Rating quality evidence, grading strength recommendation

• informative summaries for consumer

• eminently useful, but proliferation
  - Australian National and MRC
  - Oxford Center for Evidence-based Medicine
  - Scottish Intercollegiate Guidelines (SIGN)
  - US Preventative Services Task Force
  - American professional organizations
    • AHA/ACC, ACCP, AAP, Endocrine society, etc....

• cause of confusion, dismay
A common international grading system?

• international group
  - methodologists, guideline developers
  - Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC

• GRADE (Grades of recommendation, assessment, development and evaluation)

• ~ 25 meetings over last 10 years
  - (~10 – 60 attendants)
GRADE Uptake

Agencia sanitaria regionale, Bologna, Italia
Agency for Health Care Research and Quality (AHRQ)
Allergic Rhinitis and Group - Independent Expert Panel
American College of Cardiology Foundation
American College of Chest Physicians
American College of Emergency Physicians
American College of Physicians
American Endocrine Society
American Society of Gastrointestinal Endoscopy
American Society of Interventional Pain Physicians
American Thoracic Society (ATS)
BMJ Clinical Evidence
British Medical Journal
Canadian Agency for Drugs and Technology in Health Centers for Disease Control
Cochrane Collaboration
EBM Guidelines Finland
Emergency Medical Services for Children National Resource Center
European Association for the Study of the Liver
European Respiratory Society
European Society of Thoracic Surgeons
Evidence-based Nursing Sudtirol, Alta Adiga, Italy
Finnish Office of Health Technology Assessment
German Agency for Quality in Medicine
Infectious Disease Society of America
Japanese Society of Oral and Maxillofacial Radiology
Joslin Diabetes Center
Journal of Infection in Developing Countries
Kidney Disease International Guidelines Organization
National and Gulf Centre for Evidence-based Medicine
National Institute for Clinical Excellence (NICE)
National Kidney Foundation
Norwegian Knowledge Centre for the Health Services
Ontario MOH Medical Advisory Secretariat
Panama and Costa Rica National Clinical Guidelines Program
Polish Institute for EBM
Scottish Intercollegiate Guideline Network (SIGN)
Society of Critical Care Medicine
Society of Pediatric Endocrinology
Society of Vascular Surgery
Spanish Society of Family Practice (SEMFYCY)
Stop TB Diagnostic Working Group
Surviving sepsis campaign
Swedish Council on Technology Assessment in Health Care
Swedish National Board of Health and Welfare
University of Pennsylvania Health System for EB Practice
UpToDate
World Health Organization (WHO)
What are we grading?

• two components

• quality of body of evidence
  - extent to which confidence in estimate of effect adequate to support decision
    • high, moderate, low, very low

• strength of recommendation
  • strong and weak
## Quality assessment criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Quality of Evidence</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomised trial</td>
<td>High</td>
<td>Risk of bias</td>
<td>Large effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>+2 Very large</td>
</tr>
<tr>
<td>Observational study</td>
<td>Moderate</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Indirectness</td>
<td>+1 Would reduce a demonstrated effect or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Would suggest a spurious effect when results show no effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>Imprecision</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publication bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Likely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very likely</td>
<td></td>
</tr>
</tbody>
</table>
## Beta blockers in non-cardiac surgery

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of participants (studies)</th>
<th>Serious Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Reporting Bias</th>
<th>Quality</th>
<th>Relative Effect (95% CI)</th>
<th>Absolute risk difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td>10,125 (9)</td>
<td>No</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>Undetected</td>
<td>High</td>
<td>0.71 (0.57 to 0.86)</td>
<td>1.5% fewer (0.7% fewer to 2.1% fewer)</td>
</tr>
<tr>
<td>Mortality</td>
<td>10,205 (7)</td>
<td>No</td>
<td>Possible ↓</td>
<td>OK</td>
<td>Imprecise</td>
<td>Undetected</td>
<td>Moderate or low</td>
<td>1.23 (0.98 – 1.55)</td>
<td>0.5% more (0.1% fewer to 1.3% more)</td>
</tr>
<tr>
<td>Stroke</td>
<td>10,889 (5)</td>
<td>No</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>Undetected</td>
<td>High</td>
<td>2.21 (1.37 – 3.55)</td>
<td>0.5% more (0.2% more to 1.3% more)</td>
</tr>
</tbody>
</table>
Resource use: just another outcome?

• yes and no

• who benefits?
  - different payers bear costs across societies and within (age)

• costs vary much more than other outcomes
  - across/within jurisdictions, over time

• even when resource use same implications differ
  - year’s supply expensive drug
  - nurses’ salary in U.S., 6 in Poland, 30 in China

• unbearable lightness of costs

• may decide to omit from consideration
GRADE’s approach to resource use

• identify viewpoint
• identify important resource use items
• find relevant evidence
• evaluate evidence quality
  - may differ across resource use items
  - RCTs start high, observational low
  - 5 categories for rating down, 3 up
  - economic analysis criteria inapplicable
• value resources in terms of cost
GRADE approach evidence quality

• reasons for risk of bias
  - failure ITT (methadone vs buprenorphine)
  - reliance on patient recall
  - reliance on imputation

• directness often major issue
  - older studies - different practice patterns
  - only cost provided
  - modeling necessary
Evidence summary

• as other outcomes, need systematic review

• quality of evidence, summary of findings
  - “balance sheet”, special form evidence profile

• resource use and not just costs
  - can judge whether resource use applicable to local setting
  - focus on items relevant to them (pharmacy)
  - apply unit costs to local setting
Example question

• patients
  - women with pre-eclampsia

• intervention
  - intravenous magnesium

• RCT done in 33 countries
  - over 9,000 patients

• health system perspective
Table 2 | Summary of findings on whether clinicians should use magnesium sulphate to prevent eclampsia: resource use viewed from the perspective of the health system

<table>
<thead>
<tr>
<th>Resource</th>
<th>Cost*</th>
<th>Typical absolute effect (95% CI)</th>
<th>No of participants (studies)</th>
<th>Quality of evidence</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnesium sulphate ampoules (6×10 ml ampoules/patient)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income countries</td>
<td>$20 more/patient</td>
<td></td>
<td>9996</td>
<td>High†</td>
<td></td>
</tr>
<tr>
<td>Middle income countries</td>
<td>$3 more/patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income countries</td>
<td>$5 more/patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration of magnesium sulphate (1 ampoule/patient)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income countries</td>
<td>$66/patient</td>
<td></td>
<td>9996</td>
<td>High†</td>
<td>Resources for giving magnesium sulphate included midwives’ time (main cost), intravenous cannula or needles, syringes, intravenous fluids, and the drug</td>
</tr>
<tr>
<td>Middle income countries</td>
<td>$14/patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income countries</td>
<td>$8/patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other hospital resources (varied widely)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income countries</td>
<td>$12 839</td>
<td>$20 less/ patient ($0 to $60)</td>
<td>9.996</td>
<td>Moderate‡</td>
<td>Use of other hospital resources varied greatly in both intervention and control groups. Other hospital costs have been adjusted for on the basis of the influence of eclampsia to control for the many other factors that influenced these costs</td>
</tr>
<tr>
<td>Middle income countries</td>
<td>$1 416</td>
<td>$4 less/ patient ($0 to $10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income countries</td>
<td>$157</td>
<td>$2 less/ patient ($1 to $3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*$1=£0.5=€0.7.
†Evidence comes from randomised trials and there was no reason to grade down for study limitations, imprecision, inconsistency, indirectness, or publication bias.
‡The confidence interval was wide so the evidence was graded down for imprecision.
Example question 2

• patients
  - opioid dependent

• intervention
  - buprenorphine versus methadone

• 2 RCTs

• societal perspective
Table 1 - Example of resource use evidence profile

**Question:** Should Buprenorphine maintenance flexible doses vs Methadone maintenance flexible doses be used for opioid maintenance treatment? **Patient or population:** Opiate dependents; **Setting:** Outpatients in USA, Australia, Austria, Switzerland, UK; **Viewpoint:** societal

<table>
<thead>
<tr>
<th>Studies (follow up)</th>
<th>Quality assessment</th>
<th>Summary of resources and costs</th>
<th>Overall Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td>No of patients</td>
<td>Methadone per patient</td>
</tr>
<tr>
<td>Harris 2005 (1 year)*</td>
<td>RCT</td>
<td>Serious limitations</td>
<td>Some uncertainty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Doran 2003 (6 months) *</td>
<td>RCT</td>
<td>No</td>
<td>Some uncertainty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other healthcare costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris 2005 (1 year) *</td>
<td>RCT</td>
<td>Serious limitations</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Doran 2003 (6 months) *</td>
<td>RCT</td>
<td>No</td>
<td>Some uncertainty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Crime costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris 2005 (1 year) *</td>
<td>RCT</td>
<td>Very serious limitations</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

* In this example we decided not to pool resource data from different studies due to insufficient information provided. NA = not available. SD = Standard deviation. SE = Standard Error.

* Including dispensing fee.

* Includes other prescription and OTC drugs, prescriber, inpatient, outpatient, emergency, ambulance, counseling, allied health and pathology services.

* Include staff time (i.e. face-to-face contact and preparation time), diagnostic procedures and facility level (Supplies, consumables, capital, equipment, ancillary support including administration, management, security, etc.

* Healthcare costs from assault, loss of income by the victims of crime, depreciable value of property damaged, stolen or obtained fraudulently, detection, prosecution and imprisonment.

* Some limitations because of incomplete outcome data.

* Some limitations because of incomplete outcome data and crucial limitations for self reported crime data.

* All the studies were conducted within the Australia health system (while the recommendation was global).

* Doses for methadone and buprenorphine derived from Mattick 2003 study, at the 10th week.

* The average cost of crime was substantial across the sample by these reported costs were associated with just a few participants. 90% of the sample randomized to methadone and 96% of that randomized on buprenorphine reported non-involvement in property crime. Indeed the majority of patients reported non criminal activity during the trial (6/66 patients for methadone and 3/73 for buprenorphine). (Page 86)
Strength of recommendations

• degree of confidence that desirable effects of adhering to recommendation outweigh undesirable effects.

• strong recommendation
  - benefits clearly outweigh risks/hassle/cost
  - risk/hassle/cost clearly outweighs benefit
<table>
<thead>
<tr>
<th>Factor</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance between desirable and undesirable</td>
<td>The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted. The narrower the gradient, the higher the likelihood that a weak recommendation is warranted.</td>
</tr>
<tr>
<td>Quality of evidence</td>
<td>The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a weak recommendation is warranted</td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td>The higher the costs of an intervention—that is, the greater the resources consumed—the lower the likelihood that a strong recommendation is warranted</td>
</tr>
</tbody>
</table>
Significance of strong vs weak

• variability in patient preference
  - strong, almost all same choice (> 90%)
  - weak, choice varies appreciably

• interaction with patient
  - strong, just inform patient
  - weak, ensure choice reflects values

• use of decision aid
  - strong, don't bother
  - weak, use the aid

• quality of care criterion
  - strong, consider
  - weak, don't consider
Value and preference statements

• underlying values and preferences always present

• sometimes crucial

• important to make explicit
Values and preferences

Stroke guideline: patients with TIA clopidogrel over aspirin (Grade 2B).

Underlying values and preferences: This recommendation to use clopidogrel over aspirin places a relatively high value on a small absolute risk reduction in stroke rates, and a relatively low value on minimizing drug expenditures.
Values and preferences

Peripheral vascular disease: aspirin be used instead of clopidogrel (Grade 2A).

**Underlying values and preferences:**
This recommendation places a relatively high value on avoiding large expenditures to achieve small reductions in vascular events.
Summary

- GRADE provide transparent structural framework for developing and presenting recommendations

- increasingly widely adopted

- refinements required, provides framework for dealing with resource use/cost