An introduction to the GRADE approach in systematic reviews and guideline development

Associate Professor Zachary Munn
HSRAANZ Seminar
Declarations of Interest

• Employee of the Joanna Briggs Institute
• Member of the GRADE Working group
• Associate Editor for BMC Medical Research Methodology
• Editorial board for the International Journal of Evidence-based Healthcare
• No financial conflicts of interest

Zachary.Munn@adelaide.edu.au  @ZacMunn @JBIEBHC
Objectives

1. To introduce participants to the work of the GRADE working group
2. To explain the rationale behind the GRADE approach
3. To explain the key factors to consider when assessing certainty of the evidence
4. To provide advice about additional resources for guidance following the webinar
Introduction to GRADE
History of GRADE

• Began as an informal working group in 2000, largely out of McMaster University
• Informal collaboration of researchers/guideline developers with interest in methodology
• Purpose: to develop a common system for grading the quality (certainty) of evidence and the strength of recommendations that is transparent and sensible
• Website: http://www.gradeworkinggroup.org/
Over 100 organisations
From 19 countries
In Australia – Systematic Reviewers

- JBI and Cochrane explicitly endorse the use of GRADE methods and require GRADE
In Australia – Guideline Developers

Methods for rating the quality of evidence

NHMRC uses GRADE (Grading of Recommendations, Assessment, Development and Evaluation) in the development of its guidelines. GRADE is an internationally recognised approach to rate the quality of evidence and the strength of recommendations and is considered to be the standard in guideline development.

Guidelines developers seeking NHMRC approval of their guidelines may use either GRADE or the earlier NHMRC levels of evidence and grades for recommendations for developers of guidelines (2009) for rating the quality of evidence. These options are documented in the Procedures and requirements for meeting the 2011 NHMRC standard for clinical practice guidelines.

NHMRC Evidence Statement Form

It should be noted that NHMRC has no plans to update or maintain the NHMRC levels of evidence and grades for recommendations for developers of clinical practice guidelines (2009).
6. To be **evidence informed** guidelines will:
   6.1. Be informed by well conducted systematic reviews
   6.2. Consider the body of evidence for each outcome (including the quality of that evidence) and other factors that influence the process of making recommendations including benefits and harms, values and preferences, resource use and acceptability.
   6.3. Be subjected to appropriate peer review.

3. The guideline development group will:
   3.1. Be composed of an appropriate mix of expertise and experience, including relevant end users
   3.2. Have clearly defined, documented processes for reaching consensus.

4. To identify and manage conflicts of interest guideline developers will:
   4.1. Require all interests of all guideline development group members to be declared
   4.2. Establish a process for determining if a declared interest represents a conflict of interest, and how a conflict of interest will be managed.

8. To be up-to-date guidelines will:
   8.1. Ensure that the recommendation is based on an up-to-date body of evidence
   8.2. Propose a date by which the evidence and the guideline should be updated. This may be specific to each recommendation.

9. To be accessible guidelines will:
   9.1. Be easy to find
   9.2. Ideally be free of charge to the end user
   9.3. Be clearly structured, easy to navigate and in plain English
   9.4. Be available online.
1. **Train, promote, disseminate and implement GRADE within ANZ and the JBC**

2. **Act as a central hub for GRADE in Oceania**

3. **Contribute to GRADE methods**

Pictured: JBI Adelaide GRADE Center Director Associate Professor Zachary Munn (centre) with GRADE Working Group co-chairs Professor Holger Schünemann (left) and Distinguished Professor Gordon Guyatt (right)

http://grade.joannabriggs.org/
Why GRADE?
GOBSAT Method

• ‘Good old boys sat around the table’
• Initial approach to development of recommendations within guidelines
• Based on expert opinion, powerful figures, eminence based medicine
Levels of Evidence

- Designate study types
- Better study designs, with greater methodological quality, are ranked higher
- Assigned to findings of research

Grades of Recommendation

- Assist in applying research into practice
- Recommendations assigned a ‘Grade’
‘The first hierarchy of evidence quality was created, where evidence of the highest quality would have to come from at least one randomized trial, and at the bottom of that hierarchy of evidence were opinions of respected experts without any empirical evidence. That seems really simple in retrospect, but, actually, it was an incredible breakthrough to address the way we dealt with the large amount of available research evidence. It made it feasible to sift through evidence in a meaningful way and apply the principles of using the best-quality and least-biased evidence.’ Paul Glasziou

‘Eventually, the traditional hierarchies of evidence started to fall apart due to attempts to fit too many elements as well as a lack of standardization. Now, we have to move on to a new phase of trying to unify the principles’

### Levels of Evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia</td>
<td>meta-analysis or randomised control trials (RCT)</td>
</tr>
<tr>
<td>Ib</td>
<td>at least one randomised control trial (RCT)</td>
</tr>
<tr>
<td>IIa</td>
<td>at least one well designed control study without randomisation</td>
</tr>
<tr>
<td>IIb</td>
<td>at least one other type of well designed quasi-experimental study</td>
</tr>
<tr>
<td>III</td>
<td>well designed non-experimental descriptive studies (e.g., comparative correlation or case control studies)</td>
</tr>
<tr>
<td>IV</td>
<td>expert committee reports or opinions and/or clinical experience of respected authorities</td>
</tr>
</tbody>
</table>

### Grading of Recommendations

<table>
<thead>
<tr>
<th>Grade</th>
<th>Evidence Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ia/Ib</td>
<td>Requires at least one RCT as part of good quality literature consistently addressing specific recommendations</td>
</tr>
<tr>
<td>B</td>
<td>IIa/IIb/III</td>
<td>Requires well conducted clinical studies short of RCT on topic of recommendation</td>
</tr>
<tr>
<td>C</td>
<td>IV</td>
<td>Requires expert committee reports/opinions or clinical experience of respected authorities. Indicates absence of directly applicable good quality studies</td>
</tr>
</tbody>
</table>
1.3.1 LIFESTYLE ISSUES

Smoking should be discouraged.

2.1 SMOKING

Cigarette smoking is thought to be a significant risk factor for many cancers, including bladder cancer, and is likely to be detrimental to the general health of patients recovering from cancer. Cigarette smoke also has been shown to increase the risk of bladder cancer. Smoking increases the risk of bladder cancer by two to three times. 2,3,6,10-12 and stops smoking leads to a remarkable, although not statistically significant, reduction of bladder cancer risk (odds ratio, OR = 0.66; 95% confidence interval, CI 0.38-1.2). 7

While the direct relationship between continued smoking in patients with bladder cancer and a second non-bladder cancer is inadequately studied, it is recognized that stopping smoking while recovering reduces the risk of complications after surgery. 4

- Smoking should be discouraged.
- Healthcare professionals should refer smokers with bladder cancer to smoking cessation services.
Why GRADE?

<table>
<thead>
<tr>
<th>Level</th>
<th>Adjustment</th>
<th>Description by Type of Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>SR (with homogeneity) of prospective cohort study</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>Prospective cohort study with good follow-up</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>All or none case-series</td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td>SR (with homogeneity) of 2b and better studies</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>Retrospective cohort study, or poor follow-up</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>Ecological studies</td>
</tr>
<tr>
<td>3</td>
<td>a</td>
<td>SR (with homogeneity) of 3b and better studies</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>Non-consecutive cohort study, or very limited population</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Case-series or superseded reference standards</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
</tr>
</tbody>
</table>


---

### Table 1 – Classification of the procedures

<table>
<thead>
<tr>
<th>Class</th>
<th>Conditions for which conclusive evidence exists, or in its absence general consensus that the procedure is useful or effective, or both.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td></td>
</tr>
<tr>
<td>Class II A</td>
<td>Weight or evidence/opinion favoring usefulness/efficacy.</td>
</tr>
<tr>
<td>Class II B</td>
<td>Usefulness/efficacy less well established by evidence or opinion.</td>
</tr>
<tr>
<td>Class III</td>
<td>Conditions for which evidence or consensus, or both, exists that the procedure is not useful/efficient, and, in some cases, it may even be toxic.</td>
</tr>
<tr>
<td></td>
<td>Adapted from the criteria used in the guidelines of the American College of Cardiology/American Heart Association.</td>
</tr>
</tbody>
</table>

### Table 2 – American Heart Association Classification Of Recommendations And Level Of Evidence

<table>
<thead>
<tr>
<th>Levels of Evidence</th>
<th>Classes of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level A</td>
<td>Class I Benefit &gt;&gt;&gt; Risk</td>
</tr>
<tr>
<td>Level B</td>
<td>Class IIa Benefit &gt;&gt; Risk</td>
</tr>
<tr>
<td>Level C</td>
<td>Class IIb Benefit ≥ Risk</td>
</tr>
<tr>
<td>Level D</td>
<td>Class III Risk ≥ Benefit</td>
</tr>
</tbody>
</table>

- Rich body of quality RCT
- Limited body of data or high non-RCT bias
- Limited evidence
- No evidence or panel consensus judgmental

---

Table: Grade of recommendation and levels of evidence.

<table>
<thead>
<tr>
<th>n</th>
<th>Levels of evidence</th>
<th>Types of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Systematic review of homogeneous RCTs with good methodological quality</td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>Individual RCTs with narrow confidence intervals</td>
<td></td>
</tr>
<tr>
<td>1c</td>
<td>Uncontrolled studies (dramatic findings)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Systematic review of cohort studies (with homogeneity)</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>Individual cohort studies (including low quality RCTs, e.g. &lt;80% follow-up)</td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>Uncontrolled cohort studies/ecological studies</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Systematic review of non-randomized or controlled studies</td>
<td></td>
</tr>
</tbody>
</table>

---

Table 1 – Classification of the procedures

<table>
<thead>
<tr>
<th>Class</th>
<th>Conditions for which conclusive evidence exists, or in its absence general consensus that the procedure is useful or effective, or both.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td></td>
</tr>
<tr>
<td>Class II A</td>
<td>Weight or evidence/opinion favoring usefulness/efficacy.</td>
</tr>
<tr>
<td>Class II B</td>
<td>Usefulness/efficacy less well established by evidence or opinion.</td>
</tr>
<tr>
<td>Class III</td>
<td>Conditions for which evidence or consensus, or both, exists that the procedure is not useful/efficient, and, in some cases, it may even be toxic.</td>
</tr>
<tr>
<td></td>
<td>Adapted from the criteria used in the guidelines of the American College of Cardiology/American Heart Association.</td>
</tr>
</tbody>
</table>
Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group


Received: 23 January 2004 | Accepted: 22 December 2004 | Published: 22 December 2004

Open Peer Review reports

Abstract

Background

A number of approaches have been used to grade levels of evidence and the strength of recommendations. The use of many different approaches detracts from one of the main reasons for having explicit approaches: to concisely characterise and communicate this information so that it can easily be understood and thereby help people make well-informed decisions. Our objective was to critically appraise six prominent systems for grading levels of evidence and the strength of recommendations as a basis for agreeing on characteristics of a common, sensible approach to grading levels of evidence and the strength of recommendations.
Forming recommendations with GRADE

Balance between benefits, harms and burdens

Certainty of Evidence

Patients values and preferences
Equity

Resource use
Feasibility

How do we determine certainty of the evidence?
Key principle

• Important to communicate
  • Results
  • Our certainty in these results?
Magnitude of Effect (results)

Certainty/quality/confidence in the evidence

Figure 1. Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).
Our certainty in the evidence

• If not by study design:
  • How can we ascertain the ‘quality’ of the evidence?
  • What impacts our ‘confidence’ regarding the evidence?
Example meta-analysis discussion

• From the example provided, what information would increase or decrease your confidence in these results?
GRADE factors affecting certainty

• Decrease
  • Methodological limitations (risk of bias)
  • Indirectness
  • Inconsistency (heterogeneity)
  • Imprecision
  • Publication bias

• Increase
  • Large, consistent, precise effect
  • All plausible biases underestimate the effect
  • Dose response effect
Determining certainty of the evidence
GRADEing the evidence

• Pre-ranking
  • Evidence from RCTs start as high, Observational studies as low

• Certainty of evidence ranges from
  • High
  • Moderate
  • Low
  • Very low

• Can be downgraded 1 or 2 points for each area of concern

• Maximum downgrade of 3 points overall
Certainty in the evidence varies from:

- **HIGH**
- **MODERATE**
- **LOW**
- **VERY LOW**

**RCT**
- Risk of bias
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

**NRS**
- Dose-response
- Large effect
- Plausible confounding

---

**GRADE**

[The Joanna Briggs Institute]

**Better evidence. Better outcomes.**
What does this mean?

• **High quality**: We are very confident that the true effect lies close to that of the estimate of the effect

• **Moderate quality**: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

• **Low quality**: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

• **Very low quality**: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect
Factors that can decrease our certainty
1. Methodological limitations

• High risk of bias
  • No allocation concealment
  • Lack of blinding
  • Attrition bias
  • Selective reporting

• Determined by results of risk of bias assessment
### Addressing Bias

<table>
<thead>
<tr>
<th>Type of bias</th>
<th>Method to reduce bias</th>
<th>When and whom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection</td>
<td>Randomization</td>
<td>Patients, trial coordinators/investigators and allocators during the process of screening for inclusion and allocation to groups</td>
</tr>
<tr>
<td></td>
<td>Allocation concealment</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Blinding</td>
<td>Trial participants and those delivering the intervention throughout the trial period</td>
</tr>
<tr>
<td>Detection</td>
<td>Blinding</td>
<td>The participant (if self-reported outcomes) or those assessing outcomes at the time of outcome assessment</td>
</tr>
<tr>
<td>Attrition</td>
<td>Complete follow-up</td>
<td>Trial investigators collecting and analysing data</td>
</tr>
<tr>
<td></td>
<td>Intention-to-treat</td>
<td></td>
</tr>
<tr>
<td></td>
<td>analysis</td>
<td></td>
</tr>
<tr>
<td>Reporting</td>
<td>Comprehensive and full reporting of all outcomes/data</td>
<td>Trial investigators and authors following the trial</td>
</tr>
</tbody>
</table>
Steps to assess risk of bias

• Assess the risk of bias for each study providing data for an outcome
• Use tools appropriate to the question and study design
  • RCTs – Cochrane Risk of Bias Tool
  • Non-randomised studies – ROBINS-I (Cochrane), NewCastle-Ottawa
  • Diagnostic studies – QUADAS
  • Prognostic studies - QUIPS

• Consider the risk of bias across all studies providing data for an outcome, decide whether:
  • No concern (do not downgrade)
  • Serious concern (consider downgrade of 1 level)
  • Very serious concern (consider downgrade of two levels)
2. Inconsistency of results (unexplained heterogeneity)

- Widely differing estimates of treatment effect
- If inconsistency exists, look for explanation
  - Patients, intervention, comparator, outcome
- If unexplained inconsistency lower quality
Identifying heterogeneity

• Heterogeneity can be determined by:
  • Wide variance of point estimates
  • Minimal or no overlap of confidence intervals
  • Statistical tests
    • standard chi-squared test (Cochran Q test)
    • I square statistic (I2)
Example Forest Plot

![Forest Plot Diagram]

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Events</th>
<th>Control Events</th>
<th>Total Weight</th>
<th>Risk Ratio M-H, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaida 2012</td>
<td>20</td>
<td>13</td>
<td>8.6%</td>
<td>1.74 [0.88, 3.46]</td>
</tr>
<tr>
<td>Chian 2006</td>
<td>12</td>
<td>24</td>
<td>16.7%</td>
<td>0.51 [0.26, 1.01]</td>
</tr>
<tr>
<td>Dosi 2012</td>
<td>8</td>
<td>1</td>
<td>0.7%</td>
<td>9.09 [1.23, 67.06]</td>
</tr>
<tr>
<td>Gongh 2007</td>
<td>8</td>
<td>186</td>
<td>11.2%</td>
<td>0.50 [0.22, 1.15]</td>
</tr>
<tr>
<td>Hoedemaeker 2006</td>
<td>50</td>
<td>240</td>
<td>25.1%</td>
<td>1.42 [0.96, 2.09]</td>
</tr>
<tr>
<td>Lazur 2011</td>
<td>1</td>
<td>15</td>
<td>9.6%</td>
<td>0.06 [0.01, 0.50]</td>
</tr>
<tr>
<td>Rujinojndakul 2014</td>
<td>1</td>
<td>99</td>
<td>28.0%</td>
<td>0.33 [0.06, 1.68]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1601</strong></td>
<td><strong>1701</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>0.72 [0.57, 0.92]</strong></td>
</tr>
</tbody>
</table>

Total events: 100

Heterogeneity: Chi^2 = 41.59, df = 6 (P = 0.00001), I^2 = 86%

Test for overall effect: Z = 2.82, p = 0.005

---

The Joanna Briggs Institute
3. Indirectness of evidence

- Direct evidence ‘directly compares the interventions which we are interested in, delivered to the populations in which we are interested, and measures the outcomes important to patients’ (GRADE Handbook)

- Assessed by:
  - Applicability of the results (are the populations/interventions studied those we are interested in?)
  - Surrogate outcomes
  - Indirect comparisons
4. Imprecision

- Imprecise results
- Assesses
  - Are the confidence intervals wide?
  - Are there only few events?
  - Is there a small sample size?
- Optimal information size
- Imprecision is one of the more complex factors to consider – refer to GRADE Handbook for more details
- Different for SRs vs Guidelines
  - Guidelines contextualized for decision making and recommendations
  - SRs free of this context
Would you rate down?
5. Publication Bias

- Publication bias occurs when the published studies differ systematically from all conducted studies on a topic.
- It is a serious threat to the validity of systematic reviews and meta-analyses.
- Should always be suspected:
  - Only small “positive” studies
  - For profit interest
  - Various methods to evaluate – none perfect, but clearly a problem

Taken from: Sterne et al 2005
Factors that raise quality
Raising the quality

• Initially classified as low, a body of evidence from observational studies can be rated up.

• Consideration of factors reducing quality of evidence must precede consideration of reasons for rating it up.

• 5 factors for rating down quality of evidence must be rated prior to the 3 factors for rating it up.

• The decision to rate up quality of evidence should only be made when serious limitations in any of the 5 areas reducing the quality of evidence are absent.
1. Large magnitude of an effect

- Large, consistent, precise effect
- Although observational studies may overestimate the effect, bias is unlikely to explain or contribute all for a reported very large benefit (or harm)
- What is large?
  - RR of 2 (large), 5 (very large)
  - For example, odds ratio of babies sleeping on stomachs of 4.1 (95% CI of 3.1 to 5.5) for SIDS compared to sleeping on their back
  - Parachutes to prevent death when jumping from airplanes
  - May upgrade 1 level for large and 2 for very large
2. Dose-response gradient

• Dose-response gradient
  • Clear dose-response indicative of a cause-effect relationship
    • Warfarin and bleeding (clear dose response)
    • Delay in antibiotics for those presenting with sepsis (i.e. each hour delayed increases mortality)
    • Delay in removal of indwelling urinary catheter and development of UTI
3. Effect of plausible residual confounding

• Rigorous observational studies adjust/address confounding in their analysis for identified confounders
• Cannot control for ‘unmeasured or unknown’ confounders (hence why observational studies are downgraded), and other plausible confounders may not be addressed
• This ‘residual’ confounding may result in an underestimation of the true effect
• All plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
  • Sicker patients doing better
  • Not for profit vs for profit
Summary of findings
tables and evidence profiles
Evidence profiles and Summary of Findings tables

• Endpoint of the GRADE process for SRs
• Key milestone for Guideline developers on their way to make a recommendation
• Evidence profiles include outcomes, number of studies, all judgements regarding GRADE factors, assumed risk, corresponding risk, relative effect, absolute effect, overall rating, classification of outcome importance, footnotes
• SoF table includes most of the above but not all GRADE factor judgements
<table>
<thead>
<tr>
<th>No of studies (Design)</th>
<th>Limitations</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Placebo</th>
<th>Antibiotics</th>
<th>Relative risk (95% CI)</th>
<th>Control risk</th>
<th>Risk difference (95% CI)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain at 24h 5 (RCT)</td>
<td>No serious limitations</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>241/605</td>
<td>223/624</td>
<td>RR 0.9 (0.78–1.04)</td>
<td>367/1,000</td>
<td>Not Significant</td>
<td>High</td>
</tr>
<tr>
<td>Pain at 2–7 d 10 (RCT)</td>
<td>No serious limitations</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>303/1,366</td>
<td>238/1,425</td>
<td>RR 0.72 (0.62–0.81)</td>
<td>257/1,000</td>
<td>72 fewer per 1,000 (44–98)</td>
<td>High</td>
</tr>
<tr>
<td>Hearing, inferred from the surrogate outcome abnormal tympanometry—1 mo 4 (RCT)</td>
<td>No serious limitations</td>
<td>No serious inconsistency</td>
<td>Serious indirectness (because of indirectness of outcome)</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>168/460</td>
<td>153/467</td>
<td>RR 0.89 (0.75–1.07)</td>
<td>353/1,000</td>
<td>Not Significant</td>
<td>Moderate</td>
</tr>
<tr>
<td>Hearing, inferred from the surrogate outcome abnormal tympanometry—3 mo 3 (RCT)</td>
<td>No serious limitations</td>
<td>No serious inconsistency</td>
<td>Serious indirectness (because of indirectness of outcome)</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>96/398</td>
<td>96/410</td>
<td>RR 0.97 (0.76–1.24)</td>
<td>234/1,000</td>
<td>Not Significant</td>
<td>Moderate</td>
</tr>
<tr>
<td>Vomiting, diarrhea, or rash 5 (RCT)</td>
<td>No serious limitations</td>
<td>Serious inconsistency (because of inconsistency in absolute effects)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>83/711</td>
<td>110/690</td>
<td>RR 1.38 (1.09–1.76)</td>
<td>113/1,000</td>
<td>43 more per 1,000 (10–86)</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

**Abbreviations:** GRADE, Grading of Recommendations Assessment, Development, and Evaluation; RCT, randomized controlled trials; CI, confidence interval; RR, risk ratio.

* The control rate is based on the median control group risk across studies.
Summary of Findings tables

• Standard table format
  • one for each comparison (may require more than one)
  • Report all outcomes, even if no data

• Improve understanding

• Improve accessibility

• Created with GRADEpro GDT

http://www.guidelinedevelopment.org/
### Summary of findings table

**Summary of findings for the main comparison**

<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>Incidence of Diarrhea</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probiotic dose (equal to/greater than) 5 billion CFU/day</td>
<td>223 per 1000 (65 to 122)</td>
<td>RR 0.4 (0.29 to 0.55)</td>
<td>1474 (7 studies)</td>
<td><strong>⊕⊕⊙</strong> low¹⁻²</td>
<td>Control group risk estimates come from control arm of meta-analysis, based on included trials. Relative effect based on available case analysis</td>
</tr>
<tr>
<td>Follow-up: 10 days to 3 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adverse Events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up: 10 to 44 days</td>
<td>18 per 1000 (8 to 38)</td>
<td>See comment</td>
<td>1575 (11 studies)</td>
<td><strong>⊕⊕⊙</strong> low¹⁻⁴</td>
<td>Risks were calculated from pooled risk differences. Control group risk estimates come from control arm of the meta-analysis, based on included studies</td>
</tr>
<tr>
<td><strong>Duration of Diarrhea</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up: 10 days to 3 months</td>
<td>The mean duration of diarrhea in the intervention groups was 0.6 lower (1.18 to 0.02 lower)</td>
<td></td>
<td>897 (5 studies)</td>
<td><strong>⊕⊕⊙</strong> low³⁻⁶</td>
<td></td>
</tr>
</tbody>
</table>
Making Recommendations
Forming recommendations with GRADE

Balance between benefits, harms and burdens

Certainty of Evidence

Patients values and preferences

Equity

Resource use

Feasibility

University of Adelaide
When making decisions...

Guideline members use their expertise to weigh all criteria to make a recommendation

- Balance of benefits and harms
- Consideration of patient values and preferences
- Consideration of resources, feasibility, equity, and acceptability
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)
  • Strong for
  • Weak for
  • Strong against
  • Weak against
<table>
<thead>
<tr>
<th></th>
<th>Strong Recommendation</th>
<th>Weak Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>For patients</td>
<td>Most individuals in this situation would want the recommended course of action and only a small proportion would not.</td>
<td>The majority of individuals in this situation would want the suggested course of action, but many would not.</td>
</tr>
<tr>
<td>For clinicians</td>
<td>Most individuals should receive the recommended course of action. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences.</td>
<td>Recognize that different choices will be appropriate for different patients, and that you must help each patient arrive at a management decision consistent with her or his values and preferences. Decision aids may well be useful helping individuals making decisions consistent with their values and preferences. Clinicians should expect to spend more time with patients when working towards a decision.</td>
</tr>
<tr>
<td>For policy makers</td>
<td>The recommendation can be adapted as policy in most situations including for the use as performance indicators.</td>
<td>Policy making will require substantial debates and involvement of many stakeholders. Policies are also more likely to vary between regions. Performance indicators would have to focus on the fact that adequate deliberation about the management options has taken place.</td>
</tr>
</tbody>
</table>
Guideline panels

“I absolutely disagree!”
“Let’s just keep it how it has always been”
“What about that one new study?”
“Have we decided yet?”
Evidence to decision framework

• Inform panel members’ judgements about the pros and cons of each option (intervention) that is considered
• Ensure that important factors that determine a recommendation (criteria) are considered
• Provide a concise summary of the best available research evidence to inform judgements about each criterion
• Help structure discussion and identify reasons for disagreements
• Make the basis for recommendations transparent to guideline users
Decision making criteria

• Priority of problem
• Benefits and harms
• Certainty of evidence
• Values and Preferences
• Resources
• Equity
• Acceptability
• Feasibility
**Question**

Should **Tight glycemic control vs. placebo** be used for hospital?

<table>
<thead>
<tr>
<th>POPULATION:</th>
<th>hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERVENTION:</td>
<td>Tight glycemic control</td>
</tr>
<tr>
<td>COMPARISON:</td>
<td>placebo</td>
</tr>
<tr>
<td>MAIN OUTCOMES:</td>
<td>Mortality; Mortality - Nondiabetic patients; Mortality - Diabetic patients; Atrial fibrillation; Stroke; Acute renal failure; Deep sternal infection; Length of stay;</td>
</tr>
<tr>
<td>SETTING:</td>
<td></td>
</tr>
<tr>
<td>PERSPECTIVE:</td>
<td></td>
</tr>
</tbody>
</table>

**Assessment**

<table>
<thead>
<tr>
<th>JUDGEMENT</th>
<th>RESEARCH EVIDENCE</th>
<th>ADDITIONAL CONSIDERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the problem a priority?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Probably no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Probably yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Varies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Don’t know</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How substantial are the desirable anticipated effects?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Trivial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Moderate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of judgements

<table>
<thead>
<tr>
<th>JUDGEMENT</th>
<th>IMPPLICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROBLEM</td>
<td></td>
</tr>
<tr>
<td>DESIRABLE EFFECTS</td>
<td></td>
</tr>
<tr>
<td>UNDESIRABLE EFFECTS</td>
<td>MEANING</td>
</tr>
<tr>
<td>CERTAINTY OF EVIDENCE</td>
<td></td>
</tr>
<tr>
<td>VALUES</td>
<td>MEANING</td>
</tr>
<tr>
<td>BALANCE OF EFFECTS</td>
<td></td>
</tr>
<tr>
<td>ACCEPTABILITY</td>
<td></td>
</tr>
<tr>
<td>FEASIBILITY</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

Should Tight glycemic control vs. placebo be used in hospital?

<table>
<thead>
<tr>
<th>TYPE OF RECOMMENDATION</th>
<th>Strong recommendation against the intervention</th>
<th>Conditional recommendation against the intervention</th>
<th>Conditional recommendation for either the intervention or the comparison</th>
<th>Conditional recommendation for the intervention</th>
<th>Strong recommendation for the intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECOMMENDATION</td>
<td>O</td>
<td>D</td>
<td>O</td>
<td>O</td>
<td>D</td>
</tr>
</tbody>
</table>

JUSTIFICATION

SUBGROUP CONSIDERATIONS
Conclusion
Summing up: So why GRADE?

1. Transparent approach to rating certainty
2. Separation between certainty of evidence and strength of recommendation
3. Considers issues other than study design
4. Focuses on outcomes, not studies
5. Clear guidance for developing and establishing recommendations
6. Supported and endorsed by the international systematic review and guideline development community
“GRADE is much more than a rating system. It offers a transparent and structured process for developing and presenting evidence summaries for systematic reviews and guidelines in health care and for carrying out the steps involved in developing recommendations. GRADE specifies an approach to framing questions, choosing outcomes of interest and rating their importance, evaluating the evidence, and incorporating evidence with considerations of values and preferences of patients and society to arrive at recommendations. Furthermore, it provides clinicians and patients with a guide to using those recommendations in clinical practice and policy makers with a guide to their use in health policy.”  Guyatt et al 2011
Other resources/ Information

• Diagnostic test accuracy SoF tables
• Qualitative evidence synthesis  GRADE Approach – CerQual
• GRADE Handbook (http://www.guidelinedevelopment.org/handbook/)
• GIN-McMaster Guidelines checklist (http://cebgrade.mcmaster.ca/guidecheck.html)
• MAGIC App
• JBI Adelaide GRADE Workshops http://grade.joannabriggs.org/
Get involved!

GRADE WORKING GROUP NEWSLETTER SIGNUP

Please sign up to receive news on activities of the GRADE working group

email address

Subscribe

GET IN TOUCH

For inquiries, feel free to contact us through your preferred social media channel or through email.

CONTACT
mail@gradeworkinggroup.org

About Us

© 2004-2017 - The GRADE working group -

FOLLOW US

jbi@gradeworkinggroup.org