Assessing quality of evidence
The GRADE approach

Elie A. Akl, MD, MPH, PhD

Department of Internal Medicine, AUB

Department of Clinical Epidemiology & Biostatistics
McMaster University, Canada
Disclosure

• Member of the GRADE working group

• No financial conflict of interest
This session

• First of two sessions concerning the use of GRADE methodology in developing recommendations for practice guidelines:

  – How to grade the quality of evidence
  
  – How to move from evidence to recommendations
Goals and Objectives

1. Define the quality of evidence

2. Discuss the factors that affect the rating of the quality of evidence reported in a systematic review
Grades of Recommendation Assessment, Development and Evaluation

GRADE

Working Group


GRADES OF RECOMMENDATION ASSESSMENT, DEVELOPMENT AND EVALUATION

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.
70+ Organizations
Goals and Objectives

1. Define the quality of evidence

2. Discuss the factors that affect the rating of the quality of evidence reported in a systematic review
What is quality of evidence?

• Extent to which we have confidence in the estimate of effect

• The quality is assessed by outcome (and not by study)
Example

- A new medication “X” is reported to reduce mortality from heart disease by 30%
“I figure there’s a 60% chance of heart disease and a 15% chance we know what we’re talking about.”
GRADE levels of Evidence

• The confidence in effect estimate is a continuum (from 0% to 100%)

• For practical reasons, it is categorized into 4 levels:
  - High:
  - Moderate:
  - Low:
  - Very low:
GRADE levels of Evidence

- **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
Goals and Objectives

1. Define the quality of evidence

2. Discuss the factors that affect the rating of the quality of evidence reported in a systematic review
Quality of evidence

• What are the factors that determine the rating of the quality of evidence reported in a systematic review?
### Determinants of quality

#### 1. Establish initial level of confidence

<table>
<thead>
<tr>
<th>Study design</th>
<th>Initial confidence in an estimate of effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trials ➔</td>
<td>High confidence</td>
</tr>
<tr>
<td>Observational studies ➔</td>
<td>Low confidence</td>
</tr>
</tbody>
</table>

#### 2. Consider lowering or raising level of confidence

<table>
<thead>
<tr>
<th>Reasons for considering lowering or raising confidence</th>
<th>↓ Lower if</th>
<th>↑ Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of Bias</td>
<td>Large effect</td>
<td></td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Dose response</td>
<td></td>
</tr>
<tr>
<td>Indirectness</td>
<td>All plausible confounding &amp; bias</td>
<td></td>
</tr>
<tr>
<td>Imprecision</td>
<td>• would reduce a demonstrated effect or</td>
<td></td>
</tr>
<tr>
<td>Publication bias</td>
<td>• would suggest a spurious effect if no effect was observed</td>
<td></td>
</tr>
</tbody>
</table>

#### 3. Final level of confidence rating

- **High** 🟢🟢🟢🟢🟢
- **Moderate** 🟢🟢🟢🟢
- **Low** 🟢🟢🟢
- **Very low** 🟢🟢
Factors that lower quality

1. Risk of bias
   - Inappropriate sequence generation
   - Lack of allocation concealment
   - Inadequate blinding
   - Intention to treat principle violation
   - Loss to follow-up
   - Early stopping for benefit
Randomization

• Why is randomization important?

• How do you ensure an adequate randomization process?
Randomization

• Adequate sequence generation
  – Table of random number
  – Computer-generated list of random number
  – Roll of a dice; flip of a coin

• Allocation concealment
  – Central randomization office
  – Pre-numbered or coded containers
  – Sequentially numbered, sealed, opaque envelopes
Blinding

• Who do you blind?

• Why do you blind them?
Blinding

• Patients
• Providers
• Data collectors
• Outcome adjudicators
• Data analysts
• Data and safety monitoring board (DSMB)
Analysis of missing outcome data

• Can patients with missing outcome data (lost to follow-up) introduce bias?

• How do you assess whether the number of participants with missing outcome data introduces bias?
Analysis of missing outcome data

200

100 → ASA
1 missing data
10 events

100 → Placebo
1 missing data
20 events

200

100 → ASA
20 missing data
10 events

100 → Placebo
20 missing data
20 events
Analysis of missing outcome data

200

100 → ASA

1 missing data

10 events

200

100 → Placebo

1 missing data

20 events

200

100 → ASA

1 missing data

1 event

200

100 → Placebo

1 missing data

2 event
Analysis of non-adherence

• Can handling of non-adherence patients in the analysis introduce bias?

• How do you avoid bias?
Analysis of non-adherence

- **Intention to treat analysis:**
  - analyze all those randomized in the arm to which they were randomized (irrespective of adherence)

- **As treated:**
  - analyze all those randomized according to what they actually received

- **Per protocol analysis:**
  - analyze only those who were adherent
Analyzing non-adherents

RR = (Risk on ASA) / (Risk on Placebo)

- ITT: RR =
- Per protocol: RR =
- As treated RR =
Analyzing non-adherents

RR = (Risk on ASA) / (Risk on Placebo)

- ITT: \[ RR = \frac{20/100}{20/100} = 1 \]
- Per protocol: \[ RR = \]
- As treated \[ RR = \]
Analyzing non-adherents

RR = (RISK on ASA) / (RISK on Placebo)

- ITT: RR = (20/100) / (20/100) = 1
- Per protocol: RR = (10/80) / (20/100) = 0.63
- As treated RR =
Analyzing non-adherents

\[ RR = \frac{\text{Risk on ASA}}{\text{Risk on Placebo}} \]

- **ITT:** \[ RR = \frac{20}{100} / \frac{20}{100} = 1 \]
- **Per protocol:** \[ RR = \frac{10}{80} / \frac{20}{100} = 0.63 \]
- **As treated:** \[ RR = \frac{10}{80} / \frac{30}{120} = 0.5 \]
Stopping study early for benefit

• Is there a risk of bias associated with stopping a study early?
## Studies stopped early because of benefit

<table>
<thead>
<tr>
<th>Timepoint</th>
<th>Deaths/Patients</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Five courses</td>
<td>Four courses</td>
<td>Statistics (O-E)</td>
<td>Var.</td>
<td>HR &amp; 95% CI</td>
<td>Odds Redn. (SD)</td>
</tr>
<tr>
<td>1997</td>
<td>7/102</td>
<td>15/100</td>
<td>-4.6</td>
<td>5.5</td>
<td>57% (29); 2P = 0.05</td>
<td></td>
</tr>
<tr>
<td>1998 (1)</td>
<td>23/171</td>
<td>42/169</td>
<td>-12.0</td>
<td>15.9</td>
<td>53% (18); 2P = 0.003</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 1.** Hazard ratio plot of mortality in the five versus four courses randomization in the MRC AML12 trials.
Factors that lower quality

### LMWH vs VKA

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>LMWH Events Total</th>
<th>VKA Events Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cesaroni 2003</td>
<td>2 96</td>
<td>3 96</td>
<td>0.8%</td>
<td>0.67 [0.11, 3.90]</td>
</tr>
<tr>
<td>Meyer 2002</td>
<td>22 71</td>
<td>29 75</td>
<td>12.0%</td>
<td>0.80 [0.51, 1.26]</td>
</tr>
<tr>
<td>Lee 2003</td>
<td>130 336</td>
<td>136 336</td>
<td>69.5%</td>
<td>0.96 [0.79, 1.15]</td>
</tr>
<tr>
<td>Deitcher 2006</td>
<td>22 67</td>
<td>11 34</td>
<td>6.9%</td>
<td>1.01 [0.56, 1.84]</td>
</tr>
<tr>
<td>Hull 2006</td>
<td>20 100</td>
<td>19 100</td>
<td>7.6%</td>
<td>1.05 [0.60, 1.85]</td>
</tr>
<tr>
<td>Lopez Beret 2001</td>
<td>7 17</td>
<td>6 18</td>
<td>3.2%</td>
<td>1.24 [0.52, 2.94]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>687</td>
<td>659</td>
<td>100.0%</td>
<td>0.95 [0.81, 1.11]</td>
</tr>
<tr>
<td>Total events</td>
<td>203</td>
<td>204</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

### ASA/NSAIDs use

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>ASA/NSAIDs use Events Total</th>
<th>No/occasional use Events Total</th>
<th>Weight</th>
<th>Odds Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson</td>
<td>10 6012</td>
<td>60 17277</td>
<td>12.4%</td>
<td>0.48 [0.24, 0.93]</td>
</tr>
<tr>
<td>Menezes</td>
<td>17 79</td>
<td>108 327</td>
<td>13.4%</td>
<td>0.56 [0.31, 1.00]</td>
</tr>
<tr>
<td>Ratnasinghe</td>
<td>43 14838</td>
<td>35 7996</td>
<td>14.8%</td>
<td>0.66 [0.42, 1.03]</td>
</tr>
<tr>
<td>Jacobs</td>
<td>37 7769</td>
<td>3455 721041</td>
<td>16.1%</td>
<td>0.99 [0.72, 1.38]</td>
</tr>
<tr>
<td>Coogan</td>
<td>18 188</td>
<td>207 2339</td>
<td>14.2%</td>
<td>1.09 [0.66, 1.81]</td>
</tr>
<tr>
<td>Schernhammer</td>
<td>37 10292</td>
<td>153 89541</td>
<td>15.7%</td>
<td>2.11 [1.47, 3.02]</td>
</tr>
<tr>
<td>Langman</td>
<td>25 48</td>
<td>413 1286</td>
<td>13.4%</td>
<td>2.30 [1.29, 4.10]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>39226</td>
<td>839807</td>
<td>100.0%</td>
<td>1.01 [0.65, 1.55]</td>
</tr>
<tr>
<td>Total events</td>
<td>187</td>
<td>4431</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.28; Chi² = 35.73, df = 6 (P < 0.00001); I² = 83%
Test for overall effect: Z = 0.04 (P = 0.97)
Factors that lower quality

2. Inconsistency
   - Assess for inconsistency (Heterogeneity)

   - If inconsistency → look for explanation
     - patients, intervention, outcome, methods

   - If unexplained inconsistency → downgrade quality
Factors that lower quality

3. Indirectness of Evidence

- Differences in populations/patients
- Differences in interventions
- Differences in outcomes
Factors that lower quality

4. Imprecision

- If 95% CI includes both negligible effect and appreciable benefit or appreciable harm → rate down
Factors that lower quality

5. Publication bias

- Faster and multiple publication of “positive” trials
- Fewer and slower publication of “negative” trials
Funnel plot

Asymmetrical: Publication bias?
Meta-analysis contradicted by mega-trials

Publication bias
Factors that raise quality

1. Large magnitude of effect

2. Dose response relation

3. All plausible confounding may be working to reduce the demonstrated effect
Factors that raise quality

1. Large magnitude of effect
   – large (RRR 50%) can raise by one level
   – very large (RRR 80%) can raise by two levels

   – common criteria: everyone used to do badly and now almost everyone does well.
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

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials.
Relative risk reduction:
....> 99.9 % (1/100,000)
U.S. Parachute Association reported 821 injuries and 18 deaths out of 2.2 million jumps in 2007
Factors that 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%)
Determinants of quality

<table>
<thead>
<tr>
<th>Study design</th>
<th>Initial confidence in an estimate of effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trials</td>
<td>High confidence</td>
</tr>
<tr>
<td>Observational studies</td>
<td>Low confidence</td>
</tr>
</tbody>
</table>

1. Establish initial level of confidence

2. Consider lowering or raising level of confidence

   - **Lower if**
     - Risk of Bias
     - Inconsistency
     - Indirectness
     - Imprecision
     - Publication bias

   - **Higher if**
     - Large effect
     - Dose response
     - All plausible confounding & bias
       - would reduce a demonstrated effect
       - would suggest a spurious effect if no effect was observed

3. Final level of confidence rating

   - Confidence in an estimate of effect across those considerations
     - High
     - Moderate
     - Low
     - Very low
Questions?
Exercise

• Grade the quality of evidence for a systematic review on the use of parenteral anticoagulation in patients with cancer