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

Montreal, February 12, 2010
McGill University, Montreal Chest Institute

GRADE
Content

- Background about GRADE

- GRADE approach
  - Quality of evidence
  - Strength of recommendations
Evidence based clinical decisions

- Clinical state and circumstances
- Research evidence
- Expertise
- Patient values and preferences

Haynes et al. 2002
Confidence in evidence

- There always is evidence
  - “When there is a question there is evidence”
- Better research $\implies$ greater confidence in the evidence and decisions
Hierarchy of evidence based on quality

**STUDY DESIGN**
- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert Opinion
“Everything should be made as simple as possible but not simpler.”
Can you explain the following?

- Concealment of randomization
- Blinding (who is blinded in a double blinded trial?)
- Confounding, effect modification & ext. validity
- Intention to treat analysis and its correct application
- P-values and confidence intervals
Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell
Simple hierarchies are (too) simplistic

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

*Expert Opinion*

*Schünemann & Bone, 2003*
Which hierarchy?

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>B</td>
<td>Class I</td>
<td>AHA</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>ACCP</td>
</tr>
<tr>
<td>IV</td>
<td>C</td>
<td>SIGN</td>
</tr>
</tbody>
</table>
What to do?
Grades of Recommendation Assessment, Development and Evaluation

GRADE WORKING GROUP

- Since 2000
- Guideline developers, methodologists & clinicians from around the world

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide.

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
Guideline development Process

Prioritise Problems, establish panel, questions

Systematic Review

Evidence Profile

Relative importance of outcomes

Overall quality of evidence

Benefit – downside evaluation

Strength of recommendation

Implementation and evaluation of guidelines
The GRADE approach

Clear separation of 2 issues:

1) 4 categories of quality of evidence: ⊕⊕⊕⊕ (High), ⊕⊕⊕ (Moderate), ⊕⊕ (Low), ⊕ (Very low)?
   - methodological quality of evidence
   - likelihood of bias
   - by outcome and across outcomes

2) Recommendation: 2 grades – weak/conditional or strong (for or against)?
   - Quality of evidence only one factor
   - Balance of benefits and downsides, values and preferences, resource use

*www.GradeWorking-Group.org
GRADE Quality of Evidence

In the context of a systematic review
  - The quality of evidence reflects the extent to which we are confident that an estimate of effect is correct.

In the context of making recommendations
  - 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 and confidence in an outcome

Figure 1. Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).
Determinants of quality

- RCTs start ⊕⊕⊕⊕ (high)

- Observational studies start at ⊕⊕○○ (low)

- 5 factors that can lower quality
  1. Limitations in detailed 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
1. Design and Execution/Risk of Bias

- limitations
  - lack of concealment
  - intention to treat principle violated
  - inadequate blinding
  - loss to follow-up
  - early stopping for benefit
  - selective outcome reporting
Design and Execution/RoB

Regular treatment with salmeterol for chronic asthma: serious adverse events (Review)

Cates CJ, Cates MJ

Figure 4. Risk of bias summary: review authors’ judgments about each risk of bias item for each included study.
Design and Execution

Regular treatment with salmeterol for chronic asthma: serious adverse events (Review)

Cates CJ, Cates MJ
Design and Execution

Figure 3. Risk of bias graph: review authors’ judgments about each risk of bias item presented as percentages across all included studies.

Overall judgment required
Who believes the risk of bias is of concern?

Yes

No

Don’t know or undecided
## Detailed study design and execution

### Mortality, cancer and anticoagulation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Altinbas 2004</td>
<td>?</td>
<td>?</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Kakkar 2004</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Klerk 2005</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Lebeau 1994</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sideras 2006</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
</tbody>
</table>

Five trials

### Analysis 01.01. Comparison 01 Heparin vs placebo, Outcome 01 Mortality over duration of study

**Review:** Parenteral anticoagulation for prolonging survival in patients with cancer who have no other indication for anticoagulation

**Comparison:** 01 Heparin vs placebo

**Outcome:** 01 Mortality over duration of study

<table>
<thead>
<tr>
<th>Study</th>
<th>Heparin N</th>
<th>Control N</th>
<th>log [Hazard Ratio] (SE)</th>
<th>Hazard Ratio (Random) 95% CI</th>
<th>Weight (%)</th>
<th>Hazard Ratio (Random) 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 SCLC</td>
<td>138</td>
<td>139</td>
<td>-0.33 (0.12)</td>
<td></td>
<td>22.7</td>
<td>0.72 [ 0.56, 0.91 ]</td>
</tr>
<tr>
<td>Akinbas 2004</td>
<td>42</td>
<td>42</td>
<td>-0.65 (0.23)</td>
<td></td>
<td>10.8</td>
<td>0.52 [ 0.33, 0.82 ]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>34.5</td>
<td>0.65 [ 0.49, 0.87 ]</td>
</tr>
<tr>
<td>Test for heterogeneity chi-square = 1.48, df = 1, p = 0.22, P = 0.32, 41%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect z = 2.93, p = 0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**02 Advanced cancer**

<table>
<thead>
<tr>
<th></th>
<th>Heparin N</th>
<th>Control N</th>
<th>log [Hazard Ratio] (SE)</th>
<th>Hazard Ratio (Random) 95% CI</th>
<th>Weight (%)</th>
<th>Hazard Ratio (Random) 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kakkar 2004</td>
<td>190</td>
<td>184</td>
<td>-0.24 (0.11)</td>
<td></td>
<td>25.9</td>
<td>0.79 [ 0.63, 0.98 ]</td>
</tr>
<tr>
<td>Hekk 2005</td>
<td>145</td>
<td>154</td>
<td>-0.32 (0.11)</td>
<td></td>
<td>25.5</td>
<td>0.75 [ 0.60, 0.94 ]</td>
</tr>
<tr>
<td>Sideras 2006</td>
<td>68</td>
<td>69</td>
<td>0.14 (0.19)</td>
<td></td>
<td>14.1</td>
<td>1.15 [ 0.79, 1.66 ]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65.5</td>
<td>0.82 [ 0.68, 1.03 ]</td>
</tr>
<tr>
<td>Test for heterogeneity chi-square = 3.84, df = 2, p = 0.15, P = 47.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect z = 1.68, p = 0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
<td>0.77 [ 0.65, 0.91 ]</td>
</tr>
<tr>
<td>Test for heterogeneity chi-square = 7.63, df = 4, p = 0.11, P = 47.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect z = 3.01, p = 0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Plot](image)
Who believes the risk of bias is of concern?

Yes
No
Don’t know or undecided
2. Inconsistency of results (Heterogeneity)

- if inconsistency, look for explanation
  - patients, intervention, outcome
- unexplained inconsistency downgrade quality
Heparin or vitamin K antagonists for survival in patients with cancer

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>LMWH Events</th>
<th>Total</th>
<th>VKA Events</th>
<th>Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cesarone 2003</td>
<td>2</td>
<td>96</td>
<td>3</td>
<td>96</td>
<td>0.8%</td>
<td>0.67 [0.11, 3.90]</td>
<td></td>
</tr>
<tr>
<td>Meyer 2002</td>
<td>22</td>
<td>71</td>
<td>29</td>
<td>75</td>
<td>12.0%</td>
<td>0.50 [0.51, 1.28]</td>
<td></td>
</tr>
<tr>
<td>Lee 2003</td>
<td>130</td>
<td>336</td>
<td>136</td>
<td>336</td>
<td>69.5%</td>
<td>0.96 [0.79, 1.15]</td>
<td></td>
</tr>
<tr>
<td>Deitcker 2006</td>
<td>22</td>
<td>67</td>
<td>11</td>
<td>34</td>
<td>6.9%</td>
<td>1.01 [0.56, 1.84]</td>
<td></td>
</tr>
<tr>
<td>Hull 2006</td>
<td>20</td>
<td>100</td>
<td>19</td>
<td>100</td>
<td>7.6%</td>
<td>1.05 [0.60, 1.85]</td>
<td></td>
</tr>
<tr>
<td>Lopez Beret 2001</td>
<td>7</td>
<td>17</td>
<td>6</td>
<td>18</td>
<td>3.2%</td>
<td>1.24 [0.52, 2.94]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>687</strong></td>
<td><strong>659</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>659</strong></td>
<td></td>
<td><strong>0.95 [0.81, 1.11]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Total events: 203
Heterogeneity: Tau² = 0.00; Chi² = 1.24, df = 5 (P = 0.94); I² = 0%
Test for overall effect: Z = 0.62 (P = 0.53)

Who believes there is important inconsistency (rather than random error)?

Yes
No
Don’t know or undecided
Non-steroidal drug use and risk of pancreatic cancer


Meta-analysis: the use of non-steroidal anti-inflammatory drugs and pancreatic cancer risk for different exposure categories.

Who believes there is important inconsistency (rather than random error)?

Yes
No
Don’t know or undecided
Analysis 01.04. Comparison 01 Active versus placebo, Outcome 04 Nasal symptom

Review: Allergen injection immunotherapy for seasonal allergic rhinitis
Comparison: 01 Active versus placebo
Outcome: 04 Nasal symptom

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>Control</th>
<th>Standardised Mean Difference (Random)</th>
<th>Weight (%)</th>
<th>Standardised Mean Difference (Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean(SD)</td>
<td>N</td>
<td>Mean(SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Balda 1998</td>
<td>49</td>
<td>2.80 (4.29)</td>
<td>56</td>
<td>4.42 (4.40)</td>
<td>-0.37 [-0.76, 0.02]</td>
</tr>
<tr>
<td>Bousquet 1997</td>
<td>39</td>
<td>61.00 (17.90)</td>
<td>19</td>
<td>109.00 (16.80)</td>
<td>-2.70 [-3.45, -1.95]</td>
</tr>
<tr>
<td>D’Amato 1995</td>
<td>9</td>
<td>24.60 (18.40)</td>
<td>11</td>
<td>78.50 (5.60)</td>
<td>-3.98 [-5.62, -2.35]</td>
</tr>
<tr>
<td>Dolz 1996</td>
<td>18</td>
<td>0.13 (0.20)</td>
<td>10</td>
<td>1.38 (0.15)</td>
<td>-6.59 [-8.60, -4.57]</td>
</tr>
<tr>
<td>Zenner 1997</td>
<td>41</td>
<td>44.49 (32.23)</td>
<td>40</td>
<td>63.30 (38.31)</td>
<td>-0.53 [-0.97, -0.08]</td>
</tr>
<tr>
<td>Frew 2006</td>
<td>187</td>
<td>2.05 (1.52)</td>
<td>89</td>
<td>2.93 (1.93)</td>
<td>-0.53 [-0.78, -0.27]</td>
</tr>
<tr>
<td>Ferrer 2005</td>
<td>22</td>
<td>0.55 (0.39)</td>
<td>20</td>
<td>0.91 (0.63)</td>
<td>-0.68 [-1.31, -0.06]</td>
</tr>
<tr>
<td>Mirone 2004</td>
<td>11</td>
<td>4.00 (3.30)</td>
<td>12</td>
<td>8.30 (4.20)</td>
<td>-1.09 [-1.98, -0.20]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>376</td>
<td>257</td>
<td>100.0</td>
<td>-1.59 [-2.29, -0.89]</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity chi-square=81.44 df=7 p=<0.0001 I² =91.4%
Test for overall effect z=4.45 p<0.00001
Inconsistency when 1 study?

- Do not downgrade
3. Directness of Evidence

- differences in
  - populations/patients (mild versus severe COPD, older, sicker or more co-morbidity)
  - interventions (all inhaled steroids, new vs. old)
  - outcomes (important vs. surrogate; long-term health-related quality of life, short –term functional capacity, laboratory exercise, spirometry)

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

<table>
<thead>
<tr>
<th>Source of indirectness</th>
<th>Question of interest</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect comparison</td>
<td>Early emergency department systemic corticosteroids to treat acute exacerbations in adult patients with asthma</td>
<td>Both oral and intravenous routes are effective but there is no direct comparison of these two routes of administration in adults.</td>
</tr>
</tbody>
</table>
## Difference in populations

<table>
<thead>
<tr>
<th>Source of indirectness</th>
<th>Question of interest</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences in populations</td>
<td>Anti-leukotrienes plus inhaled glucocorticosteroids vs. inhaled glucocorticosteroids alone to prevent asthma exacerbations and nighttime symptoms in patients with chronic asthma and allergic rhinitis.</td>
<td>Trials that measured asthma exacerbations and nighttime symptoms did not include patients with allergic rhinitis.</td>
</tr>
</tbody>
</table>
### Differences in intervention

<table>
<thead>
<tr>
<th>Source of indirectness</th>
<th>Question of interest</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences in intervention</td>
<td>Avoidance of pet allergens in non-allergic infants or preschool children to prevent development of allergy.</td>
<td>Available studies used multifaceted interventions directed at multiple potential risk factors in addition to pet avoidance.</td>
</tr>
</tbody>
</table>
## Differences in outcomes

<table>
<thead>
<tr>
<th>Source of indirectness</th>
<th>Question of interest</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences in outcomes of interest</td>
<td>Intranasal glucocorticosteroids vs. oral H$_1$-antihistamines in children with seasonal allergic rhinitis</td>
<td>In the available study parents were rating the symptoms and quality of life of their teenage children, instead the children themselves</td>
</tr>
</tbody>
</table>
Funnel plot

Symmetrical:
No publication bias
Funnel plot

Asymmetrical: Publication bias?

File drawer problem
No interest in publishing or being published
5. Imprecision

- small sample size
  - small number of events
- wide confidence intervals
  - uncertainty about magnitude of effect
- extent to which confidence in estimate of effect adequate to support decision
**Fluoroquinolone prophylaxis in neutropenia: infection-related mortality**

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Quinolones, n/n</th>
<th>Placebo, n/n</th>
<th>RR (Fixed) (95% CI)</th>
<th>Weight, %</th>
<th>RR (Fixed) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleijfer et al., 1980 (23)</td>
<td>0/53</td>
<td>9/52</td>
<td></td>
<td>25.80</td>
<td>0.05 (0.00–0.87)</td>
</tr>
<tr>
<td>Karp et al., 1987 (16)</td>
<td>6/35</td>
<td>3/33</td>
<td></td>
<td>8.31</td>
<td>1.89 (0.51–6.93)</td>
</tr>
<tr>
<td>Schroeder et al., 1992 (22)</td>
<td>0/40</td>
<td>2/35</td>
<td></td>
<td>7.16</td>
<td>0.18 (0.01–3.54)</td>
</tr>
<tr>
<td>Talbot et al., 1993 (24)</td>
<td>1/62</td>
<td>2/57</td>
<td></td>
<td>5.61</td>
<td>0.46 (0.04–4.93)</td>
</tr>
<tr>
<td>Moreau et al., 1995 (94)</td>
<td>0/44</td>
<td>0/44</td>
<td></td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>Carlson et al., 1997 (13)</td>
<td>0/45</td>
<td>0/45</td>
<td></td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>Thomas et al., 2000 (25)</td>
<td>5/99</td>
<td>5/52</td>
<td></td>
<td>17.64</td>
<td>0.53 (0.16–1.73)</td>
</tr>
<tr>
<td>Menova et al., 2001 (20)</td>
<td>0/36</td>
<td>5/34</td>
<td></td>
<td>15.21</td>
<td>0.09 (0.00–1.50)</td>
</tr>
<tr>
<td>Tjan-Heijnen et al., 2001 (26)</td>
<td>0/82</td>
<td>5/79</td>
<td></td>
<td>15.07</td>
<td>0.09 (0.00–1.56)</td>
</tr>
<tr>
<td>Lee et al., 2002 (17)</td>
<td>2/46</td>
<td>2/49</td>
<td></td>
<td>5.21</td>
<td>1.07 (0.16–7.25)</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>542</strong></td>
<td><strong>480</strong></td>
<td></td>
<td><strong>100.00</strong></td>
<td><strong>0.38 (0.21–0.69)</strong></td>
</tr>
</tbody>
</table>

Total events: 14 (quinolones), 33 (placebo)

Test for heterogeneity: chi-square = 11.41 (P = 0.12), I² = 38.6%

Test for overall effect: Z = 3.20 (P = 0.001)
Example: Bleeding in the hospital

What can raise quality?

1. large magnitude can upgrade (RRR 50%)
   - very large two levels (RRR 80%)
   - common criteria
     - everyone used to do badly
     - almost everyone does well
   - oral anticoagulation for mechanical heart valves
   - insulin for diabetic ketoacidosis
   - hip replacement for severe osteoarthritis
   - parachutes to prevent death when jumping from airplanes
What can raise quality?

2. dose response relation
   - (higher INR – increased bleeding)
   - childhood lymphoblastic leukemia
     - risk for CNS malignancies 15 years after cranial irradiation
     - no radiation: 1% (95% CI 0% to 2.1%)
     - 12 Gy: 1.6% (95% CI 0% to 3.4%)
     - 18 Gy: 3.3% (95% CI 0.9% to 5.6%)

3. all plausible confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
All plausible confounding would result in an underestimate of the treatment effect

- Higher death rates in private for-profit versus private not-for-profit hospitals
  - patients in the not-for-profit hospitals likely sicker than those in the for-profit hospitals
  - for-profit hospitals are likely to admit a larger proportion of well-insured patients than not-for-profit hospitals (and thus have more resources with a spill over effect)
All plausible confounding would result in an overestimate of effect

- Hypoglycaemic drug phenformin causes lactic acidosis
- The related agent metformin is under suspicion for the same toxicity.
- Large observational studies have failed to demonstrate an association
  - Clinicians would be more alert to lactic acidosis in the presence of the agent
# Quality assessment criteria

<table>
<thead>
<tr>
<th>Quality of evidence</th>
<th>Study design</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Randomised trials</td>
<td>Study quality:</td>
<td>Strong association:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serious limitations</td>
<td>Strong, no plausible confounders</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very serious limitations</td>
<td>Very strong, no major threats to validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Important inconsistency</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Observational studies</td>
<td></td>
<td>Evidence of a dose response gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Directness:</td>
<td>All plausible confounders would have reduced the effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Some uncertainty</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Major uncertainty</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sparse or imprecise data</td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td></td>
<td>High probability of publication bias</td>
<td></td>
</tr>
</tbody>
</table>
### Evidence Profiles/Summaries

**Question 42 [profile 2]**
**Date:** 2007-05-27
**Question:** Should ketotifen be used for long-term control of asthma and wheeze in children?
**Bibliography:** 1. Bessler D., Mila A., Ducharme F.M., Forster J., Schwarzer G. Ketotifen alone or as additional medication for long-term control of asthma and wheeze in children. Cochrane database of systematic reviews (Online), 2004.CD001584

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
<th>No of patients</th>
<th>Effect</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No of studies</strong></td>
<td><strong>Design</strong></td>
<td><strong>Limitations</strong></td>
<td><strong>Inconsistency</strong></td>
<td><strong>Indirectness</strong></td>
</tr>
<tr>
<td><strong>Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>very serious</td>
</tr>
<tr>
<td><strong>Asthma exacerbations (follow-up 12 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>serious</td>
</tr>
<tr>
<td><strong>Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>serious</td>
</tr>
<tr>
<td><strong>Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>serious</td>
</tr>
<tr>
<td><strong>Efficacy evaluated by physicians (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 randomised trial</td>
<td>serious</td>
<td>serious</td>
<td>very serious</td>
<td>no serious imprecision</td>
</tr>
<tr>
<td><strong>Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>serious</td>
</tr>
<tr>
<td><strong>Sedation (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>very serious</td>
</tr>
<tr>
<td><strong>Weight gain (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>very serious</td>
</tr>
<tr>
<td><strong>Withdrawal from study due to side effects (follow-up 10 to 16 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 randomised trial</td>
<td>no serious limitations</td>
<td>no serious inconsistency</td>
<td>serious</td>
<td>very serious</td>
</tr>
</tbody>
</table>

---

1. Most trials have been published before 1990, when reporting of methods were not as stringent as they are now, which may lead to inadequate reporting of good methods rather than bad methods per se.
2. Inhaled corticosteroids were allowed as additional treatment in eight trials. There was not enough information in the studies to assess the effect of ketotifen as an add-on therapy to inhaled corticosteroids that are the mainstay of therapy of asthma today.
3. Results include small or large effect.
4. Very small trials with few events.
5. Only four trials reported that outcome, but we did not downgrade since we already downgraded for very serious imprecision.
## Evidence Profiles/Summaries

### 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><strong>Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²</td>
<td>very serious¹,²,⁴</td>
<td>none⁵,⁶</td>
</tr>
<tr>
<td><strong>Asthma exacerbations (follow-up 12 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²,⁷</td>
<td>serious⁴</td>
<td>none⁵</td>
</tr>
<tr>
<td><strong>Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²</td>
<td>serious⁴</td>
<td>none⁵,⁶</td>
</tr>
<tr>
<td><strong>Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²</td>
<td>serious⁴</td>
<td>none⁵</td>
</tr>
<tr>
<td><strong>Efficacy evaluated by physicians (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>randomised trial</td>
<td>serious⁹</td>
<td>serious¹⁰</td>
<td>very serious²,¹¹</td>
<td>no serious imprecision</td>
<td>reporting bias¹²</td>
</tr>
<tr>
<td><strong>Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²</td>
<td>serious⁴</td>
<td>none⁵</td>
</tr>
<tr>
<td><strong>Sedation (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>very serious¹,³,⁵</td>
<td>none⁵</td>
</tr>
<tr>
<td><strong>Weight gain (follow-up 10 to 26 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>very serious¹,³,⁵</td>
<td>none⁵,⁶</td>
</tr>
<tr>
<td><strong>Withdrawal from study due to side effects (follow-up 10 to 16 weeks)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>randomised trial</td>
<td>no serious limitations¹</td>
<td>no serious inconsistency</td>
<td>serious²</td>
<td>very serious¹,⁷</td>
<td>none⁵,⁶</td>
</tr>
</tbody>
</table>
### Evidence Profiles/Summaries

#### Question 42 [profile 2]
**Date:** 2007-05-27  
**Question:** Should ketotifen be used for long-term control of asthma and wheeze in children?  
**Bibliography:** Bessler D., Mila A., Ducharme F.M., Forster J., Schwarzer G. Ketotifen alone or as additional medication for long-term control of asthma and wheeze in children. Cochrane database of systematic reviews (Online), 2004 CD001564

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No of studies</strong></td>
<td><strong>Design</strong></td>
<td><strong>Limitations</strong></td>
</tr>
<tr>
<td><strong>Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)</strong></td>
<td>4</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Asthma exacerbations (follow-up 12 weeks)</strong></td>
<td>2</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)</strong></td>
<td>4</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)</strong></td>
<td>7</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Efficacy evaluated by physicians (follow-up 10 to 26 weeks)</strong></td>
<td>10</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)</strong></td>
<td>12</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Sedation (follow-up 10 to 26 weeks)</strong></td>
<td>7</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Weight gain (follow-up 10 to 26 weeks)</strong></td>
<td>5</td>
<td>Randomised trial</td>
</tr>
<tr>
<td><strong>Withdrawal from study due to side effects (follow-up 10 to 16 weeks)</strong></td>
<td>3</td>
<td>Randomised trial</td>
</tr>
</tbody>
</table>

¹ Most trials have been published before 1990, when reporting of methods were not as stringent as they are now, which may lead to inadequate reporting or good methods rather than bad methods per se.  
² Inhaled corticosteroids were allowed as additional intervention in eight trials. There was not enough information in the studies to assess the effect of ketotifen as an add-on therapy to inhaled corticosteroids that are the mainstay of therapy of asthma today.  
³ Results include small or large effect.  
⁴ Very small trials with few events.  
⁵ Only four trials reported that outcome, but we did not downgrade since we already downgraded for very serious imprecision.
# Evidence Profiles/Summaries

## Summary of findings

<table>
<thead>
<tr>
<th>No of patients</th>
<th>Effect</th>
<th>Quality</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ketotifen</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>control</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relative (95% CI)</strong></td>
<td><strong>Absolute</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>72</td>
<td>76</td>
<td>-</td>
<td>SMD -0.49 (-0.16 to -0.82)</td>
</tr>
<tr>
<td>10/105</td>
<td>32/104</td>
<td>RR 0.31 (0.19 to 0.59)</td>
<td>213 fewer per 1000 (from 126 fewer to 249 fewer)</td>
</tr>
<tr>
<td>21/156</td>
<td>73/150</td>
<td>RR 0.28 (0.13 to 0.58)</td>
<td>351 fewer per 1000 (from 205 fewer to 424 fewer)</td>
</tr>
<tr>
<td>101/301</td>
<td>143/298</td>
<td>RR 0.71 (0.52 to 0.96)</td>
<td>139 fewer per 1000 (from 19 fewer to 230 fewer)</td>
</tr>
<tr>
<td>113/310</td>
<td>188/315</td>
<td>RR 0.6 (0.46 to 0.79)</td>
<td>239 fewer per 1000 (from 125 fewer to 322 fewer)</td>
</tr>
<tr>
<td>50/76</td>
<td>21/73</td>
<td>RR 2.39 (1.64 to 3.48)</td>
<td>400 more per 1000 (from 184 more to 714 more)</td>
</tr>
<tr>
<td>45/218</td>
<td>26/221</td>
<td>RR 1.69 (1.11 to 2.59)</td>
<td>81 more per 1000 (from 13 more to 188 more)</td>
</tr>
<tr>
<td>38/142</td>
<td>24/141</td>
<td>RR 1.42 (1.02 to 1.99)</td>
<td>71 more per 1000 (from 3 more to 168 more)</td>
</tr>
<tr>
<td>5/129</td>
<td>3/100</td>
<td>RR 1.22 (0.3 to 4.92)</td>
<td>6 more per 1000 (from 20 fewer to 110 more)</td>
</tr>
</tbody>
</table>
GRADE Evidence Profiles

Should subcutaneous specific immunotherapy be used for treatment of allergic rhinitis in children without concomitant asthma?

**Author(s):** JLB & HJS  
**Date:** 2007-06-06

**Question:** Should subcutaneous immunotherapy be used in children with allergic rhinitis?

**Settings:** 

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of studies</td>
<td>No of patients</td>
</tr>
<tr>
<td><strong>nasal symptoms (follow-up 3 to 5 years; range of scores: 0-4; Better indicated by less)</strong></td>
<td></td>
</tr>
<tr>
<td>2 randomised trials</td>
<td>serious(^1)</td>
</tr>
<tr>
<td><strong>development of asthma (follow-up 5 years)</strong></td>
<td></td>
</tr>
<tr>
<td>1 randomised trial</td>
<td>serious(^3)</td>
</tr>
<tr>
<td><strong>non-life threatening systemic adverse events</strong></td>
<td></td>
</tr>
<tr>
<td>13 randomised trials</td>
<td>no serious limitations</td>
</tr>
<tr>
<td><strong>anaphylactic shock</strong></td>
<td></td>
</tr>
<tr>
<td>9 randomised trials</td>
<td>no serious limitations</td>
</tr>
<tr>
<td><strong>adrenaline use for systemic reaction</strong></td>
<td></td>
</tr>
<tr>
<td>13 randomised trials</td>
<td>no serious limitations</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 14,985 injections  
\(^11\) 1 event in 2276 injections
Self management for patients with chronic obstructive pulmonary disease

**Patient or population:** patients with chronic obstructive pulmonary disease  
**Settings:** primary care, community, outpatient  
**Intervention:** self management  
**Comparison:** usual care

| Outcomes | Illustrative comparative risks*  
(95% CI)  
Assumed risk  
usual care | Corresponding risk  
self management | Relative  
effect  
(95% CI) | No of Participants  
(studies) | Quality of  
evidence  
(GRADE) | Comments |
|---|---|---|---|---|---|---|
| Quality of Life  
St George's Respiratory Questionnaire.  
Scale from: 0 to 100.  
(follow-up: 3 to 12 months) | The mean quality of life ranged across control groups from 38 to 60 points  
The mean quality of life in the intervention groups was 2.58 lower  
(5.14 to 0.02 lower) | 698  
(7) | 450 moderate | Lower score indicates better quality of life. A change of less than 4 points is not shown to be important to patients. |
| Dyspnoea  
Borg Scale. Scale from: 0 to 10.  
(follow-up: 3 to 6 months) | The mean dyspnoea ranged across control groups from 1.2 to 4.1 points  
The mean dyspnoea in the intervention groups was 0.53 lower  
(0.96 to 0.1 lower) | 144  
(2) | 4 low | Lower score indicates improvement |
| Number and severity of exacerbations | See comment | See comment | 591  
(3) | See comment | Effect is uncertain |
| Respiratory-related hospital admissions  
(follow-up: 3 to 12 months) | Low risk population  
10 per 100  
(5 to 9) | OR 0.64  
(0.47 to 0.89) | 966  
(8) | 4 moderate | |
| High risk population  
50 per 100  
(32 to 47) | The mean emergency department visits for lung diseases in the intervention groups was 0.1 higher  
(0.2 lower to 0.3 higher) | 328  
(4) | 4 moderate | |
| Emergency department visits  
for lung diseases  
(follow-up: 6 to 12 months) | The mean doctor and nurse visits ranged across control groups from 0.2 to 0.7 visits per person per year | The mean doctor and nurse visits in the intervention groups was 0.02 higher  
(1 lower to 1 higher) | 629  
(5) | 4 moderate | |
| Doctor and nurse visits  
(follow-up: 8 to 12 months) | The mean doctor and nurse visits ranged across control groups from 1 to 5 visits per person per year | The mean doctor and nurse visits in the intervention groups was 0.02 higher  
(1 lower to 1 higher) | | |

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

CI: Confidence interval; OR: Odds ratio;
How to create a summary of findings table or evidence profile

- GRADEpro – software to create SoF
  - Available from Cochrane IMS website
- Import data from RevMan 5 into GRADEpro
- Create table – author makes suggestions about information to present and GRADEs the evidence
Creating a new GRADEpro file
Content

- Background about GRADE

- GRADE approach
  - Quality of evidence
  - Strength of recommendations
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
Quality of evidence & strength of recommendation

- Linked but no automatism
- Other factors beyond the quality of evidence influence our confidence that adherence to a recommendation causes more benefit than harm
- Systems/approaches failed to make this explicit
- GRADE separates quality of evidence from strength of recommendation
Developing recommendations

**Strength of Recommendations**

**Evaluating desirable and undesirable effects**

- Desirable << Undesirable effects
- Desirable ?< Undesirable effects
- Desirable ?> Undesirable effects
- Desirable >> Undesirable effects

**Formulating a recommendation**

- Against
  - Strong
  - Weak
- For
  - Weak
  - Strong

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.
Implications of a strong recommendation

- **Patients:** Most people in your 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/conditional recommendation

- **Patients**: The majority of people in your 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.

- **Policy makers**: There is a need for substantial debate and involvement of stakeholders.
Factors determining 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 the certainty for that benefit, the more likely is a weak recommendation.</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>
Current state of recommendations

The Yale Guideline Recommendation Corpus: A representative sample of the knowledge content of guidelines

Tamseela Hussain*, George Michel, Richard N. Shiffman

Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, United States
Current state of recommendations

- Reviewed 7527 recommendations
  - 1275 randomly selected
- Inconsistency across/within
- 31.6% did not recommendations clearly
  - Most of them not written as executable actions
- 52.7% did not indicated strength
Challenges in wording recommendations

- Need to express (two) levels
  - Strong vs weak/conditional
- Need to express direction
- Differences across languages

<table>
<thead>
<tr>
<th></th>
<th>Wording 1</th>
<th>Wording 2</th>
<th>Wording 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strong recommendation for</strong></td>
<td>We recommend...</td>
<td>Clinicians should...</td>
<td>We recommend...</td>
</tr>
<tr>
<td><strong>Weak recommendation for</strong></td>
<td>We suggest</td>
<td>Clinicians might...</td>
<td>We conditionally recommend...</td>
</tr>
<tr>
<td><strong>Weak recommendation against</strong></td>
<td>We suggest...not</td>
<td>Clinicians might not...</td>
<td>We conditionally recommend...not</td>
</tr>
<tr>
<td><strong>Strong recommendation against</strong></td>
<td>We recommend ...not</td>
<td>Clinicians should not...</td>
<td>We recommend ...not</td>
</tr>
</tbody>
</table>

- Need codes (letters, symbols, numbers)
### 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 consideration</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Severe symptoms of CMA (severe laryngeal edema, severe asthma, anaphylaxis) (follow-up 12 and 24 months)

- 2 randomised trials
- No of patients: 0/125 (0/117)
- Relative (95% CI): Not estimable
- Absolute: Not estimable
- Quality: ✽✽✽tempts
- Importance: CRITICAL

#### Allergic reaction to formula (follow-up 12 and 24 months)

- 2 randomised trials
- No of patients: 2/125 (1.6%) vs 0/125
- Relative (95% CI): RR 0.18 (0.05 to 0.71)
- Absolute: 91 fewer per 1000 (from 32 fewer to 106 fewer)
- Quality: ✽✽✽tempts
- Importance: CRITICAL

#### Moderate symptoms of CMA (mild laryngeal edema or mild asthma)

- 2 randomised trials
- No of patients: 0/125 (0/117)
- Relative (95% CI): Not estimable
- Absolute: Not estimable
- Quality: ✽✽✽tempts
- Importance: CRITICAL

#### Enteropathy or enteroctitis/proctocolitis (follow-up 12 and 24 months)

- 2 randomised trials
- No of patients: 0/125 (0/117)
- Relative (95% CI): Not estimable
- Absolute: Not estimable
- Quality: ✽✽✽tempts
- Importance: CRITICAL

#### Failure to thrive (measured as: length for age z-score) (follow-up 12 months; Better indicated by higher values)

- 1 randomised trials
- No of patients: 31 vs 32
- Relative (95% CI): MD 0.27 higher (0.19 lower to 0.73 higher)
- Quality: ✽✽✽tempts
- Importance: CRITICAL

#### Failure to thrive (measured as: weight for age z-score) (follow-up 12 months; Better indicated by higher values)

- 1 randomised trials
- No of patients: 31 vs 32
- Relative (95% CI): MD 0.23 higher (0.01 to 0.45 higher)
- Quality: ✽✽✽tempts
- Importance: CRITICAL

---

1. No events in both studies.
2. Allocation concealment was not reported and studies were not blinded. One study reported the results of per protocol analysis only.
3. The study to what extent a length for age z-score reflects a change in growth that would have an important consequence for a patient.
4. Only 62 children. There is uncertainty to what extent a weight for age z-score reflects a change in growth that would have an important consequence for a patient.
5. Allocation concealment was not reported and studies were not blinded. One study outcome was measured only in patients who developed symptoms.
6. One additional study (Salpietro 2005) included children with cow's milk allergy (23%) or intolerance and reported a relative risk of secondary sensitisation to extensively hydrolysed casein formula compared to soy formula of 1.33 (95% CI: 0.37 to 4.82).
7. It is uncertain how important is sensitization alone.
8. Only 11 events. Only 4 events.
9. Only 106 events.
Example recommendation from the cow milk allergy guidelines

In children with IgE-mediated cow’s milk allergy, we suggest extensively hydrolysed milk formula rather than soy formula (conditional recommendation | ⌧○○○○/very low quality evidence).

Underlying values and preferences

This recommendation places a relatively high value on avoiding adverse reactions to soy formula, and a relatively low value on an inferior acceptance of the extensively hydrolysed formula and resource utilization. In settings where relative importance of resource expenditure is lower an alternative choice may be equally reasonable.
Relevant healthcare 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

WHO Avian Influenza GL. 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
**Systematic review**

- Formulate question
- Select outcomes
- Rate importance
- Outcomes across studies
- Create evidence profile with GRADEpro
- Rate quality of evidence for each outcome

**Guideline development**

**Formulate recommendations:**
- For or against (direction)
- Strong or weak (strength)

**By considering:**
- Quality of evidence
- Balance benefits/harms
- Values and preferences

**Revise if necessary by considering:**
- Resource use (cost)

**Rate overall quality of evidence across outcomes based on lowest quality of critical outcomes**

- “We recommend using...”
- “We suggest using...”
- “We recommend against using...”
- “We suggest against using...”
GRADE – conclusions

- Do we need grading?
  - Impossible to train *all* practitioners in research methods and the related complexity quickly

- GRADE: Unifying system to evaluate quality of evidence & strength of recommendations

- Clear separation of 2 issues
  - Quality of evidence – 4 levels
  - Recommendations – conditional or strong

- Transparent

- Systematic by and across outcomes
Thank you!