

Use of systematic reviews and the GRADE approach in WHO policy and guideline development

Karin Weyer

Coordinator, TB Diagnostics & Laboratory Strengthening
Stop TB Department
World Health Organization, Geneva, Switzerland

Holger Schünemann

Chair, Department of Clinical Epidemiology & Biostatistics
Michael Gent Chair in Healthcare Research
McMaster University, Hamilton, Canada

Post-graduate Course: Introduction to systematic reviews and meta-analyses

Berlin, 12 November 2010



Content

- WHO guidelines and GRADE
- Background about GRADE
- Grading quality of evidence
- Going from evidence to recommendations
- Example of recent WHO policy formulation



What is a WHO guideline?

- "Guidelines are recommendations intended to assist providers and recipients of health care and other stakeholders to make informed decisions. Recommendations may relate to clinical interventions, public health activities, or government policies."

WHO 2003, 2007



World Health Organization

3

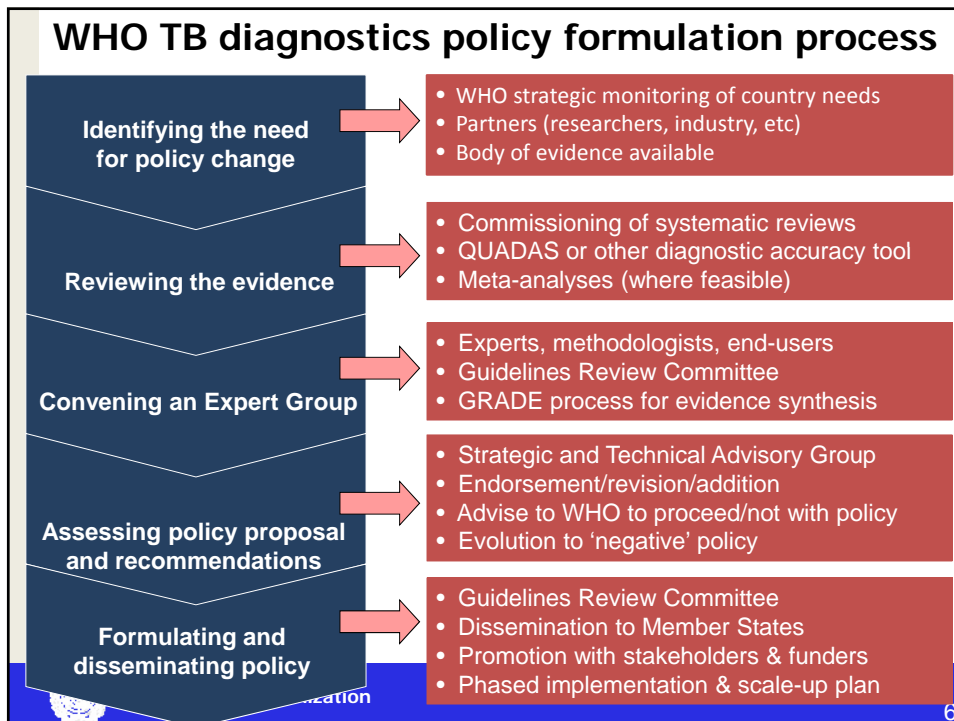
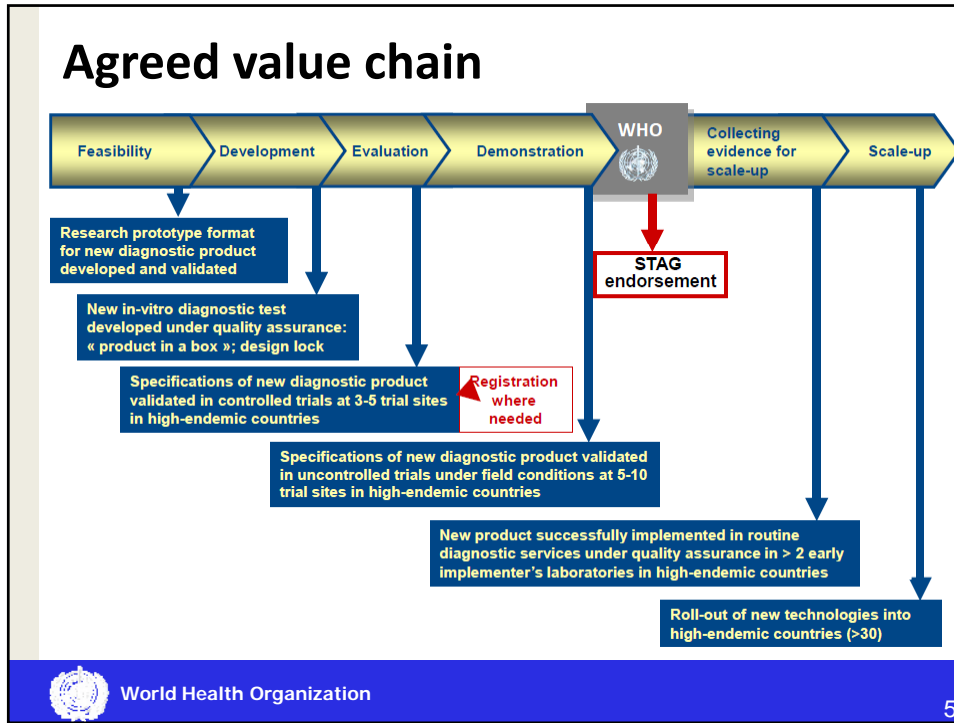
New frontiers

- **Tools development** : >20 new technologies in various stages of development and evaluation
- **Accelerated WHO policy formulation**, with evolution to 'negative' policy guidance
- **Expanded access** to new diagnostics and laboratory strengthening (eg. GLI, EXPAND-TB)
- Require **adaptation to local context and rapid uptake at country level** to maximise patient benefit



World Health Organization

4



Working with evidence

- For key recommendations:
 - Search for and retrieve all available evidence
 - Identify relevant systematic reviews
 - Formally assess quality of evidence
 - GRADE (systematic and transparent approach)



World Health Organization

7

GRADE Working Group

Grades of Recommendation Assessment, Development and Evaluation

- Aim: to develop a common, transparent and sensible system for grading the quality of evidence and the strength of recommendations
- International group of guideline developers, methodologists & clinicians from around the world (>100 contributors) – since 2000
- International group: ACCP, AHRQ, Australian NMRC, BMJ Clinical Evidence, CC, CDC, McMaster, NICE, Oxford CEBM, SIGN, UpToDate, USPSTF, WHO



World Health Organization

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005,
AJRCCM 2006, Chest 2006, BMJ 2006

GRADE Uptake

- World Health Organization
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Physicians
- European Respiratory Society
- European Society of Thoracic Surgeons
- British Medical Journal
- Infectious Disease Society of America
- American College of Chest Physicians
- UpToDate®
- National Institutes of Health and Clinical Excellence (NICE)
- Scottish Intercollegiate Guideline Network (SIGN)
- Cochrane Collaboration
- Infectious Disease Society of America
- Clinical Evidence
- Agency for Health Care Research and Quality (AHRQ)
- Partner of GIN
- Over 40 major organizations



World Health Organization

9

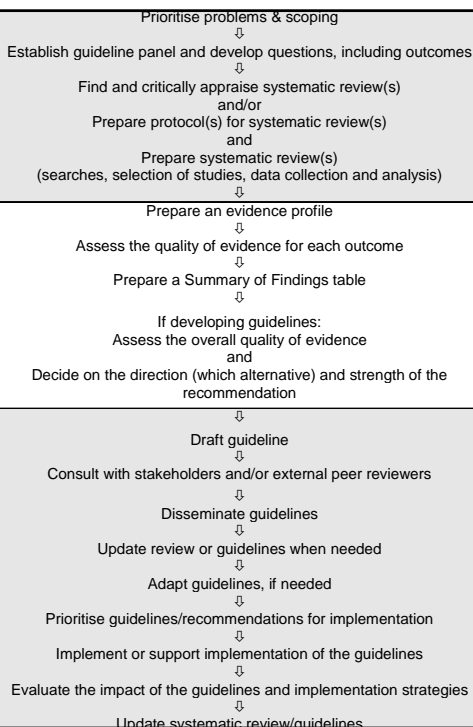
Guideline Development Process

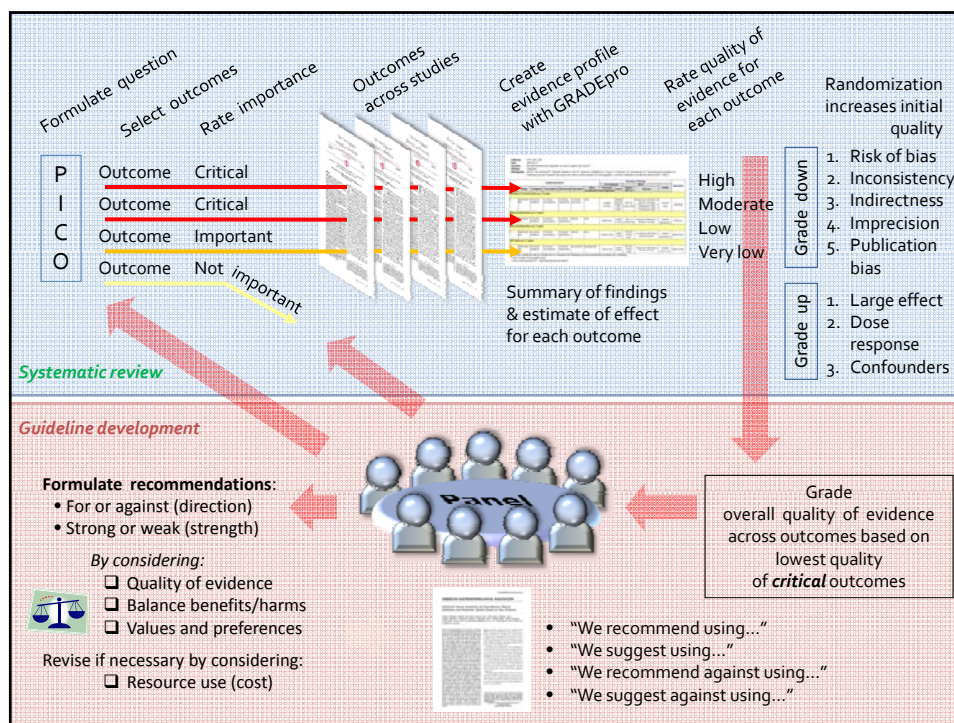
Health Research Policy

Review
Improving the use of research introduction
 Andrew D Oxman*1, Atle Fretheim

Review
Improving the use of research I. Guidelines for guidelines
 Holger J Schünemann*1, Atle Fretheim

Published: 21 November 2006
 Health Research Policy and Systems 2006, 4:13 doi:10.1186/1478-4491-4-13
 This article is available from: <http://www.health-policy-systems.com>





Why bother about grading?

- People always draw conclusions about:
 - Quality of evidence
 - Strength of recommendation
- Systematic and explicit approaches can help:
 - Protect against errors
 - Resolve disagreements
 - Facilitate critical appraisal
 - Communicate information



Getting from evidence to recommendations - GRADE

- Recommendations are judgments:
 - Quality of evidence
 - Trade off between benefits and harms
 - Values and preferences
 - Resource use



Types of questions

Background Questions

Definition: *What is LTBI?*

Mechanism: *How does IGRA work?*

Foreground Questions

Benefit > harm: *In patients suspected of LTBI, does use of test A versus B improve survival, ...?*



Framing a foreground question

Population: Individuals with LTBI

Intervention: Test A

Comparison: No test/test B

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



Choosing outcomes

- Desirable outcomes
 - lower mortality
 - reduced hospital stay
 - reduced duration of disease
 - reduced resource expenditure
- Undesirable outcomes
 - adverse reactions
 - the development of resistance
 - costs of treatment
- Every decision comes with desirable and undesirable consequences
 - Developing recommendations must include a consideration of desirable and undesirable outcomes



Relative importance of outcomes

- Decision makers (and guideline authors) need to consider the relative importance of outcomes when balancing these outcomes to make a recommendation
- Relative importance vary across populations
- Relative importance may vary across patient groups within the same population
- When considered critical - evaluate



GRADE: recommendation – quality of evidence

Clear separation:

- 1) Recommendation: 2 grades – weak/conditional/optional or strong (for or against an intervention)
 - Balance of benefits and downsides, values and preferences, resource use and quality of evidence

- 2) 4 categories of quality of evidence: ⊕⊕⊕⊕ (High), ⊕⊕⊕○ (Moderate), ⊕⊕○○ (Low), ⊕○○○ (Very low)
 - methodological quality of evidence
 - likelihood of bias
 - by outcome and across outcomes



GRADE Quality of Evidence

In the context of making recommendations:

- The quality of evidence reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.



Likelihood
of and
confidence
in an
outcome



Definition of grades of evidence

- ⊕⊕⊕⊕/A/High: Further research is **very unlikely to change confidence** in the estimate of effect.
- ⊕⊕⊕○/B/Moderate: Further research is **likely** to have an important impact on confidence in the estimate of effect and **may change** the estimate.
- ⊕⊕○○/C/Low: Further research is **very likely** to have an important impact on confidence in the estimate of effect and is **likely to change** the estimate.
- ⊕○○○/D/Very low: Any estimate of effect is **very uncertain**.



Determinants of quality for diagnostic questions

- RCTs and observational studies: direct ⊕⊕⊕⊕
- 5 factors that can lower quality
 1. limitations in design and execution (*risk of bias criteria*)
 2. Inconsistency (*or heterogeneity*)
 3. Indirectness (*PICO and applicability*)
 4. Imprecision (*number of events and confidence intervals*)
 5. Publication bias
- 3 factors can increase quality
 1. Large magnitude of effect
 2. All plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
 3. Dose-response gradient



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.”

- Strong or weak/conditional



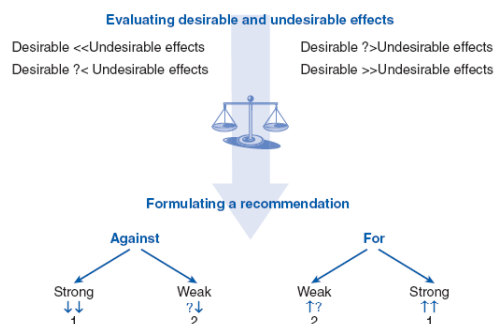
Factors determining strength of recommendation

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 a strong recommendation warranted. The smaller the net benefit and the lower the certainty for that benefit, the more likely is a weak recommendation.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely 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



Developing recommendations

Strength of Recommendations



The figure describes the balance between important benefits and downsides relate to a recommendation. The process begins by evaluating whether desirable effects outweigh undesirable effects or vice versa. Moving on to making a recommendation requires a decision: if the balance is clear, a strong recommendation for or against an action follows (<< and >> denote a clear balance). If the balance is not clear, a weak recommendation for or against an action follows (?< and ?> denote a balance that is not clear). Widely differing values (the importance or preference patients assign to a certain health state) can also lead to a less clear balance of benefits versus downsides.



True positives (TP):

True negatives (TN):

False positives (FP):

False negatives (FN):

Inconclusive results:

Complications of a test:

Resource utilization (cost):



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 *conditional/weak* recommendation

- Patients: The majority of people in this situation would want the recommended course of action, but many would not
- Clinicians: Be more prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision making
- Policy makers: There is a need for substantial debate and involvement of stakeholders



Lessons learnt

- Evolution in systematic review methods allow refinement of evidence-synthesis and GRADE
- If aimed at WHO policy, systematic review scope, protocols and methods need to be agreed beforehand by Guideline Group
- Systematic reviewers need to present findings objectively, without bias, and preferably refrain from making policy inferences
- WHO conflict of interest policy also apply to systematic reviewers ('observer status' at Expert Group and STAG-TB meetings)



Summary

- Clinical practice guidelines should be based on the **best available** evidence to be evidence based
- GRADE combines what is known in health research methodology and provides a structured approach to improve communication
- Criteria for evidence assessment across questions and outcomes
- Criteria for moving from evidence to recommendations
- Transparent, systematic
 - four categories of quality of evidence
 - two grades for strength of recommendations
- Transparency in decision making and judgments is key

