be confident that desirable effects of an intervention outweigh undesirable effects, or vice versa, across the range of patients for whom the recommendation is intended. Desirable effects can include favourable health outcomes and an improved quality of life, and undesirable effects – harms, burden and costs.

Although the balance between desirable and undesirable effects, and thus the strength of a recommendation, is a continuum, the GRADE system classifies recommendations for or against an intervention into two categories: strong or weak. Categorizing recommendations as 'strong' or 'weak' is inevitably arbitrary. The GRADE Working Group believes that the simplicity and behavioural implications of this explicit grading outweigh the disadvantages. It reflects the way clinicians think or behave: in many situations clinicians are very certain about what to do and apply or recommend a given management option (corresponding to a strong recommendation for or against an action) or they consider a management option in more detail, balancing the pros and cons, before they undertake an action (corresponding to a weak recommendation for or against an action). This simple classification also aids interpretation (7, 20). The GRADE Working Group suggests specific implications of strong or weak recommendations (Table 2) to aid their interpretation.

**Factors that influence the strength of a recommendation**

The balance between desirable and undesirable effects

The balance between the desirable and undesirable effects of the alternative health interventions is based on the best estimates of these effects (Table 3). The larger the difference between the desirable and undesirable effects of the alternative interventions, the more appropriate is a strong recommendation for or against one of the considered interventions. When desirable and undesirable effects are closely balanced, a weak recommendation is more appropriate. If ARIA guideline panel members were confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects, or vice versa, they made a strong recommendation for or against an intervention. When they were not confident, they concluded that a weak recommended is more appropriate.

Quality of evidence

The quality of evidence determines our confidence in the estimates of the effects. Lower quality evidence (i.e. evidence that is at increased risk of leading to biased results) decreases a guideline panel's confidence in the results of research and the estimates of the effects. If panel members are uncertain of the magnitude of the benefits or downsides of alternative interventions, it is unlikely that they make a strong recommendation for or against that intervention. Thus, even when there is an apparent large difference in the balance of advantages and disadvantages, guideline panel members will be appropriately reluctant to make a strong recommendation in the face of a low quality of the evidence. In general the higher the quality of the evidence, the more likely it is for the

Table 2. Implications of strong and weak recommendations

	Strong recommendation	Weak recommendation
Implications		
For patients	Most individuals in this situation would want the recommended course of action and only a small proportion would not. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences	The majority of individuals in this situation would want the suggested course of action, but many would not
For clinicians	Most individuals should receive the intervention. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator	Recognize that different choices will be appropriate for different patients, and that you must help each patient arrive at a management decision consistent with her or his values and preferences. Decision aids may well be useful helping individuals making decisions consistent with their values and preferences
For policy makers	The recommendation can be adapted as policy in most situations	Policy making will require substantial debates and involvement of many stakeholders
Example	We recommend intranasal glucocorticosteroids for treatment of allergic rhinitis in adults and children (strong recommendation based on high quality evidence)	<p><i>Question</i></p> <p>Should antigen avoidance diet be used in pregnant or breastfeeding women to prevent development of allergy in children?</p> <p><i>Recommendation</i></p> <p>For pregnant or breastfeeding women, we suggest no antigen avoidance diet (weak recommendation based on very low quality evidence)</p>

Table 3. Determinants of strength of recommendation

Factors that influence the strength of a recommendation	Comment
Balance between desirable and undesirable effects	A strong recommendation is more likely as the difference between the desirable and undesirable consequences becomes larger. A weak recommendation is more likely as the net benefit becomes smaller and the certainty around that net benefit decreases
Quality of the evidence	A strong recommendation becomes more likely with higher quality of evidence
Values and preferences	A strong recommendation is more likely as the variability of or uncertainty about patient values and preferences decreases. A weak recommendation is more likely as the variability or uncertainty about patient values and preferences increases
Costs (resource allocation)	A weak recommendation is more likely as the incremental costs of an intervention (more resources consumed) increase

recommendation to be strong. Conversely, if the quality is low or very low a weak recommendation is likely to follow. Strong recommendations based on low or very low quality evidence are rare (Table 4).

Variability of or uncertainty about patient values and preferences

There always is a trade-off between advantages and disadvantages of alternative health interventions. Thus, wording and the strength of any recommendation heavily depends on how guideline panel members value related benefits, risks, or inconvenience. Uncertainty or variability of values and preferences may weaken the strength of a recommendation. Although it would be ideal to elicit the estimates of patient values and

preferences associated with a recommended management strategy directly from large scale population studies, such studies are usually not available. When the ARIA guideline panel felt that a judgement about values or preferences was particularly important for the formulation or interpretation of a recommendation, they described the key values they have attributed in making this recommendation. These values, therefore, reflect the values of the panel.

Costs

The panel could have considered cost or resource expenditure as one of the outcomes when weighing up desirable (less resource expenditure) and undesirable (more resource expenditure) consequences of competing interventions for each recommendation. Resource expenditure, however, is much more variable over jurisdictions, time, and availability of the over-the-counter medications (21, 22). In addition, the resource implications also vary widely. Thus, while higher costs will reduce the likelihood of a strong recommendation in favour of a particular intervention, the context of the recommendation can be critical. In considering resource allocation issues, guideline panels must thus be very specific about the setting to which a recommendation applies and the perspective that they had chosen, i.e. which costs were considered and whether resource expenditure was considered from the perspective of the patient (depending on insurance status these cost differ) or society within a given health care system (this includes indirect or opportunity cost saved or incurred by following a recommendation). Furthermore, recommendations that are heavily influenced by costs are likely to change over time as resource

Table 4. Grading the strength of recommendations and quality of evidence in the ARIA guidelines according to the GRADE system [adapted from Schünemann et al. (8)]

Notation*	Strength of recommendation and quality of evidence	Clarity of balance between desirable and undesirable effects	Quality of supporting evidence	Implications
1A	Strong recommendation High-quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research is unlikely to change our confidence in the estimate of effect
1B	Strong recommendation Moderate-quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or unusually strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
1C	Strong recommendation Low-quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence for at least one critical outcome from RCTs with serious flaws, observational studies, or indirect evidence	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
1D	Strong recommendation Very low-quality evidence (very rarely applicable)	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence for at least one of the critical outcomes from unsystematic clinical observation or very indirect evidence	Recommendation may change when higher quality evidence becomes available. Any estimate of the effect for at least one critical outcome is very uncertain
2A	Weak recommendation High-quality evidence	Desirable effects closely balanced with undesirable effects	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies	The best action may differ depending on circumstances or patients' or societal views. Further research is very unlikely to change our confidence in the estimate of the effect
2B	Weak recommendation Moderate-quality evidence	Desirable effects closely balanced with undesirable effects	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or unusually strong evidence from unbiased observational studies	Alternative approach likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence of the estimate of effect and may change the estimate
2C	Weak recommendation Low-quality evidence	Uncertainty in the estimates of desirable and undesirable effects; desirable effects may be closely balanced with undesirable effects	Evidence for at least one critical outcome from RCTs with serious flaws, observational studies, or indirect evidence	Other alternatives may be equally reasonable. Further research is very likely to have important impact on our confidence in the estimate of effect and is likely to change the estimate
2D	Weak recommendation Very low-quality evidence	Major uncertainty in the estimates of desirable and undesirable effects; desirable effects may be closely balanced with undesirable effects	Evidence for at least one critical outcome from unsystematic clinical observation or very indirect evidence	Other alternatives may be equally reasonable. Any estimate of the effect for at least one critical outcome is very uncertain

RCT, randomized controlled trial.

\*Notation is not a part of GRADE system; it was adopted for the ARIA guideline to simplify the presentation.

implications evolve. The ARIA guideline panel considered cost implicitly and not in detail, mostly because ARIA guidelines are intended for users around the world and drug costs are highly variable between

different regions. However, when resource utilization was likely to play a role and influenced a recommendation, this was labelled in the values and preferences section.

### Formulation of recommendations

Guidelines make a set of recommendations advising the clinicians, patients, and other healthcare professionals which of the alternative management strategies is likely to be most beneficial for patients. Consequently, wording of the recommendations in the ARIA guidelines update clearly states what is the proposed course of action. In this document we followed the GRADE working group's advice to use the phrase 'we recommend' for strong recommendations and a less definitive 'we suggest' for weak recommendations. Wording of strong and weak recommendations was particularly important since the ARIA guidelines are intended for patients and clinicians in different regions, cultures, traditions, and usage of language. When appropriate, recommendations are supplemented with an explanatory statement of values and preferences the members of a guideline panel considered in formulating the recommendation and determining its strength. Further remarks occasionally follow, when the panel thought that additional statements are justified (such as dosing), but are not the recommended actions *per se*.

### Panel meeting

Based on the available evidence, the estimates of effect and their gradients, assumed values and preferences and resource utilization issues, members of the ARIA guideline panel made decisions regarding the strength of each recommendation. To achieve this, guideline panel held a 1-day meeting on June 20th 2007 under the auspices of Agenzia Italiana del Farmaco (AIFA) in Rome, Italy, including 6 members of ARIA to discuss the procedures and draft recommendations and discussed a first draft of the guidelines and two members of AIFA (SB and GR). After having agreed on the procedures and additional work that needed to be done, a second meeting was arranged to discuss the final the recommendations based on a second draft of the updated guidelines. During this subsequent meeting on 15 September 2007 in Stockholm, Sweden, members of the ARIA guideline panel reviewed these judgements and made decisions on the quality of evidence, the balance of benefits and downsides (harms, burden and occasionally cost) of a considered management strategy and on the final strength of each recommendation. Recommendations are based on a consensus of the panel. For the formulation and discussion about the final recommendations, all panel members were asked to consider their own and other conflicts during the discussion and decision making as well as to abstain from discussion and voting if necessary (see supplemental information). Subsequent interaction and discussion took place through email but recommendations were not changed after the final meeting except for minor wording changes or correction of factual errors. We will submit

the final draft of the ARIA guidelines to a panel of consultants. Consultants will be specifically asked to consider only errors of fact.

### Limitations of these guideline development methods

Limitations of these guidelines include that we did not consistently and explicitly define all the important outcomes that the panel should have considered for each recommendation, and no formal decision on the relative importance of these outcomes. Secondly, there is a small possibility that we missed studies, because we did not perform full systematic reviews (in particular reviews of observational studies) for all recommendations. Thirdly, we did not search EMBASE and some trials are published in journals not referenced in MEDLINE. As a result, methodological evaluation of the available studies was not always as rigorous as the GRADE approach warrants. The primary reason for not conducting complete systematic reviews for each relevant question was the lack of time. However, due to the extensive knowledge of the ARIA panel of the published literature, it is thought all important studies have been discussed. Fourth, when the systematic reviews were not available, we performed only few statistical pooling exercises of primary study results. Finally, sparse data on how patients value the outcomes and what are patient preferences for recommended interventions is an additional limitation inherent to most clinical practice guidelines.

### Future updates of the ARIA guideline

The ARIA guideline panel decided that future iterations and updates of the guidelines will include better quality evidence summaries. Each recommendation will have a well done systematic review as the basis for decision making. Evidence profiles will be available for every recommendation and detailed prior judgements about the importance of outcomes will be available that include the views of consumers. The ARIA panel has already decided to encourage critical appraisal of the guidelines using established tools such as the AGREE Instrument (23).

### Summary

The 2008 update of the ARIA guidelines has become more evidence-based and followed a structured approach to evaluating the quality of evidence, the balance of desirable and undesirable consequences of patient management, values and preferences and, on occasion, resources. Future iterations of the guidelines will further be improved by following the described processes even closer, such as ensuring availability of updated high quality systematic reviews for each question. To allow

judgements about the quality and transparency guideline panel members have declared potential conflicts of interest in greater detail.

### Potential conflict of interest statement

The following statements follow the template of declaring potential conflict of interests for the World Health Organization guideline panels. J. Brozek is an editor of and receives a salary from a clinical journal that advertises for drugs including those that are the subject of this guideline. J. Brozek received honoraria for speaking at conferences from GlaxoSmithKline and UCBPharma and is a member of the GRADE working group. C. Baena-Cagnani has received fees for consultancy, speaker bureau participation, lectures and research grants from Sanofi-Aventis, Novartis, GSK, Schering Plough, ALK and Abello. S. Bonini has received research grants, consultancies, fees for editorial activity, honoraria from participation in speakers bureau from Alcon, Allergopharma, Allergy Therapeutics, ALK, Almirall, Aventis, AstraZeneca, Bioallergy, Bioxell, GlaxoSmithKline, Lofarma, Merck Sharp & Dohme, Novartis, Pfizer, Sanofi-Synthelabo, Schering-Plough, Stallergens, UCB, Zambon. S. Bonini also declares membership of the AIFA R & D panel. R.G. van Wijk has received fees for lectures and expert panel participation from Allmiral, Alcon, Merck Sharp & Dome, Novartis, Stallergènes and UCB. T. Zuberbier has no conflict of interest. J. Bousquet has received fees and honoraria for lectures, expert panel participation and consultations from Allmiral, AstraZeneca, Centocor, GlaxoSmithKline, Merck Sharp and Dohme, Novartis, Nycomed-Altana, Pfizer, Roche, Stallergènes, Schering Plough, UCB and Uriach. G.W. Canonica has received fees and honoraria for lectures, expert panel participation and consultations and research support from A.Menarini, Alcon, Alk-Abellò,

Almirall, Altana, Anallergo, AstraZeneca, Aventis Pharma, Bayer, Biofutura PharmaBoehringer Ingelheim, Chiesi Farmaceutici, Chiron, Essex, Fujisawa, Genentech, Gentili, Glaxo Smith Kline, Glaxo Smith Kline, Lofarma, Merck Sharp & Dome, Novartis, Pfizer, Pharmacia & Upjohn, Schering Plough, SigmaTau, Stallergenes, Yamanuchi, UCB Pharma and Valeas. G.R. Guido is a member of the AIFA Executive Board. G. Guyatt and H.J. Schünemann are members of the GRADE working group and UpToDate® EBM consultant; honoraria related to this work are deposited into research accounts. H.J. Schünemann has not personal payments from for-profit sponsors, but he received research grants and honoraria that were deposited into research accounts or received by a research group that he belongs to from AstraZeneca, Amgen, Chiesi Foundation, Lilly, and Pfizer, Roche and UnitedBioSource for development or consulting regarding quality of life instruments for chronic respiratory diseases and as lecture fees related to the methodology of evidence based practice guideline development and research methodology. Institutions or organizations that H.J. Schünemann is affiliated with likely receive funding from for-profit sponsors that are supporting infrastructure and research that may serve his work.

### Acknowledgments

H.J. Schünemann is funded by a 'The human factor, mobility and Marie Curie Actions Scientist Reintegration' European Commission Grant: IGR 42194 – 'GRADE'. Members of the GA<sup>2</sup>LEN (Global Allergy and Asthma European Network) are supported by EU Framework programme for research, contract n ° FOOD-CT-2004-506378. We thank the Agenzia Italiana del Farmaco (AIFA) in Rome for hosting one of the meetings of the ARIA group. The authors wish to thank Enrico Compalati (University of Genoa), Paola Aita (AIFA) and Carlo Tomino (AIFA) for participating in and supporting that meeting.

### References

- 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.
- COMPUS. Proposed evaluation tools for COMPUS. 2005. [http://www.cadth.ca/media/pdf/COMPUS\\_Evaluation\\_Methodology\\_draft\\_e.pdf](http://www.cadth.ca/media/pdf/COMPUS_Evaluation_Methodology_draft_e.pdf); accessed 15 June 2007.
- Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schunemann H. An emerging consensus on grading recommendations? *ACP J Club* 2006;**144**: A8–A9.
- Bousquet J, Van Cauwenberge P, Khaltaev N. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol* 2001;**108**(5 Suppl.): S147–S334.
- Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. *BMJ* 1999;**318**:593–596.
- Guyatt GH, Hayward R, Richardson WS, Green L, Wilson M, Sinclair J et al. Moving from evidence to action. In: Guyatt GH, Rennie D, editors. *Users' guides to the medical literature: a manual for evidencebased practice*. Chicago, IL: AMA Press, 2002: 175–199.
- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490.

8. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A et al. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006;**174**:605–614.
9. Global Programme on Evidence for Health Policy. Guidelines for WHO guidelines. EIP/GPE/EQC/2003.1. Geneva: World Health Organization, 2003.
10. Boyd EA, Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. *Health Res Policy Sys* 2006;**4**:16.
11. World-Health-Organisation. [http://whqlibdoc.who.int/hq/2003/EIP\\_GPE\\_EQC\\_2003\\_1.pdf](http://whqlibdoc.who.int/hq/2003/EIP_GPE_EQC_2003_1.pdf); accessed 1 May 2007.
12. Custovic A, Wijk RG. The effectiveness of measures to change the indoor environment in the treatment of allergic rhinitis and asthma: ARIA update (in collaboration with GA(2)LEN). *Allergy* 2005;**60**:1112–1115.
13. Passalacqua G, Bousquet PJ, Carlsen KH, Kemp J, Lockey RF, Niggemann B et al. ARIA update: I—Systematic review of complementary and alternative medicine for rhinitis and asthma. *J Allergy Clin Immunol* 2006;**117**:1054–1062.
14. Passalacqua G, Durham SR. Allergic rhinitis and its impact on asthma update: allergen immunotherapy. *J Allergy Clin Immunol* 2007;**119**:881–891.
15. Bonini S, Bonini M, Bousquet J, Brusasco V, Canonica GW, Carlsen KH et al. Rhinitis and asthma in athletes: an ARIA document in collaboration with GA2LEN. *Allergy* 2006;**61**:681–692.
16. Bousquet J, van Cauwenberge P, Ait Khaled N, Bachert C, Baena-Cagnani CE, Bouchard J et al. Pharmacologic and anti-IgE treatment of allergic rhinitis ARIA update (in collaboration with GALEN). *Allergy* 2006;**61**:1086–1096.
17. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988;**138**:697–703.
18. Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG et al. WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. *Lancet Infect Dis* 2007;**7**:21–31.
19. Schünemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R, Stockman L et al. Transparent development of the WHO rapid advice guidelines. *PLoS Medicine* 2007;**4**:e119.
20. Guyatt GH, Cook DJ, Jaeschke R, Schünemann HJ, Pauker S. Grading recommendations: a qualitative approach. In: Guyatt GH, Rennie D, editors. *Users' guides to the medical literature: a manual for evidencebased practice*. Chicago, IL: AMA Press, 2002:599–608.
21. Sullivan PW, Nair KV, Patel BV. The effect of the Rx-to-OTC switch of loratadine and changes in prescription drug benefits on utilization and cost of therapy. *Am J Manag Care* 2005;**11**:374–382.
22. Brass EP. Implications of a switch from prescription to over-the-counter status for allergy drugs. *Curr Allergy Asthma Rep* 2004;**4**:245–250.
23. AGREE-Collaboration. Appraisal of guidelines for research & evaluation (AGREE) instrument. <http://www.agreecollaboration.org>; accessed 2 September 2007.