Advice on global guideline development by SURE to the WHO

Holger Schünemann
State University of New York at Buffalo
Italian National Cancer Institute „Regina Elena“
CLARITY Group, McMaster University, Canada

KDIGO
New York, NY October 12, 2007

Disclosure

Relevant Financial Relationships

- Member of the GRADE working group: honoraria related to this work and for giving lectures on research methodology deposited into research accounts
  - UpToDate®, Pfizer, Lilly, Chiesi, AstraZeneca
- Institutions or organizations that I am affiliated with likely receive funding from for-profit sponsors that are supporting infrastructure and research that may serve his work

Off label medication

- None mentioned

Content

- Background and rationale for this work
- Methods of the SURE review
- Main findings
- Recommendations to WHO
- What is happening at WHO and other organizations

Background

- WHO develops advice (recommendations/ guidelines) “all the time”
- Format differs, methods differ, much criticism
- Increasingly governments, professional and consumer organizations are demanding more rigorous processes to ensure that health decisions are well informed by the best available research evidence.
- May 2005 World Health Assembly resolution
  - WHO Director-General "to undertake an assessment of WHO's internal resources, expertise and activities in the area of health research, with a view to developing a position paper on WHO's role and responsibilities in the area of health research, and to report through the Executive Board to the next World Health Assembly."

Background

- Advisory Committee on Health Research (ACHR)
- Advice on how WHO can improve the use of research evidence in the development of recommendations, guidelines and policies.
- Subcommittee for the Use or Research Evidence (SURE) to do the work

Methods

- Prepare series of reviews for WHO to develop guideline handbook
- 3 member secretariat (ADO, AF, HJS)
- Vetting of most important topics and questions that should be addressed among authors and ACHR SURE
- Performed semi-systematic reviews of existing literature and databases
  - What are others doing?
  - What is WHO doing?
  - What should WHO do?
- Presentation of results to ACHR and WHO DG

Key topics

1. Guidelines for guidelines
2. Priority setting
3. Group composition and consultation process
4. Managing conflicts of interest
5. Group processes
6. Determining which outcomes are important
7. Deciding what evidence to include
8. Synthesis and presentation of evidence
9. Grading evidence and recommendations
10. Integrating values and consumer involvement
11. Incorporating considerations of cost-effectiveness, affordability and resource implications
12. Incorporating considerations of equity
13. Adaptation, applicability and transferability
14. Reporting guidelines
15. Disseminating and implementing guidelines
16. Evaluation

Key topics

1. Guidelines for guidelines
2. Priority setting
3. Group composition and consultation process
4. Managing conflicts of interest
5. Group processes
6. Determining which outcomes are important
7. Deciding what evidence to include
8. Synthesis and presentation of evidence
9. Grading evidence and recommendations
10. Integrating values and consumer involvement
11. Incorporating considerations of cost-effectiveness, affordability and resource implications
12. Incorporating considerations of equity
13. Adaptation, applicability and transferability
14. Reporting guidelines
15. Disseminating and implementing guidelines
16. Evaluation

Guidelines for guidelines

- Lack of standardized guideline development leads to widely varying recommendations
- Identified 19 components that guidelines for guidelines should cover, but:
  - to make a guideline for guideline credible and acceptable, individuals with expertise in methodology, process and implementation of guidelines should be involved in developing a guideline for guideline document
  - Clinicians?

Guidelines for guidelines II

- Standardize methods beyond organization
- Obtain approval – e.g. board
- Publish
- Training/software
- Living document
  - Monitor methodological literature
  - Examples

Group composition

- One systematic review (Murphy et al. 1998)
- Composition of panel influences recommendations
  - Members of a specialty are more likely to advocate techniques that involve their specialty
- Balanced groups
  - Select the appropriate group leader
- Necessary technical skills
  - including information retrieval, systematic reviewing, health economics, group facilitation, project management, writing and editing
- Include or have access to content experts
- No SR on how to obtain consultation, but logical reasons support this

Managing COI (Boyd and Bero)

- No SR, no RCT comparing methods to obtain COI
- No agreement on amount, period of recall, type (own, family)
- Possible management:
  - disclosure of the financial tie(s) in publications and presentations (primarily used strategy)
  - reducing equity holdings
  - altering consulting agreements to
  - eliminating the financial tie; appointing oversight committees
  - recusal
- But does this help?
  - No empirical evidence for COI policy enforcement

Which outcomes?

- What methods, what type, what ranking?
- Little evidence!
  - Systematic methods of question formulation improve search for evidence
- Questions/panel should identify outcomes a priori
- Ranking outcomes by their relative importance, separated into benefits and downsides can help to focus attention and clarify disagreements.
- Research on values and preferences should guide the ranking
- If varies across cultures ranking by people in a specific setting
- If evidence is lacking -> acknowledge

Case scenario

A 13 year old girl who lives in rural Indonesia presented with flu symptoms and developed severe respiratory distress over the course of the last 2 days. She required intubation. The history reveals that she shares her living quarters with her parents and her three siblings. At night the family's chicken stock shares this room too and several chicken had died unexpectedly a few days before the girl fell sick.

Relevant clinical question?

Clinical question:

Population: Avian Flu/influenza A (H5N1) patients

Intervention: Oseltamivir (or Zanamivir)

Comparison: No pharmacological intervention

Outcomes: Mortality, hospitalizations, resource use, adverse outcomes, antimicrobial resistance

Schunemann et al., PLOS Med, 2007

Methods – WHO Rapid Advice Guidelines for management of Avian Flu

- Applied findings the work in front of you
- Group composition (including panel of 13 voting members):
  - clinicians who treated influenza A(H5N1) patients
  - infectious disease experts
  - basic scientists
  - public health officers
  - methodologists
- Independent scientific reviewers:
  - Identified systematic reviews, recent RCTs, case series, animal studies related to H5N1 infection

Evidence Profile

Oseltamivir for treatment of H5N1 infection:

Quality assessment						Summary of findings					
						No of patients		Effect			
No of studies (Ref)	Design	Limitations	Consistency	Directness	Other considerations	Oseltamivir	Placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Healthy adults:											
Mortality											
0	-	-	-	-	-	-	-	-	-		9
Hospitalisation (Hospitalisations from influenza – influenza cases only)											
5 (TJ 06)	Randomised trial	No limitations	-	Major uncertainty (-2)[1]	Imprecise or sparse data (-1)	-	-	OR 0.22 (0.02 to 2.16)	-	⊕○○○ Very low	6
Duration of hospitalization											
0	-	-	-	-	-	-	-	-	-	-	7
LRTI (Pneumonia - influenza cases only)											
5 (TJ 06)	Randomised trial	No limitations	-	Major uncertainty (-2)[1]	Imprecise or sparse data (-1)[2]	2/982 (0.2%)	9/662 (1.4%)	RR 0.149 (0.03 to 0.69)	-	⊕○○○ Very low	8
Duration of disease (Time to alleviation of symptoms/median time to resolution of symptoms – influenza cases only)											
5[3] (TJ 06) (DT 03)	Randomised trials	No limitations[4]	Important inconsistency (-1)[5]	Major uncertainty (-2)[1]	-	-	-	HR 1.30[3] (1.13 to 1.50)	-	⊕○○○ Very low	5
Viral shedding (Mean nasal titre of excreted virus at 24h)											
2[6] (TJ 06)	Randomised trials	No limitations	-[7]	Major uncertainty (-2)[1]	None	-	-	-	WMD -0.73[8] (-0.99 to -0.47)	⊕⊕○○ Low	4
Outbreak control											
0	-	-	-	-	-	-	-	-	-	-	4
Resistance											
0	-	-	-	-	-	-	-	-	-	-	7
Serious adverse effects (Mention of significant or serious adverse effects)											
0[9]	-	-	-	-	-	-	-	-	-	-	7
Minor adverse effects [10] (number and seriousness of adverse effects)											
3[11] (TJ 06)	Randomised trials	No limitations	-[12]	Some uncertainty (-1)[13]	Imprecise or sparse data (-1)[14]	-	-	OR range[15] (0.56 to 1.80)	-	⊕⊕○○ Low	
Cost of drugs											
0	-	-	-	-	-	-	-	-	-	-	4

Oseltamivir for Avian Flu

Summary of findings:

- No clinical trial of oseltamivir for treatment of H5N1 patients.
- 4 systematic reviews and health technology assessments (HTA) reporting on 5 studies of oseltamivir in seasonal influenza.
  - Hospitalization: OR 0.22 (0.02 – 2.16)
  - Pneumonia: OR 0.15 (0.03 - 0.69)
- 3 published case series.
- Many in vitro and animal studies.
- No alternative that is more promising at present.
- Cost: ~ Euro 50 per treatment course

Example: Oseltamivir for Avian Flu

Recommendation: In patients with confirmed or strongly suspected infection with avian influenza A (H5N1) virus, clinicians should administer oseltamivir treatment as soon as possible (????? recommendation, very low quality evidence).

Schunemann et al., The Lancet ID, 2007

Example: Oseltamivir for Avian Flu

Recommendation: In patients with confirmed or strongly suspected infection with avian influenza A (H5N1) virus, clinicians should administer oseltamivir treatment as soon as possible (strong recommendation, very low quality evidence).

Values and Preferences

Remarks: This recommendation places a high value on the prevention of death in an illness with a high case fatality. It places relatively low values on adverse reactions, the development of resistance and costs of treatment.

Schunemann et al., The Lancet ID, 2007

Other explanations

Remarks: Despite the lack of controlled treatment data for H5N1, this is a strong recommendation, in part, because there is a lack of known effective alternative pharmacological interventions at this time.

The panel voted on whether this recommendation should be strong or weak and there was one abstention and one dissenting vote.

Schunemann et al., The Lancet ID, 2007

Deciding what evidence to include?

- Globalize the Evidence (localize the decisions) – J. Eisenberg
- Begin by searching for high quality systematic reviews
- Different questions and outcomes, different study designs
- When high quality available, do not search for other data

Adverse event data

Table 1: Pros and cons of different approaches for incorporating adverse effect data in a systematic review*

Method	Look in the trials/studies included in the systematic review of benefit.	Look in all retrieved trials/studies of that intervention, even in those excluded from the analysis of benefit	Look for studies that specifically evaluate adverse effects of the intervention
Protocol	Should usually be the minimum recommendation	Studies rejected from analysis of benefit (e.g. because beneficial outcomes are measured in a different way, which cannot be combined with other studies), may be included to allow adverse effect data collection. Two sets of inclusion criteria will be needed – for benefit, and for adverse effects	Design separate strategy to identify studies that report adverse effects, including those that do not look at beneficial effects. Might amount to a separate review nested within a traditional Cochrane review
Pros	Less demanding on time and resources	More comprehensive than just looking at included trials	Most comprehensive
	Does not require new literature search strategy	Can potentially cover a more representative group of patients	May be able to evaluate rare, or long-term, or previously unrecognized adverse effects
Cons	Data may be very limited and biased towards common, short-term harms	Relatively time consuming as full-text articles of all potentially relevant studies need checking Data may be limited to well-recognized and commonly seen adverse effects.	Time and resource intensive
		Benefit and harm cannot be compared directly as the data come from different sources	Special techniques required in synthesizing data from a diverse range of sources Increased quantity of data but greater risk of biased and poor quality data Benefit and harm cannot be compared directly as the data come from different sources.

Loke et al, Cochrane Handbook, 2005

Grading evidence

The GRADE approach

Clear separation of 2 issues:

- 4 categories of quality of evidence: very low, low, moderate, or high quality?
  - methodological quality of evidence
  - likelihood of bias
- Recommendation: weak or strong (for or against)?
  - Quality of evidence only one factor

*www.GradeWorking-Group.org

Synthesis and Presentation

- Use existing reviews
  - Check quality (e.g. AMSTAR)
- Concise summaries
  - Recommendations
  - Summary of Findings (SOF) tables
  - Evidence profiles
  - Text
- Make systematic review available

Summary of Findings table

Summary of findings:

Compression stockings compared with no compression stockings for people taking long flights

Patients or population: Anyone taking a long flight (lasting more than 6 hours)
Settings: International air travel
Intervention: Compression stockings[1]
Comparison: Without stockings

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk Without stockings	Corresponding risk With stockings				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies.
Symptom-less deep vein thrombosis	**Low risk population** [2]		**RR 0.10** (0.04 to 0.25)	2637 (9 studies)	⊕⊕⊕⊕ **High**	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population [2]					
	30 per 1000	**3 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**RR 0.45** (0.18 to 1.13)	1804 (8 studies)	⊕⊕⊕◯ **Moderate**[3]	
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6 to 9**.	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI −4.5 to −4.9).		1246 (6 studies)	⊕⊕◯◯ **Low**[4]	
Pulmonary embolus	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed pulmonary embolus in these studies.
Death	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants died in these studies.
Adverse effects	See comment	See comment	Not estimable	1182 (4 studies)	See comment	The tolerability of the stockings was described as very good with no complaints of side effects in 4 studies.[5]

*The basis for the **assumed risk** is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the intervention group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see explanations)

Summary of findings:

Compression stockings compared with no compression stockings for people taking long flights

Patients or population: Anyone taking a long flight (lasting more than 6 hours)
Settings: International air travel
Intervention: Compression stockings[1]
Comparison: Without stockings

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk **Without stockings**	Corresponding risk **With stockings**				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies.
Symptom-less deep vein thrombosis	**Low risk population** [2]		**RR 0.10** (0.04 to 0.25)	2637 (9 studies)	⊕⊕⊕⊕ **High**	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population [2]					
	30 per 1000	**3 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**RR 0.45** (0.18 to 1.13)	1804 (8 studies)	⊕⊕⊕◯ **Moderate**[3]	
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6 to 9**.	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI −4.5 to −4.9).		1246 (6 studies)	⊕⊕◯◯ **Low**[4]	

Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6 to 9**.	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI −4.5 to −4.9).		1246 (6 studies)	⊕⊕◯◯ **Low**[4]	
Pulmonary embolus	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed pulmonary embolus in these studies.
Death	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants died in these studies.
Adverse effects	See comment	See comment	Not estimable	1182 (4 studies)	See comment	The tolerability of the stockings was described as very good with no complaints of side effects in 4 studies.[5]

*The basis for the **assumed risk** is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the intervention group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see explanations)

How to develop recommendations?

1. What is the "strength of a recommendation" and what determines the strength?
2. What are the implications of strong and weak recommendations for patients, clinicians and policy makers?
3. Should guideline panels make recommendations in the face of very low quality evidence and can these recommendations be used for performance measures?
4. How should recommendations be formulated and presented?

Strength of recommendation

- “The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects.”

Desirable and undesirable effects

- Desirable effects
  - Mortality
  - improvement in quality of life, fewer exacerbations
  - reduction in the burden of treatment
  - reduced resource expenditure
- Undesirable effects
  - deleterious impact on morbidity, mortality or quality of life, increased resource expenditure

Factors determining strength of recommendation

Factors that can strengthen a recommendation	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely is it that a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely is a weak recommendation warranted.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely is a weak recommendation warranted.
Costs (resource allocation)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted.

How to develop recommendations?

1. What is the “strength of a recommendation” and what determines the strength?
2. What are the implications of strong and weak recommendations for patients, clinicians and policy makers?
3. Should guideline panels make recommendations in the face of very low quality evidence and can these recommendations be used for performance measures?
4. How should recommendations be formulated and presented?

Implications of a *strong* recommendation

- Patients: Most people in this situation would want the recommended course of action and only a small proportion would not
- Clinicians: Most patients should receive the recommended course of action
- Policy makers: The recommendation can be adapted as a policy in most situations

Implications of a *weak* recommendation

- Patients: The majority of people in this situation would want the recommended course of action, but many would not

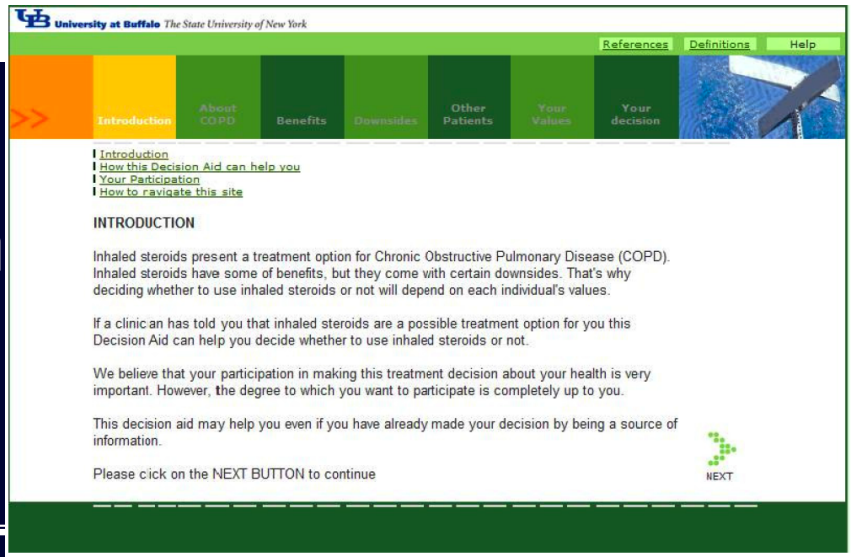
Research article

Open Access

A Decision Aid for COPD patients considering inhaled steroid therapy: development and before and after pilot testing

Elie A Akl*[1,2], Brydon JB Grant[1,2,3,4,5], Gordon H Guyatt[6,7], Victor M Montori[8] and Holger J Schünemann[9]

- Policy makers: There is a substantial debate and in stakeholders



How to develop recommendations?

1. What is the "strength of a recommendation" and what determines the strength?
2. What are the implications of strong and weak recommendations for patients, clinicians and policy makers?
3. Should guideline panels make recommendations in the face of very low quality evidence and can these recommendations be used for performance measures?
4. How should recommendations be formulated and presented?

3. Should guideline panels make recommendations in the face of very low quality evidence/performance indicators?

- would fail one of their fundamental missions: to provide guidance and solutions for the clinician who requires answers to pertinent clinical questions
- panels are in the best position to to make specific and unambiguous recommendations
- higher quality evidence may never be obtained
- physicians need guidance regardless of the quality of the underlying evidence
- one may disagree with this conclusion (in view of believe by some clinicians that all recommendations require immediate implementation)

Clinicians and patients want to know!

- 1) UpToDate® Users
- 2) Mini Medical School attendees*:
  - Participants preferred to know about the uncertainty relating to outcomes of a treatment or a test
  - more interested in knowing about uncertainty relating to benefits than harms (96% vs. 90%; $P<0.001$).
  - strong preference to be informed about the quality of evidence that supports a recommendation.

*Akl et al. J Clin Epi, 2007, in press

How to develop recommendations?

1. What is the “strength of a recommendation” and what determines the strength?
2. What are the implications of strong and weak recommendations for patients, clinicians and policy makers?
3. Should guideline panels make recommendations in the face of very low quality evidence and can these recommendations be used for performance measures?
4. How should recommendations be formulated and presented?

- Few written standards exist
- For strong recommendations, the GRADE working group has suggested adopting terminology such as, “We recommend…” or “Clinicians should…”.
- For weak recommendation, they should use less definitive wording, “We suggest…” or “Clinicians might…”.

Conclusions

- Standardized process: guidelines for guidelines
- Globalize the evidence
- Strong partnerships (systematic reviewers!)