Practice guidelines: overview of methodology with focus on GRADE

Benjamin Djulbegovic, M.D., Ph.D.
Distinguished Professor
USF, Dpt of Internal Medicine
Chief, Division of Evidence-based Medicine and Health Outcome Research

CTSI, Director of
Center for Evidence-based Medicine and Health Outcome Research
Guideline Definition

“Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances”

Institute of Medicine, 1990
Types of Guidelines

• What to do?
  – Pathways/Algorithms
  – Boundary Guidelines

• How to do it. . .
  – Critical Care Paths

Courtesy of Dr. R. Winn
Methods of Developing Guidelines

• Informal consensus
• Formal consensus
• Evidence-based medicine approach
• Explicit approach
Consensus

Although it may capture collective knowledge, it is also vulnerable to the possibility of capturing collective ignorance

-- Murphy, 1998
Importance of grading quality of evidence and strength of recommendations

- Patients and physicians using clinical practice and other recommendations need to know how much confidence they can place in the recommendations.
- Clinical guidelines are only good as good as the evidence and judgments that are based on.
- Systematic and explicit methods of making judgments can reduce errors and improve communication.
What does the patient and his physician need?

Evidence-based principles

- **Evidence: selective citation vs. totality of evidence**
  - Need for systematic reviews of the totality of research evidence
  - Simultaneous instead of separate presentations of evidence (benefits and harms)

- **Assessing the quality**
  - Critical appraisal of the quality of research is central to informed decision in health care
  - Quantity, quality (internal validity), consistency

- **Benefits and harms**
  - Relative effect measures vs. absolute effect measures (NNT)
    - Minimizing framing effect
    - Probability vs. certainty

- **Patient-oriented evidence vs. disease-oriented evidence**
  - Evidence on survival, DFS, QOL is more important than evidence on tumor response, markers etc

- **Help with decisions**
  - Effective health-care recommendations vs. preference-sensitive health-care recommendations (decisions)
The need for research synthesis

• Health care decision makers need to access research evidence to make informed decisions on diagnosis, treatment and health care management for both individual patients and populations.

• There are few important questions in health care which can be informed by consulting the result of a single empirical study.
Systematic reviews of the totality research evidence represents a scientific foundation for development of clinical practice guidelines and health technology assessments.
The need for better methods of research synthesis: the rise of systematic reviews

- **Systematic Review**
  - "The application of strategies that limit bias in the **assembly**, **critical appraisal**, and **synthesis** of all relevant studies on a specific topic. Meta-analysis may be, but is not necessary, used as part of this process."

- **Meta-Analysis**
  - "The statistical synthesis of the data from separate but similar, i.e. comparable studies, leading to a quantitative summary of the pooled results."

Last JM. Dictionary of Epidemiology, 2001
All major theories of choice agree that rational decision-making requires integrations of:

- **Benefits** (gains)
- **Harms** (losses)

Theories of decision-making differ in the proposal how benefits and harms should be integrated in a given decision.
Patient or population: newly diagnosed (previously untreated) patients with multiple myeloma
Intervention: High-dose chemotherapy with single autologous transplant
Control: Chemotherapy
Outcomes: Overall Survival, Progression-free survival, Treatment related mortality

Denotes quantity of evidence
Denotes quality of evidence
Consistency
Generalizability
Power of the evidence
Relative effect measures
Absolute effect measures
Denotes Quality

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of patients</td>
<td>Effect</td>
</tr>
<tr>
<td>No of studies</td>
<td>Design</td>
</tr>
<tr>
<td>Benefits</td>
<td>Overall Survival (presented as total mortality):</td>
</tr>
<tr>
<td>9</td>
<td>Randomized trials</td>
</tr>
<tr>
<td>Benefits</td>
<td>Progression-free survival:</td>
</tr>
<tr>
<td>9</td>
<td>Randomized trials</td>
</tr>
<tr>
<td>Harms</td>
<td>Treatment related mortality:</td>
</tr>
<tr>
<td>9</td>
<td>Randomized trials</td>
</tr>
</tbody>
</table>

Footnotes:
1. Significant heterogeneity among included studies (Q statistic of 27.65; p<0.01). All the exploratory sensitivity analysis results did not differ from the main result of excluding benefit for survival with high-dose therapy plus transplant versus chemotherapy only.
2. Significant heterogeneity among included studies (Q statistic of 51.57; p<0.01). However different sensitivity analysis based on either excluding non standrad trials or source of stem cells or length of follow-up etc. did not result in any significant difference from the original results.
3. Assuming 60% PFS at 3 yrs
4. Assuming baseline risk of death of 1% in control group

Analytic Framework: does platelet transfusion administered at different target values of platelet count result in different clinical outcomes?

Linking guidelines/HTA to systematic reviews

**PICO**

**Patient population**

1. Non-bleeding patients with low platelets
   a) Hypoproliferative thrombocytopenia --AML --chemotherapy
2) Actively bleeding patient
   a) ITP
   b) TTP etc
3. Non-bleeding patient undergoing surgical procedures

**Comparisons**

Transfusion at Platelets <5K vs. >10-5K

**Management Decisions**

Is platelet count trigger associate with different management decisions?

Harms of RBC transfusion (HIV, hep B,C, HTLV, TRALI etc)

Impact on subsequent management decisions (e.g., treatment for TRALI, iron overload etc)

Outcomes:

1. Blood loss/Bleeding (major)
2. Mortality
3. Nonfatal myocardial infarction
4. Stroke (hemorrhagic and ischemic)
5. RBC transfusion requirements.
6.?
Evidence & Decision making

• Evidence is necessary but not sufficient for optimal decision-making
• Making categorical recommendations (considered judgments)
• Qualitative exercise
  – Occasionally is supplemented with quantitative (decision-analytic) modeling
Effective health-care recommendations
- Effective health-care (strong recommendations) when benefits >>>harms: candidate for quality criteria

Preference-sensitive recommendations
- Judgments about benefits/harm ratio uncertain, depend on patient values and preferences
- May be based on quantitative or qualitative judgments about (explicitly) summarized evidence

decision-making process must be transparent and explicit with clarity regarding the critical criteria that informed recommendations; based on shared deliberation & must include appeal process

“Accountability for reasonableness”
- which may help legitimize specific choices that may favor one set of stakeholders over others

After O’Connors, 2003; Teutsch et al, 2005; Daniels at al, 1997
Systematic review

Guideline development

Grade recommendations
• For or against (direction) ↓ ↑
• Strong or conditional/weak (strength)

By considering balance of:
  □ Quality of evidence
  □ Balance benefits/harms
  □ Values and preferences

Revise if necessary by considering:
  □ Resource use (cost)

Formulate Recommendations (↓ ↑ | ⊕ ...)
• “We recommend using...” | “Clinicians should...”
• “We suggest using...” | “Clinicians might...”
• “We suggest not using...” | “Clinicians ... not...”
• “We recommend not using...” | “Clinicians should not...”

Panel

Guideline

Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence

Summary of findings & estimate of effect for each outcome

Grade up
Grade down

Input?

Input?

Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes

Grade down
Grade up

Summary of findings & estimate of effect for each outcome

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence

Summary of findings & estimate of effect for each outcome

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence

Summary of findings & estimate of effect for each outcome

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence

Summary of findings & estimate of effect for each outcome

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence

Summary of findings & estimate of effect for each outcome

Risk of bias
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Incomplete evidence
Formulation of guidelines: main principles

• Separate evidence from decision-making

• **Quality of evidence** indicates the extent to which one can be confident that an estimate of effect is correct
  – represented on a *continuum scale* of credibility

• **Strength of recommendations** indicates the extent to which one can be confident that adherence to a recommendation will do more good than harm
  – Represent decision-making about choice and is *categorical exercise* (we recommend or do not)
From Evidence to Decision-making (recommendations)

Continuum from Study Quality Through Strength of Evidence to Guideline Development

Quality of Studies

Strength of Evidence

Clinical Practice Guidelines

Central role of evidence

I

Quality
Magnitude
Consistency

II

Categorical recommendations (statements)

Broad Categories
- Good vs. High
- Fair vs. Moderate
- Poor vs. Low
- Very low

* FOR
* AGAINST
* Can’t recommend
# Guidelines development process

## Prior steps in developing guidelines
- Prioritise problems, establish panel

## Preparatory steps
- Systematic review

## Grading the quality of evidence and the strength of recommendations
- Quality of evidence for each outcome
  - Relative importance of outcomes
  - Overall quality of evidence
  - Balance of benefits and harms
    - (Does the intervention do more good than harm?)
  - Balance of net benefits and costs
    - (Are incremental health benefits worth the costs?)
  - Strength of recommendation

## Subsequent steps
- Implementation and evaluation
Hierarchy of outcomes according to importance to patients to assess effect of phosphate lowering drugs in patients with renal failure and hyperphosphataemia

What are we assessing/grading?

• two components

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

• strength of recommendation
  – The extent to which we can be confident that the desirable effects of an intervention outweigh the undesirable effects
The importance of context: conclusions vs. decisions

- **Quality of evidence** (="conclusions")
  - The extent of confidence that an estimate of effect is correct i.e. representing the "truth"
    - Important for systematic reviews
  - The extent to which confidence in an estimate of the effect is adequate to support recommendations
    - Importance for the guidelines panels

- **Making recommendations**
  - The extent to which we can be confident that the desirable effects of an intervention outweigh the undesirable effects
    - Important for guidelines panels
    - **NB as long as there is judgment that benefits>>>harms, recommendation can be strong even if the quality of evidence is low or very low**
      - assumes that the error making a strong recommendation will be regretted less than the error making a weak recommendation
GRADE: categories of quality

- **High**: Considerable confidence in the estimate of effect.
  - True effect likely lies close to our estimate of the effect
  - Further research unlikely to change our confidence in estimate

- **Moderate**: moderately confident that the estimate is close to the truth
  - Further research likely to have important impact on confidence in estimate, may change estimate.

- **Low**: confidence in the effect limited. True effect may be substantially different from the estimate
  - Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

- **Very low**: little confidence in the effect estimate
  - Any estimate of effect is very uncertain.
### GRADE quality assessment criteria: therapeutic studies

<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 trial</td>
<td>Risk of bias:</td>
<td>Strong association:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious limitations</td>
<td>+1 Large effect (Strong, no plausible confounders,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious limitations</td>
<td>consistent and direct evidence)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Inconsistency</strong></td>
<td>+2 Very large effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>(Very strong, no major threats to validity and direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>evidence)***</td>
</tr>
<tr>
<td>Moderate</td>
<td>Quasi-randomised trial</td>
<td>Indirectness:</td>
<td>+1 Evidence of a Dose response gradient</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>+1 All plausible confounders would have reduced the</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Imprecision</strong></td>
<td>effect</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>Low</td>
<td>Observational study</td>
<td>Reporting bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 likely</td>
<td>+1 Evidence of a Dose response gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very likely</td>
<td>+1 All plausible confounders would have reduced the</td>
</tr>
<tr>
<td>Very low</td>
<td>Any other evidence</td>
<td></td>
<td>effect</td>
</tr>
</tbody>
</table>

* 1 = move up or down one grade (for example from high to intermediate)
  2 = move up or down two grades (for example from high to low)

** A statistically significant relative risk of $>2$ ($<0.5$), based on consistent evidence from two or more observational studies, with no plausible confounders

*** A statistically significant relative risk of $>5$ ($<0.2$) based on direct evidence with no major threats to validity
Sources of bias: Rx

- Choice of the control intervention

**Selection bias**
- systematic differences in comparison groups

**Performance bias/information bias**
- systematic differences in care provided apart from the intervention being evaluated
- systematic error in the measurement of information on exposure or outcome

**Attrition bias**
- (systematic differences in withdrawals from the trial)

**Detection bias/Recall bias**
- (systematic differences in outcome assessment)

**Specimen handling bias**
- (systematic differences in analysis of specimens)

Target Population (baseline state)

Allocation

- Intervention Group
  - Exposed to Intervention
    - Follow-up
      - Outcomes

- Control Group
  - Not exposed to Intervention
    - Follow-up
      - Outcomes

**Analysis appropriateness**
- (was analysis reflective of the problem at hand? ITT vs. PP)
Controlling for selection bias

• Randomized controlled trials
  – **Generation** of allocation sequence
    • *In RCTs this is usually done by computer using any number of available methods (usually block randomization, etc)*
  – **Concealing** treatment assignment until after the treatment has been allocated

• Observational research
  – *In a cohort study: are participants in the exposed and unexposed groups similar in all important respects except for the exposure?*
    • Control for confounders
  – *In a case-control study: are cases and controls similar in all important respects except for the disease in question?*
    • matching
Controlling for performance bias

• RCTs
  – Are we controlling for co-intervention/contamination?
  – a method to prevent that those who providing and receiving care do not know to which intervention group the recipients of care have been allocated
  – use of “blinding/masking”

• Observational research
  – Accounting for information (measurement) /recall bias
    – In a cohort study: is information about outcome obtained in the same way for those exposed and unexposed?
    – In a case-control study, is information about exposure gathered in the same way for cases and controls?
Controlling for attrition bias

• RCTs
  – Complete follow-up
  – Baseline characteristics of participants lost to follow-up and those included in the analysis should be reported separately

• Observational research
  – Cohort/case-control studies: Completeness of follow-up
  – Baseline characteristics of participants lost to follow-up and those included in the analysis should be reported separately
“Intention to treat” vs. ‘per protocol’ analysis

• All patients should be analysed in the arm to which they were allocated at randomisation, regardless of whether they receive the allocated treatment (‘Intention-to-treat’ analysis).
Observer bias

- The biases that lead to misperceptions that we have detected, seen or experienced something that actually isn't there

  - **Placebo/masking technique to control for observer bias**
    - Mesmerism and Franklin’s commission appointed by Louis XVI in 1784 to investigate the medical claims of "animal magnetism", or "mesmerism".
    - The people being studied felt the effects of mesmerism only when they were "told" and felt no effects when they were not told, whether or not they were receiving the treatment.
Confounding by indication: important quality issue in transfusion medicine

• Results from the conscious choice of different treatments for patients with different prognosis
  – According to severity of disease

• Probably the most important bias in clinical research
  – Observational studies
  – Can be avoided by performing a well designed RCT
Blood transfusions: Good or Bad?

Sick patients

More transfusions  Spurious association  Poor outcome
Confounding by indications: apparent protective effect of liberal platelet transfusion strategy on poor outcomes (e.g., bleeding, etc)

More transfusions
platelets 5-20 k/ccu

Less transfusions
plat 15-20 k/ccu

Sick patients

Less transfusion (plt 15-20) / More transfusions (plt 5-20) = # Poor outcomes / # Poor outcomes <1

Spurious association

Adopted from Transfusion 2010; 50:1181-1183
Factors that might decrease quality of evidence

- Study limitations (risk of bias)
  - Inadequacy of allocation concealment; lack of blinding, large drop-outs, failure to perform ITT, failure to report outcomes,
  - Inconsistency of results
    - Variability or heterogeneity in results due true differences in treatment effect (due to P-I-C-O)
    - Statistical: large I² (e.g. >50%); clinical: (PICO)

- Indirectness of evidence (2 types)
  - Lack of head-to-head comparisons
  - differences in treatment effect (due to P-I-C-O)

- Imprecision
  - A few events (<200-300?), small studies (N<400), wide confidence intervals consistent with important differences in both directions or no effects or all

- Reporting (publication) bias

- Other factors
  - Carryover effect in crossover trials, use of unvalidated outcome measures, recruitment bias in cluster RCT etc

Factors that might increase quality of evidence

- Large magnitude of effect
  - A statistically significant relative risk of > 5 (< 0.2)

- Plausible confounding, which would reduce a demonstrated effect is accounted for without affecting treatment effect

- Dose-response gradient
Quality of reporting compared with actual methodological quality

Restrictive vs. liberal RBC transfusion: effect of assessors’ blinding (MI)

7.12.1 low risk for bias (assessors blinded)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Restrictive</th>
<th></th>
<th>Liberal</th>
<th></th>
<th></th>
<th></th>
<th>Risk Ratio</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson 2011</td>
<td>38</td>
<td>1009</td>
<td>23</td>
<td>1007</td>
<td>38.2%</td>
<td></td>
<td>1.65 [0.99, 2.75]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foss 2009</td>
<td>1</td>
<td>60</td>
<td>0</td>
<td>60</td>
<td>6.0%</td>
<td></td>
<td>3.00 [0.12, 72.20]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grover 2005</td>
<td>0</td>
<td>109</td>
<td>1</td>
<td>109</td>
<td>5.9%</td>
<td></td>
<td>0.33 [0.01, 8.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson 1992</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>18</td>
<td>6.1%</td>
<td></td>
<td>0.30 [0.01, 6.97]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lotke 1999</td>
<td>1</td>
<td>62</td>
<td>0</td>
<td>65</td>
<td>6.0%</td>
<td></td>
<td>3.14 [0.13, 75.72]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td><strong>1260</strong></td>
<td></td>
<td><strong>1259</strong></td>
<td></td>
<td><strong>62.2%</strong></td>
<td></td>
<td><strong>1.57 [0.97, 2.55]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>40</td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 2.35, df = 4 (P = 0.67); I² = 0%
Test for overall effect: Z = 1.82 (P = 0.07)

7.12.2 high-risk for bias (assessors not blinded)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Restrictive</th>
<th></th>
<th>Liberal</th>
<th></th>
<th></th>
<th></th>
<th>Risk Ratio</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bracey 1999</td>
<td>1</td>
<td>212</td>
<td>0</td>
<td>216</td>
<td>5.9%</td>
<td></td>
<td>3.06 [0.13, 74.61]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bush 1997</td>
<td>1</td>
<td>50</td>
<td>2</td>
<td>49</td>
<td>9.7%</td>
<td></td>
<td>0.49 [0.05, 5.23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebert 1999</td>
<td>3</td>
<td>418</td>
<td>12</td>
<td>420</td>
<td>22.2%</td>
<td></td>
<td>0.25 [0.07, 0.88]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td><strong>680</strong></td>
<td></td>
<td><strong>685</strong></td>
<td></td>
<td><strong>37.8%</strong></td>
<td></td>
<td><strong>0.39 [0.13, 1.17]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>5</td>
<td></td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.06; Chi² = 2.10, df = 2 (P = 0.35); I² = 5%
Test for overall effect: Z = 1.69 (P = 0.09)

Total (95% CI) 1940 1944 100.0% 0.88 [0.38, 2.04]

Total events 45 39
Heterogeneity: Tau² = 0.41; Chi² = 10.42, df = 7 (P = 0.17); I² = 33%
Test for overall effect: Z = 0.29 (P = 0.77)
Test for subgroup differences: Chi² = 5.18, df = 1 (P = 0.02), I² = 80.7%
Assessment of the quality of evidence in RCTs testing restrictive vs. liberal transfusion strategy
(based on Cochrane review by Carless et al)

- **Use of RBC transfusion:**
  - High
    - None of the potential flaws appear to have significant effect on the results

- **30 days mortality**
  - High
    - The results between high-quality trials and those with the flaws consistent. Hence, none of the potential flaws appear to have significant effect on the results

- **Myocardial infarction/Cardiac events**
  - Very low
    - Trials in which outcome assessors were not blinded favored restricted strategy, which may have incorporated biased assessment of outcomes

- **Walking independently at 60 days**
  - Low
    - Based on self-reporting from one (high-quality) trial (sparse data)

- **Length of stay**
  - Moderate
    - Decision to discharge may be a function of knowledge of treatment group

- **CHF**
  - Very low
    - It is not clear how CHF was diagnosed; a few data provided in the Cochrane review
      - Pulmonary edema clinically can encompass many conditions including TRALI, CHF etc
GRADE recommends that the guideline developers consider the quality of evidence across outcomes as that associated with the **critical outcome with the lowest quality evidence**.

- GRADE requires **guideline developers, but not systematic review authors**, to make an overall rating of evidence quality **across outcomes** deemed critical for decision-making.

- [NB The principle is that if there is higher quality evidence from some critical outcomes to support a decision in favour of an intervention (that is, benefits on critical outcomes clearly outweigh undesirable effects of the intervention, for which there is also high quality evidence) one needn’t rate down the quality because of lower quality evidence regarding other critical outcomes that support the same recommendation]

BMJ 2008
From: Red Blood Cell Transfusion: A Clinical Practice Guideline From the AABB*


Figure Legend:
Adverse effects of RBC transfusion contrasted with other risks.
From evidence to recommendations

- Evidence is necessary but not sufficient for optimal decision-making
- Making categorical recommendations (considered judgments)
- Qualitative exercise
  - Occasionally is supplemented with quantitative (decision-analytic) modeling
  - Driven by normative/prescriptive principles
Deliberations/decisions

Trade-offs between benefits (B) and harms (H) under uncertainties

Quality of evidence

Values/Preferences (constructed as a response to information on B&H)

Setting/resource use ("non-clinical" factors including emotions/affect)

Factors affecting decision-making

GRADE 2008
### Determinants of strength of recommendation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Balance between desirable and undesirable effects</strong></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><strong>Quality of evidence</strong></td>
<td>The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted.</td>
</tr>
<tr>
<td><strong>Values and preferences</strong></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><strong>Costs (resource allocation)</strong></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>
Representations of quality of evidence and strength of recommendations

<table>
<thead>
<tr>
<th>Quality of evidence</th>
<th>Strength of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High quality</td>
<td>Strong recommendation for using an intervention: ↑ ↑ or 1</td>
</tr>
<tr>
<td>Moderate quality</td>
<td>Weak recommendation for using an intervention: ↑ ? or 2</td>
</tr>
<tr>
<td>Low quality</td>
<td>Weak recommendation against using an intervention: ↓ ? or 2</td>
</tr>
<tr>
<td>Very low quality</td>
<td>Strong recommendation against using an intervention: ↓ ↓ or 1</td>
</tr>
</tbody>
</table>

Guyatt, G. H et al. BMJ 2008;336:1049-1051
## Factors that affect the strength of a recommendation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Examples of strong recommendations</th>
<th>Examples of weak recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of evidence</td>
<td>Many high quality randomised trials have shown the benefit of inhaled steroids in asthma</td>
<td>Only case series have examined the utility of pleurodesis in pneumothorax</td>
</tr>
<tr>
<td>Uncertainty about the balance between desirable and undesirable effects</td>
<td>Aspirin in myocardial infarction reduces mortality with minimal toxicity, inconvenience, and cost</td>
<td>Warfarin in low risk patients with atrial fibrillation results in small stroke reduction but increased bleeding risk and substantial inconvenience</td>
</tr>
<tr>
<td>Uncertainty or variability in values and preferences</td>
<td>Young patients with lymphoma will invariably place a higher value on the life prolonging effects of chemotherapy than on treatment toxicity</td>
<td>Older patients with lymphoma may not place a higher value on the life prolonging effects of chemotherapy than on treatment toxicity</td>
</tr>
<tr>
<td>Uncertainty about whether the intervention represents a wise use of resources</td>
<td>The low cost of aspirin as prophylaxis against stroke in patients with transient ischemic attacks</td>
<td>The high cost of clopidogrel and of combination dipyridamole and aspirin as prophylaxis against stroke in patients with transient ischaemic attacks</td>
</tr>
</tbody>
</table>
Has a mistake been made? Explicitly taking consequences into guidelines considerations

• **We can always** make a mistake
  – Recommend ineffective treatments
    • Regret of commission
  – Fail to recommend effective treatments
    • Regret of omission

• **Sense of loss, or regret**
  – How many times regret of commission is worse than regret of omission
# Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive

If you need to vote: Insert the number of votes for the recommendation in each category

<table>
<thead>
<tr>
<th>Assessors’ view of the balance between desirable and undesirable consequences of the intervention</th>
<th>Desirable consequences clearly outweigh undesirable consequences</th>
<th>Desirable consequences probably outweigh undesirable consequences</th>
<th>Undesirable consequences probably outweigh desirable consequences</th>
<th>Undesirable consequences clearly outweigh desirable consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strength of recommendation</strong></td>
<td>Strong for an intervention</td>
<td>Conditional (weak) for an intervention</td>
<td>Conditional (weak) against an intervention</td>
<td>Strong against an intervention</td>
</tr>
<tr>
<td><strong>Wording of a recommendation</strong></td>
<td>We recommend to “do something”</td>
<td>We suggest (conditionally recommend) to “do something”</td>
<td>We suggest (conditionally recommend) not to “do something”</td>
<td>We recommend not to “do something”</td>
</tr>
<tr>
<td><strong>Number of votes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NB typically one defines the rules advance. For example, one suggested rule is that recommendation for or against a particular intervention (compared with a specific alternative) will be made if at least 50% of the panel members vote in favor, with less than 20% preferring comparator. Failure to meet this criterion result in no recommendation (or “only in research”). For recommendations to be graded as strong vs. weak, at least 70% of the panel members should endorse it as “strong”.)
Making recommendations

Strength of the recommendation:
- Strong
- Conditional (weak)

Final recommendation:

Strength:   Quality of evidence:

Assumptions about underlying values and preferences

Remarks
Systematic review

Formulate question
- Select outcomes
- Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

<table>
<thead>
<tr>
<th>Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Very low</td>
</tr>
</tbody>
</table>

Grade down

1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up
1. Large effect
2. Dose response
3. Opposing bias & Confounders

Summary of findings & estimate of effect for each outcome

**PICOC**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>Critical</td>
</tr>
<tr>
<td>Outcome</td>
<td>Important</td>
</tr>
<tr>
<td>Outcome</td>
<td>Not important</td>
</tr>
</tbody>
</table>

**Guideline development**

Grade recommendations
- For or against (direction) ↓↑
- Strong or conditional/weak (strength)

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

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

Formulate Recommendations (↓↑ | ⊕…)
- “We recommend using...” | “Clinicians should...”
- “We suggest using...” | “Clinicians might...”
- “We suggest not using...” | “Clinicians ... not...”
- “We recommend not using...” | “Clinicians should not...”
| Evidentiary Standards: Clinical, Judicial, GRADE and FDA | Beyond reasonable doubt \([= \text{when both in the worst } (P_{\text{worst}}) \text{ (skeptic) and best } (P_{\text{best}}) \text{ (enthusiast) case scenarios probability that intervention will exceed clinically important thresholds } > 95\%]\) | Criminal cases | “Substantial Evidence” \((\text{FDA marketing approval})\) | GRADE: Strong recommendation for intervention \((\text{high quality of evidence})\) | Regret (of wrongly ) recommending \(< \ll \text{regret of not recommending} \)

| Clear and convincing evidence \([P_{\text{worst}} < 95\%; P_{\text{best}} > 95\%]\) | Malpractice litigation | Strong recommendation \((\text{moderate quality of evidence})\) |  |

| Preponderance of evidence \([P_{\text{worst}} > 50\%, P_{\text{best}} < 95\%]\) | Civil trials | AA \((DOE \rightarrow POE)\) | Strong vs. weak recommendation \((\text{low quality of evidence}) \text{(context-dependent)}\) | Regret Rx \(< \ll \text{regret NoRx} \)

| Reasonable to believe \([P_{\text{worst}} < 50\%; P_{\text{best}} > 50\%]\) | Search warrants, reasonable suspicion | “Reasonable to believe” \((\text{EUA})\) | Weak recommendation \((\text{very low quality of evidence})\) | Regret Rx \(< \ll \text{regret NoRx} \)

| Insufficient evidence \([P_{\text{worst}} < 50\%, P_{\text{best}} < 50\%]\) |  | Do not recommend vs. “Only in research” \((\text{context-dependent})\) | Regret Rx \(\geq \text{regret NoRx} \) |