Systematic reviews and other strategies to protect against bias in guidelines development

Holger Schünemann, MD, PhD

Professor of Clinical Epidemiology & Biostatistics
Associate Professor of Medicine
American Thoracic Society Documents Editor
State University of New York
Italian National Cancer Institute, Rome, Italy

IOM, May 22, 2008
Disclosure

Relevant Financial Relationships

- Member of the GRADE working group: honoraria from for-profit sponsors related to this work and for giving lectures on research methodology deposited into research accounts.

- Institutions or organizations that I am affiliated with likely receive funding from for-profit sponsors that are supporting infrastructure and research that may serve my work.
Content

- Why systematic reviews to protect against bias?
- Explicit evaluation of quality of evidence – transparent judgments
- Transparent development of recommendations
A prominent COPD guideline

7.6. Mucolytic/antioxidant therapy

These include drugs such as:

- ambroxol
- erdosteine
- carbocysteine
- iodinated glycerol

The regular use of these drugs has been evaluated in a number of studies with little evidence of any effect on lung function.

Data from a Cochrane review of the studies supports a role for these drugs in reducing the number of exacerbations of chronic bronchitis [33].

There is better evidence that N-acetylcysteine, a drug with mucolytic and anti-oxidant actions, can reduce the number of exacerbations of COPD and this is currently under study in a large prospective trial [34].
Another prominent COPD guideline

Mucolytic (mucokinetic, mucoregulator) agents (ambroxol, erdosteine, carbocysteine, iodinated glycerol). The regular use of mucolytics in COPD has been evaluated in a number of long-term studies with controversial results\(^{150-152}\). Although a few patients with viscous sputum may benefit from mucolytics\(^{153, 154}\), the overall benefits seem to be very small, and the widespread use of these agents cannot be recommended at present (Evidence D).

Antioxidant agents. Antioxidants, in particular N-acetylcysteine, have been reported in small studies to reduce the frequency of exacerbations, leading to speculation that these medications could have a role in the treatment of patients with recurrent exacerbations\(^{155-158}\) (Evidence B). However, a large randomized controlled trial found no effect of N-acetylcysteine on the frequency of exacerbations, except in patients not treated with inhaled glucocorticosteroids\(^{159}\).
What to do?
Reasons for standardized methods

- Non-rigorous guidelines:
  - Create noise & bias
  - Make more aggressive recommendations
  - Can impair credibility of professional societies
- Link between evidence and recommendations until now not transparent
- Systematic approach to identify and select evidence will provide transparency and opportunity for control when the quality is doubtful

**Health Research Policy and Systems**

This article is available from: http://www.health-policy-systems.com/content/4/1/12

Review

**Improving the use of research evidence in guideline development: introduction**

Andrew D Oxman*, Atle Fretheim¹, Holger J Schünemann² and SURE³
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
What makes Guidelines Evidence-Based in 2008?

Critical appraisal: Appraisal of Guidelines for ResEarch and Evaluation (AGREE) collaboration

6 domains to be covered in guideline development
  n Scope and purpose
  n Stakeholder involvement
  n Rigor of development (including searching and grading)
  n Clarity and presentation
  n Applicability
  n Editorial independence
What makes Guidelines Evidence-Based in 2008?

Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization

Checklist for reporting: 18 items

<table>
<thead>
<tr>
<th>Table: The COGS Checklist for Reporting Clinical Practice Guidelines*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
</tr>
<tr>
<td>1. Overview material</td>
</tr>
<tr>
<td>2. Focus</td>
</tr>
<tr>
<td>3. Goal</td>
</tr>
<tr>
<td>4. Users/setting</td>
</tr>
<tr>
<td>5. Target population</td>
</tr>
<tr>
<td>6. Developer</td>
</tr>
<tr>
<td>7. Funding source/sponsor</td>
</tr>
<tr>
<td>8. Literature search</td>
</tr>
<tr>
<td>9. Recommendation grading criteria</td>
</tr>
<tr>
<td>10. Evidence collection</td>
</tr>
<tr>
<td>11. Problem review</td>
</tr>
<tr>
<td>12. Update plan</td>
</tr>
<tr>
<td>13. Definitions</td>
</tr>
<tr>
<td>14. Recommendations and rationale</td>
</tr>
<tr>
<td>15. Patient preferences</td>
</tr>
<tr>
<td>16. Algorithm</td>
</tr>
<tr>
<td>17. Implementation considerations</td>
</tr>
</tbody>
</table>

* COGS = Conference on Guideline Standardization.

8. Evidence collection - Describe the methods used to search the scientific literature, including the range of dates and databases searched, and criteria applied to filter the retrieved evidence.

Traditional “literature reviews” are not systematic

- 158 reviews of clinical topics published in 6 general medical journals in 1996
- 2 satisfied all 10 methodologic criteria
- Median number of criteria satisfied = 1
- Less than ¼ described how evidence was identified, evaluated, or integrated
- 34% addressed a focused clinical question
- 111 (70%) made treatment recommendations
  - 48% provided an estimate of the magnitude of potential benefits
- The methodologic quality of clinical review articles is highly variable, and many of these articles do not specify systematic methods.

McAlister et al., Ann Int Med (1999)
In the United States, the NIH Consensus Development Program (www.consensus.nih.gov), which was started in 1977, sponsors evidence-based assessments of important medical issues. At present, each assessment includes a systematic literature review, prepared through the AHRQ; a public conference that features
What is a systematic review?

- Establish objectives + selection criteria
- Search for *eligible* studies
- Systematically apply selection criteria
  - In duplicate, reproducible, transparent
- Assess study quality + assemble data
  - Standard criteria
- Analyse results using meta-analysis, if appropriate and possible
  - Detect heterogeneity, reporting bias
  - Increase precision
- Prepare report
Systematic reviews - why?
Thrombolysis in acute myocardial infarction

Cumulative Odds Ratio (Log Scale)

Year  RCTs  Pts
1960   1     23
      2     65
1965   3     149
      7    1763
      10   2544
      11   2651
      15   3311
      17   3929
      22   5452
1980   23    5767
      27    6125
1985   30    6346
      33    6571
      43    21059
      54    22051
      65    47185
1990   67    47531
      70    48154

Textbook/Review Recommendations
Routine  Specific  Rare/Never  Experimental  Not Mentioned
M     1        2        9     5
M     1        1        7     3
M     5        2        2     1
M     15       8        1
M     6        1

Antman et al., JAMA (1993)
Selecting all relevant studies
Beta-blockers for acute Myocardial Infarction

Antman et al., JAMA (1992)
What makes Guidelines Evidence-Based in 2008?

Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization Checklist for reporting: 18 items

<table>
<thead>
<tr>
<th>Table: The COGS Checklist for Reporting Clinical Practice Guidelines*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
</tr>
<tr>
<td>1. Overview material</td>
</tr>
<tr>
<td>2. Focus</td>
</tr>
<tr>
<td>3. Goal</td>
</tr>
<tr>
<td>4. User/setting</td>
</tr>
<tr>
<td>5. Target population</td>
</tr>
<tr>
<td>6. Develop</td>
</tr>
<tr>
<td>7. Funding source/sponsor</td>
</tr>
<tr>
<td>8. Evidence collection</td>
</tr>
<tr>
<td>9. Recommendation grading criteria</td>
</tr>
<tr>
<td>10. Method for synthesizing evidence</td>
</tr>
<tr>
<td>11. Peer review</td>
</tr>
<tr>
<td>12. Update plan</td>
</tr>
<tr>
<td>13. Definitions</td>
</tr>
<tr>
<td>14. Recommendations and rationale</td>
</tr>
<tr>
<td>15. Potentially benefits and harms</td>
</tr>
<tr>
<td>16. Patient preferences</td>
</tr>
<tr>
<td>17. Algorithm</td>
</tr>
<tr>
<td>18. Implementation considerations</td>
</tr>
</tbody>
</table>

*COGS = Conference on Guideline Standardization.

9. Recommendation grading criteria - Describe the criteria used to rate the quality of evidence that supports the recommendations and the system for describing the strength of recommendations.
Evidence: Grading and Judgments

- Black box until recently
- Clear clinical question (PICO)
  - Population
  - Intervention
  - Comparison
  - Outcomes
- Explicit criteria for quality evaluation
  - Factors known for both experimental and observational studies
Quality of evidence

“The quality of evidence reflects the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation”

= Likelihood of bias explaining an effect
Hierarchy of evidence

**STUDY DESIGN**
- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations

**BIAS**
- Expert Opinion
Which grading system?

Recommendation for use of oral anticoagulation in patients with atrial fibrillation and rheumatic mitral valve disease

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Recommendation</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>n  B</td>
<td>Class I</td>
<td>AHA</td>
</tr>
<tr>
<td>n  A</td>
<td>1</td>
<td>ACCP</td>
</tr>
<tr>
<td>n  IV</td>
<td>C</td>
<td>SIGN</td>
</tr>
</tbody>
</table>

The same evidence – different classification
Grades of Recommendation Assessment, Development and Evaluation

GRADE Working Group

GRADE Uptake

- World Health Organization
- National Institute Clinical Excellence (NICE)
- Agency for Health Care Research and Quality (AHRQ)
- Canadian Agency for Drugs and Technology in Health
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Chest Physicians
- UpToDate
- British Medical Journal
- American College of Physicians
- Cochrane Collaboration
- European Society of Thoracic Surgeons
- Clinical Evidence
- Many other organizations
Determinants of quality

- RCTs start high
- Observational studies start low
- Factors that lower quality (bias)
- Factors that increase quality (bias is unlikely to explain observed effect)
Determinants of quality

What lowers quality/confidence?

5 factors

- shortcoming: detailed design and execution
- inconsistency
- indirectness
- reporting bias
- imprecision
Determinants of quality

- What raises quality?
  - 3 factors (mostly observational studies)
    - strong association
    - bias unlikely to explain an existing effect or its absence
    - dose effect
What makes Guidelines Evidence-Based in 2008?

Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization

Checklist for reporting: 18 items

| Table: The CGS Checklist for Reporting Clinical Practice Guidelines* |
|---|---|
| **Topic** | **Description** |
| 1. Overview material | Provide a structured abstract that includes the guideline's release date, status (original, revised, updated), and print and electronic access. |
| 2. Focus | Describe the primary disease/condition and intervention/service/technology that the guideline addresses. Include any alternative preventative, diagnostic, or therapeutic interventions that were considered during development. |
| 3. Goal | Describe the goal that following the guideline is expected to achieve, including the rationale for development of a guideline on the topic. |
| 4. Users/setting | Describe the intended users of the guideline (e.g., provider types, patients) and the settings in which the guideline is intended to be used. |
| 5. Target population | Describe the patient population eligible for guideline recommendations and list any exclusion criteria. |
| 6. Developer | Identify the organization(s) responsible for guideline development and the names/credentials/potential conflicts of interest of individuals involved in the guideline's development. |
| 7. Funding source/sponsor | Identify the funding source/sponsor and describe its role in developing and/or reporting the guideline. Disclose potential conflicts of interest. |
| 8. Evidence collection | Describe the methods used to search the scientific literature, including the range of dates and databases searched, and criteria applied to filter the retrieved evidence. |
| 9. Recommendation grading criteria | Describe the criteria used to judge the quality of evidence that supports the recommendations and the system for describing the strength of the recommendations. Recommendation strength communicates the importance of adherence to a recommendation and is based on both the quality of the evidence and the magnitude of anticipated benefits or harms. |
| 10. Method for synthesizing evidence | Describe how evidence was used to create recommendations, e.g., evidence tables, meta-analysis, decision analysis, evidence tables, meta-analysis, decision analysis. |
| 11. Procedure review | Describe the guideline developer reviewed and/or tested the guidelines prior to release. |
| 12. Update plan | State whether or not there is a plan to update the guideline and, if applicable, an expected date for the version of the guideline. |
| 13. Definitions | Define key terms and those critical to correct application of the guideline that might be subject to misinterpretation. |
| 14. Recommendations and rationale | State the recommended action precisely and the specific circumstances under which to perform it, identify each recommendation by describing the linkage between the recommendation and its supporting evidence, indicate the quality of evidence and the recommendation strength, based on the criteria described in 9. |
| 15. Potential benefits and harms | Describe anticipated benefits and potential risks associated with implementation. |
| 16. Patient preferences | Describe the role of patient preferences when a recommendation involves a substantial element of personal choice or value. |
| 17. Algorithm | Provide (when appropriate) a graphical description of the stages and decisions in clinical care described by the guideline. |
| 18. Implementation considerations | Describe anticipated barriers to application of the recommendations. Provide reference to any auxiliary documents for providers or patients that are intended to facilitate implementation. Support review criteria for measuring changes in care when the guideline is implemented. |

* CGS = Conference on Guideline Standardization.

10. Method for synthesizing evidence - Describe how evidence was used to create recommendations, e.g., evidence tables, meta-analysis, decision analysis.
### Should subcutaneous specific immunotherapy be used for treatment of allergic rhinitis in children without concomitant asthma?

**Author(s):** JLB & H/JS  
**Date:** 2007-08-08  
**Question:** Should subcutaneous immunotherapy be used in children with allergic rhinitis?  

**Settings:**  

#### Quality assessment

<table>
<thead>
<tr>
<th>No of studies</th>
<th>Design</th>
<th>Limitations</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>randomised trial</td>
<td>serious¹</td>
<td>serious²</td>
<td>no serious indirectness</td>
<td>serious³</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>development of asthma (follow-up 5 years)</td>
<td>serious⁵</td>
<td>no serious indirectness</td>
<td>serious⁶</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious indirectness</td>
<td>serious⁷</td>
<td>no serious imprecision</td>
<td>none</td>
</tr>
<tr>
<td>9</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious indirectness</td>
<td>serious⁸</td>
<td>serious⁹</td>
<td>none</td>
</tr>
<tr>
<td>13</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious indirectness</td>
<td>serious¹⁰</td>
<td>no serious imprecision</td>
<td>none</td>
</tr>
</tbody>
</table>

#### Summary of findings

<table>
<thead>
<tr>
<th></th>
<th>subcutaneous immunotherapy</th>
<th>control</th>
<th>Relative (95% CI)</th>
<th>Absolute</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>63</td>
<td>64</td>
<td>-</td>
<td>not pooled⁴</td>
</tr>
<tr>
<td>13</td>
<td>15/75</td>
<td>29/67</td>
<td>RR 0.46 (0.27 to 0.77)</td>
<td>234 fewer per 1000 (from 100 fewer to 316 fewer)</td>
</tr>
<tr>
<td>9</td>
<td>43/615⁹</td>
<td>3/453⁶</td>
<td>not pooled</td>
<td>not pooled</td>
</tr>
<tr>
<td>13</td>
<td>0/417⁸</td>
<td>0/303³</td>
<td>not pooled</td>
<td>not pooled</td>
</tr>
<tr>
<td></td>
<td>0/0¹⁰</td>
<td>0/0¹¹</td>
<td>not pooled</td>
<td>not pooled</td>
</tr>
</tbody>
</table>

¹ One old small trial and a subgroup analysis of one small recent trial.
² One trial showed benefit, while the other did not.
³ One trial found no difference between the SCIT and placebo groups (improvement in 11/25 and 11/26 children, respectively) and the other found improvement in symptom score of 14 mm on a 100 mm visual analog scale.
⁴ Reporting of symptoms did not allow for meta-analysis. One trial found no difference between the treated and placebo groups (improvement in 11/25 and 11/26 children, respectively). Second found that the symptoms scores measured on a visual analog scale improved more in the SCIT group compared to placebo (-21.5 mm vs -7.4 mm).
⁵ Post hoc subgroup analysis
⁶ Small trial with small number of events
⁷ Extrapolated from trials in adults
⁸ Most studies reported number of adverse events, rather than the number of participants in which one or more adverse events were observed.
⁹ Very small number of events
¹⁰ 19 events of 14,085 injections
¹¹ 1 event in 8,278 injections
# Transparent quality assessment

<table>
<thead>
<tr>
<th>Study Type</th>
<th>No of studies</th>
<th>Design</th>
<th>Limitations</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal symptoms (follow-up 3 to 5 years; Better indicated by less)</td>
<td>2</td>
<td>randomised trial</td>
<td>serious(^1)</td>
<td>serious(^2)</td>
<td>no serious indirectness</td>
<td>serious(^3)</td>
<td>none</td>
</tr>
<tr>
<td>Development of asthma (follow-up 5 years)</td>
<td>1</td>
<td>randomised trial</td>
<td>serious(^5)</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>serious(^6)</td>
<td>none</td>
</tr>
<tr>
<td>Non-life threatening systemic adverse events</td>
<td>13</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious(^7)</td>
<td>no serious imprecision</td>
<td>none</td>
</tr>
<tr>
<td>Anaphylactic shock</td>
<td>9</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious(^7)</td>
<td>serious(^9)</td>
<td>none</td>
</tr>
<tr>
<td>Adrenaline use for systemic reaction</td>
<td>13</td>
<td>randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious(^7)</td>
<td>no serious imprecision</td>
<td>none</td>
</tr>
</tbody>
</table>

\(^1\) One old small trial and a subgroup analysis of one small recent trial.

\(^2\) One trial showed benefit while the other did not.
# Summary of Findings Tables – Cochrane Collaboration

## Erythropoiesis Stimulants (epo) Compared to Placebo for Anemia from Cancer Chemotherapy

**Patient or population:** Patients with anemia from cancer chemotherapy  
**Settings:** Outpatient cancer treatment  
**Intervention:** Erythropoiesis stimulants (epo)  
**Comparison:** Placebo

### Outcomes

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All cause mortality</strong> (follow-up: 0-36 months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td>HR 1.11 (1 to 1.22)</td>
<td>6918</td>
<td>☢☢☢</td>
<td></td>
</tr>
<tr>
<td>Low risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>268 per 1000</td>
<td>293 per 1000 (269 to 317)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 per 1000</td>
<td>110 per 1000 (100 to 121)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 per 1000</td>
<td>537 per 1000 (500 to 571)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Thromboembolic events</strong></td>
<td></td>
<td>RR 1.69 (1.36 to 2.1)</td>
<td>6032</td>
<td>☢☢☢</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td>RR 1.69 (1.36 to 2.1)</td>
<td>6032</td>
<td>☢☢☢</td>
<td></td>
</tr>
<tr>
<td>Low risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41 per 1000</td>
<td>69 per 1000 (58 to 86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 per 1000</td>
<td>17 per 1000 (14 to 21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80 per 1000</td>
<td>135 per 1000 (103 to 163)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Complete response of tumor to chemotherapy</strong></td>
<td>613 per 1000 (584 to 574)</td>
<td>RR 1.0 (0.92 to 1.1)</td>
<td>688</td>
<td>☢☢☢</td>
<td></td>
</tr>
</tbody>
</table>

---

1. Overall heterogeneity not significant, but underlying clinical heterogeneity due to risk of VTE, treatment regimens, and epo protocols (starting and stopping Hb)  
2. CI includes no effect and clinically important increase in mortality  
3. Criteria for determining and reporting VTE variable in studies; trials reporting varying combinations of DVT, PE, TIA, stroke, and MI  
4. Only 5 trials reported this outcome; does not include the largest trials powered for mortality benefit  
5. Tests of heterogeneity I square were significant. Reduced risk of transfusion evidence in subgroups defined by different starting Hb level, but size of benefit differs. Clinical heterogeneity in control rate transfusions, tumor type and chemo regimen, and protocols for determining transfusion need  
6. All trials support substantial benefit but significant heterogeneity in magnitude of benefit; clinical heterogeneity in starting Hb levels, underlying chemo regimens and tumor types, and risk of anemia  
7. Size of RR (3.4 pooled, range 2 to 9) would qualify as large effect
**Erythropoiesis stimulants (epo) compared to placebo for anemia from cancer chemotherapy**

**Patient or population:** patients with anemia from cancer chemotherapy  
**Settings:** Outpatient cancer treatment  
**Intervention:** Erythropoiesis stimulants (epo)  
**Comparison:** placebo

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All cause mortality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(follow-up: 4 - 36 months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>268 per 1000</td>
<td>293 per 1000</td>
<td>HR 1.11 (1 to 1.22)</td>
<td>631     (35)</td>
<td></td>
<td>low 1,2</td>
</tr>
<tr>
<td>Low risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 per 1000</td>
<td>110 per 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 per 1000</td>
<td>537 per 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thromboembolic events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41 per 1000</td>
<td>89 per 1000</td>
<td>RR 1.89 (1.36 to 2.1)</td>
<td>8292     (30)</td>
<td></td>
<td>low 1,3</td>
</tr>
</tbody>
</table>
Strength of recommendation

“The degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects”

Desirable effects
- health benefits
- less burden
- savings

Undesirable effects
- harms
- more burden
- costs
Strength of recommendation

**STRONG**

Almost all patients would choose
Informed consent
Consider as quality performance measure

**WEAK**

Informed patient choice varies
Participatory decision making
Not ready as quality measure
### Determinants of the strength of recommendation

<table>
<thead>
<tr>
<th>Factors that can strengthen a recommendation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the evidence</td>
<td></td>
</tr>
<tr>
<td>Balance between desirable and undesirable effects</td>
<td></td>
</tr>
<tr>
<td>Values and preferences</td>
<td></td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td></td>
</tr>
</tbody>
</table>
Determinants of the strength of recommendation

<table>
<thead>
<tr>
<th>Factors that can strengthen a recommendation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the evidence</td>
<td>The higher the quality of evidence, the more likely is a strong recommendation.</td>
</tr>
<tr>
<td>Balance between desirable and undesirable effects</td>
<td>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 certainty for that benefit, the more likely weak recommendation warranted.</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.</td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td>The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted.</td>
</tr>
</tbody>
</table>
Systematic reviews

DO:
- reduce bias in our assessment of the evidence

CAN:
- provide statistical summaries of effects
- increase precision and certainty around effect estimates

BUT:
- cost up to 50,000 $ (low compared to trials)
- may take a long time to complete
Experience

- WHO
- American Thoracic Society
- Allergic Rhinitis in Asthma
- American College of Chest Physicians
Summary

- Systematic reviews – requirement for high quality guidelines
- Judgments about quality need to be transparent
- Criteria for moving from evidence to recommendations
- Simple, systematic
  - four categories of quality of evidence
  - two grades for strength of recommendations
- Transparency in decision making and judgments is key
Thank you
Design and Execution

- Limitations
  - Lack of concealment
  - Intention to treat principle violated
  - Inadequate blinding
  - Loss to follow-up
  - Early stopping for benefit

- 13 RCTs bacterial extract (immunomodulation) for preventing exacerbation
  - Unclear concealment of randomization
  - Questionable intention to treat
  - Inadequate attention to loss to follow-up
Consistency of results

- consistency of results
- if inconsistency, look for explanation
  - patients, intervention, outcome, methods
- unexplained inconsistency downgrade quality
  - oxygen for day-to-day dyspnea in COPD with exercise hypoxemia
  - five cross-over RCTs oxygen versus placebo
  - 4 no benefit, 1 substantial benefit
Directness of Evidence

- indirect comparisons
  - interested in A versus B
  - have A versus C and B versus C
  - formoterol versus salmeterol versus tiotropium

- differences in
  - patients (mild versus severe COPD)
  - interventions (all inhaled steroids)
  - outcomes (long-term health-related quality of life, short-term functional capacity, laboratory exercise, spirometry)
Summary

- Many challenges – many of them financial
- Collaboration and coordination!
- Criteria for moving from evidence to recommendations
- Simple, transparent, systematic
  - four categories of quality of evidence
  - two grades for strength of recommendations
- Transparency in decision making and judgments is key
Thank you
Determinants of the strength of recommendation

<table>
<thead>
<tr>
<th>Factors that can strengthen a recommendation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the evidence</td>
<td>Low</td>
</tr>
<tr>
<td>Balance between desirable and undesirable effects</td>
<td>Moderately balance</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>Fair bit of variability</td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td>Relatively high cost</td>
</tr>
</tbody>
</table>
Another COPD guidelines

1.2.14 Mucolytic therapy

1.2.14.1 Mucolytic drug therapy should be considered in patients with a chronic cough productive of sputum.

1.2.14.2 Mucolytic therapy should be continued if there is symptomatic improvement (for example, reduction in frequency of cough and sputum production).

1.2.15 Anti-oxidant therapy

1.2.15.1 Treatment with alpha-tocopherol and beta-carotene supplements, alone or in combination, is not recommended.
Determinants of the strength of recommendation

Factors that can weaken the strength of a recommendation. Example:

<table>
<thead>
<tr>
<th></th>
<th>Decision</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower quality evidence</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty about the balance of benefits versus harms and burdens</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty or differences in values</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty about whether the net benefits are worth the costs</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
</tbody>
</table>

Table. Decisions about the strength of a recommendation
Frequent “yes” answers will increase the likelihood of a weak recommendation
Questions for you

1. If you have been part of guideline panels, do you have prepared an evidence profile?
2. Are systematic reviews for every recommendation in your guidelines a reality/possibility?
3. What about cost – how do you deal with cost and how should we deal with it?
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.
- **Clinicians**: Be 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.
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...”.
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.

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
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., Lancet ID, 2007
GRADE Quality of Evidence

Extent to which confidence in estimate of effect adequate to support decision

- high: considerable confidence in estimate of effect.
- moderate: further research likely to have impact on confidence in estimate, may change estimate.
- low: further research is very likely to impact on confidence, likely to change the estimate.
- very low: any estimate of effect is very uncertain
There always is evidence

The better the research and the evidence, the more confident the decision

Evidence alone is never sufficient to make a clinical decision
Do evidence based guidelines make a difference?

Non-rigorous guidelines:
- Create noise & bias
- Make more aggressive recommendations
- Can harm patients and impair research efforts
- Can reduce credibility of professional societies

Evidence-based clinical practice guidelines can:
- reduce delivery of inappropriate care
- support introduction of new knowledge into clinical practice

“Practice guidelines ... have been demonstrated to improve patient outcomes and lower cost”

...be based on sound scientific evidence and implemented in an effective manner

S. Weingarten. Hospital Medicine 2005
Evidence-based Medicine

The conscientious and judicious use of current best evidence from clinical care research in making health care decisions
Confidence in evidence

- There always is evidence
  - “When there is a question there is evidence”
- Evidence alone is never sufficient to make a clinical decision
- Better research $\Rightarrow$ greater confidence in the evidence and decisions
Hierarchy of evidence

**STUDY DESIGN**
- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert Opinion

**BIAS**
- Expert Opinion
<table>
<thead>
<tr>
<th>Guideline development process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritise Problems, establish panel</td>
</tr>
<tr>
<td>Systematic Review</td>
</tr>
<tr>
<td>Evidence Profile</td>
</tr>
<tr>
<td>Relative importance of outcomes</td>
</tr>
<tr>
<td>Overall quality of evidence</td>
</tr>
<tr>
<td>Benefit – downside evaluation</td>
</tr>
<tr>
<td>Strength of recommendation</td>
</tr>
<tr>
<td>Implementation and evaluation of guidelines</td>
</tr>
</tbody>
</table>

GRADE
## Guideline development process

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritise Problems, establish panel</td>
<td></td>
</tr>
<tr>
<td>Systematic Review</td>
<td></td>
</tr>
<tr>
<td>Evidence Profile</td>
<td></td>
</tr>
<tr>
<td>Relative importance of outcomes</td>
<td></td>
</tr>
<tr>
<td>Overall quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Benefit – downside evaluation</td>
<td></td>
</tr>
<tr>
<td>Strength of recommendation</td>
<td></td>
</tr>
<tr>
<td>Implementation and evaluation of guidelines</td>
<td></td>
</tr>
</tbody>
</table>
Evidence based clinical decisions

Clinical state and circumstances

Expertise

Patient values and preferences

Research evidence

Equal for all

Haynes et al. 2002