Basics of GRADE

UW CLEAR Center / PM&R Journal Methods Webinar
10/21/2019

M. Hassan Murad, MD, MPH
Professor of Medicine
Mayo Clinic

Outline

- What is GRADE and for whom?
- Historical background
- First part of GRADE, certainty
- Second part of GRADE, action

What is GRADE?

- A system to:
  - Determine how trustworthy is the evidence (certainty in evidence)
  - Move from evidence to recommendations (action)

Who is GRADE for?

- Knowledge of GRADE is critical for:
  - Guideline developers (obviously)
  - Systematic reviewers (obviously)
  - Editors, peer reviewers
  - Media reporters
  - Researchers
  - ? Patients

Historical background

- Guidelines: statements of action telling us what to do with the aim of improving patient care and standardization
- 1970s Guidelines based on consensus of experts & cherry-picked evidence (if any)
- Late 80s-90s: evidence based medicine movement led to guidelines based on research, dependent on study design:
  - RCTs→ compelling recommendations
  - Non-RCTs→ vague recommendations

Historical background

- 2000s GRADE was developed based on the observations:
  - Not all RCTs are good
  - Some non RCTs provide compelling evidence
  - Evidence alone is not enough for decisionmaking

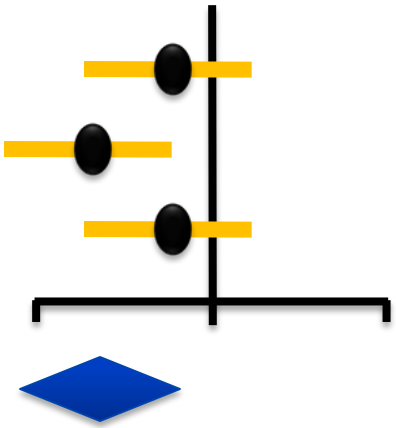
Historical background

- Currently GRADE has become the gold standard with >100 organizations/entities and >10,000 publications
- Guideline methodology research almost solely on GRADE
- A few organizations not using but borrowing GRADE principles

Reminder: hat is GRADE?

- A system to:
  - Determine how trustworthy is the evidence (certainty in evidence)
  - Move from evidence to recommendations (action)

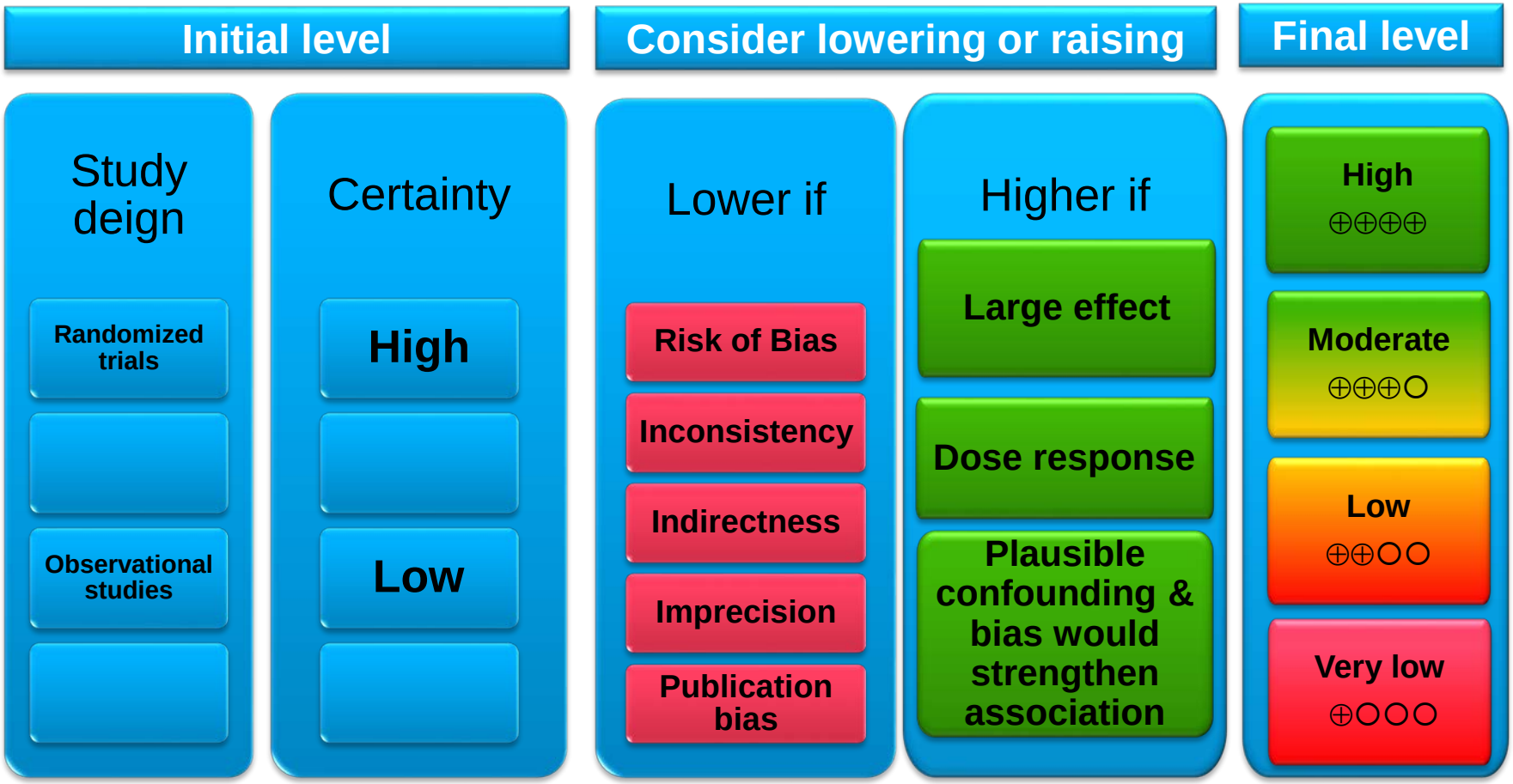
Certainty in what?



OR 0.70 (0.50-0.90)

- It is NOT quality of a study (risk of bias)
- Rather, certainty is about that the true effect is in a certain range of values (that may warrant an action)

Level of certainty (quality, confidence, strength)



1. Study limitations

- Also called risk of bias
- Per study, per outcome
- Then, overall for all the studies
- Tools:
  - RCTs: Cochrane tool
  - Observational: Newcastle Ottawa, ROBINS
  - Diagnostic: QUADS2
  - etc...

2. Inconsistency/Heterogeneity

- Differences in results between individual trials
- Causes
  - Clinical differences (patients, interventions, outcomes)
  - Methods (risk of bias)
  - Chance
- Hypotheses to explain inconsistency are better made before seeing results (*a priori*)

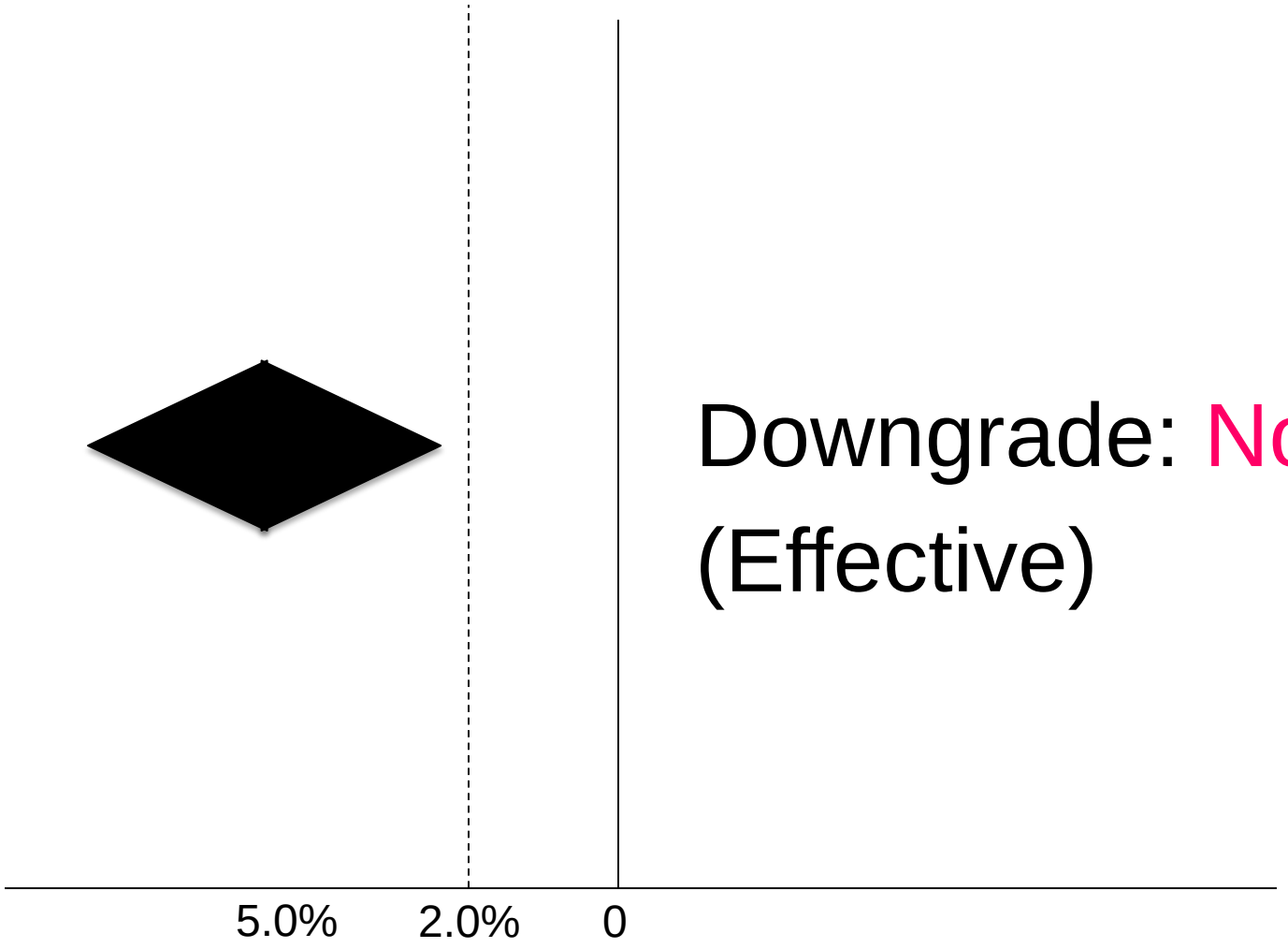
Inconsistency/Heterogeneity

- Common visual representation in forest plots
- Eye ball test
  - Variation in effect size
  - Overlap of confidence intervals
- Statistical tests for heterogeneity
  - Tau-squared
  - Chi-squared
  - I^2 (preferred)
    - % of Heterogeneity not attributable to chance
    - Heterogeneity between studies/between studies+ within studies

3. Imprecision

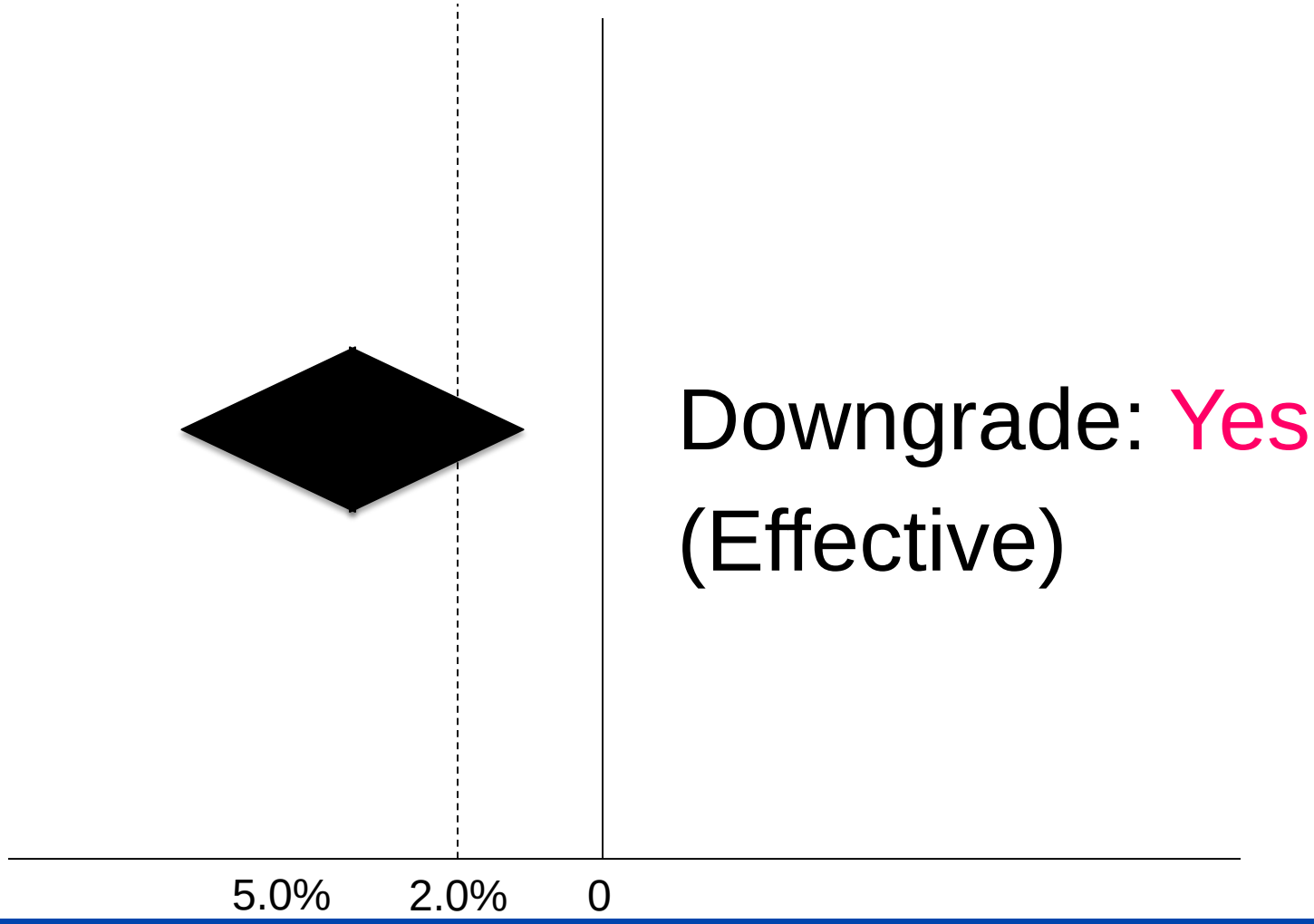
- Uncertainty about “true” effect size as reflected by:
  - Low event rate/small sample size
  - Wide confidence intervals
- Downgrade for imprecision when:
  - Decisions would differ if the truth was upper vs lower boundary

Imprecision



Downgrade: No (Effective)

Imprecision



Downgrade: Yes
(Effective)

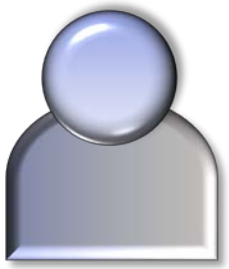
4. Indirectness of Evidence

- Consider the PICO elements :
  - Population/patients (eg, old versus young)
  - Interventions (eg, intravenous vs oral drug administration)
  - Outcomes
    - Patient important vs surrogates
    - Long follow up vs short follow up)

PICO
Patients
Interventions
Comparisons
outcomes

?

PICO
Patients
Interventions
Comparisons
outcomes



5. Publication Bias

- Faster publication of “positive” trials
- Slower or no publication of “negative” trials
- SR needs to search “grey literature” for unpublished studies

How common?

- 20-30% of RCTs submitted to FDA are never published. In one case (RCTs of reboxetine), 75% of data were unpublished
- Not at random: published data overestimate benefit & underestimate harm
- How to detect it:
  - Non-statistically (registries, protocols, FDA records)
  - Statistically (important limitations such having enough studies and some homogeneity)

Level of certainty

Preliminary Level		Modify Level		Final Level
Design	Certainty	Lower if	Higher if	QoE
Randomized trials	High	Risk of Bias	Large effect	High ⊕⊕⊕⊕
		Inconsistency	Dose response	Moderate ⊕⊕⊕⊖
Observational studies	Low	Indirectness	Residual plausible confounding & bias strengthen association	Low ⊕⊕⊖⊖
		Imprecision		Very low ⊕⊖⊖⊖
		Publication bias		

GRADE NETWORK

Scenario 1:

- Total hip replacement for people with disabling hip arthritis
- No RCT compared THR to no THR
- How certain are we that THR helps in reducing OA pain and disability?
- Many other examples from observational studies

Scenario #1 Consider rating up for a large effect

- Modeling and empirical evidence suggests that confounding (from nonrandom allocation) alone:
  - unlikely to explain associations with RR >2 (< 0.5)
  - very unlikely to explain associations with RR>5 (<0.2)
- Confidence in the estimates is then rated up once or twice
- Particularly when considering rapidity and trajectory

THE U.S. GRADE NETWORK

J Chronic Dis 1967;20:487-95.
Cochrane Database Syst Rev 2007;2. MR000012.

Scenario 2: Dose-effect gradient

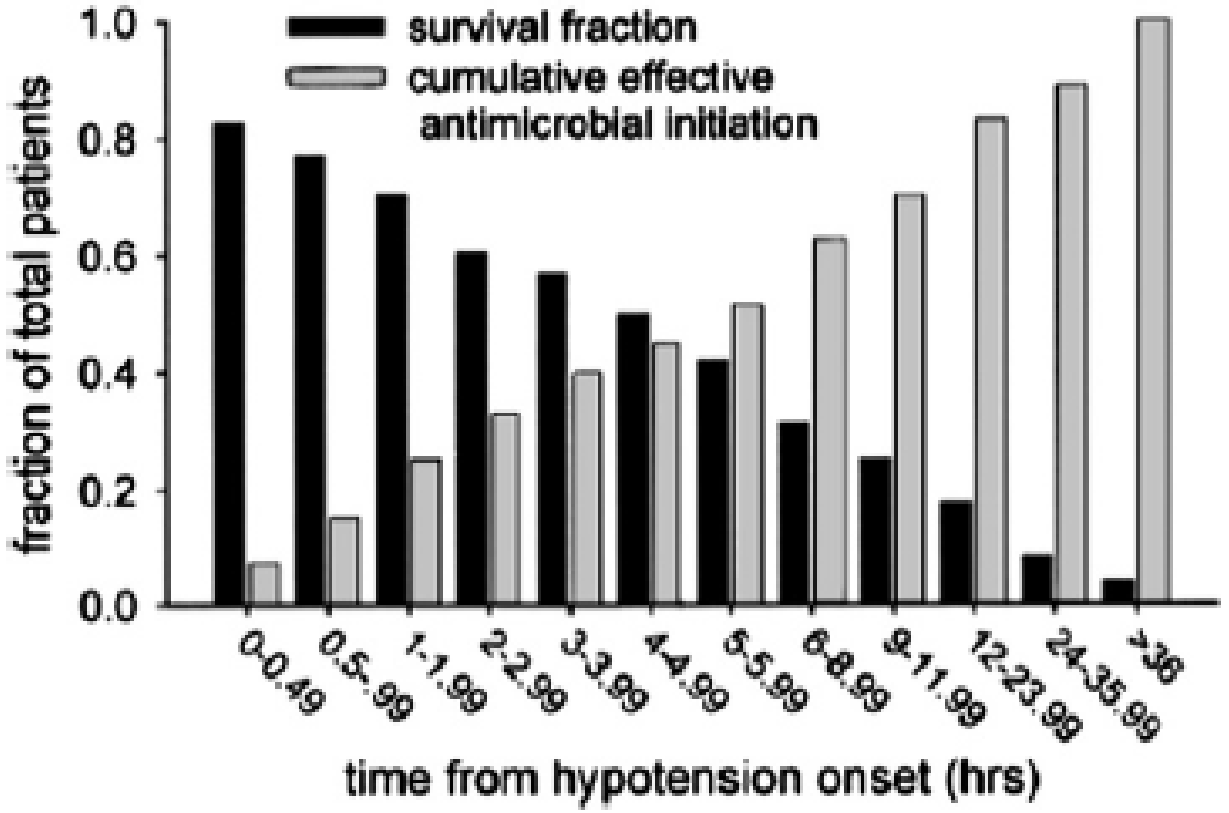


Fig. 1. Cumulative effective antimicrobial initiation following onset of septic shock-associated hypotension and associated survival. The *x*-axis represents time (h) following first documentation of septic shock-associated hypotension. *Black bars* represent the fraction of patients surviving to hospital discharge for effective therapy initiated within the given time interval. The *gray bars* represent the cumulative fraction of patients having received effective antimicrobials at any given time point.

THE U.S. GRADE NETWORK

Scenario 3:

- Well-done observational studies showed no association between autism and vaccinations
- However, these studies suffer from recall bias (parents of children with autism more likely to remember the proximity of vaccination to onset of autism)
- How does this recall bias affect our certainty about lack of association?

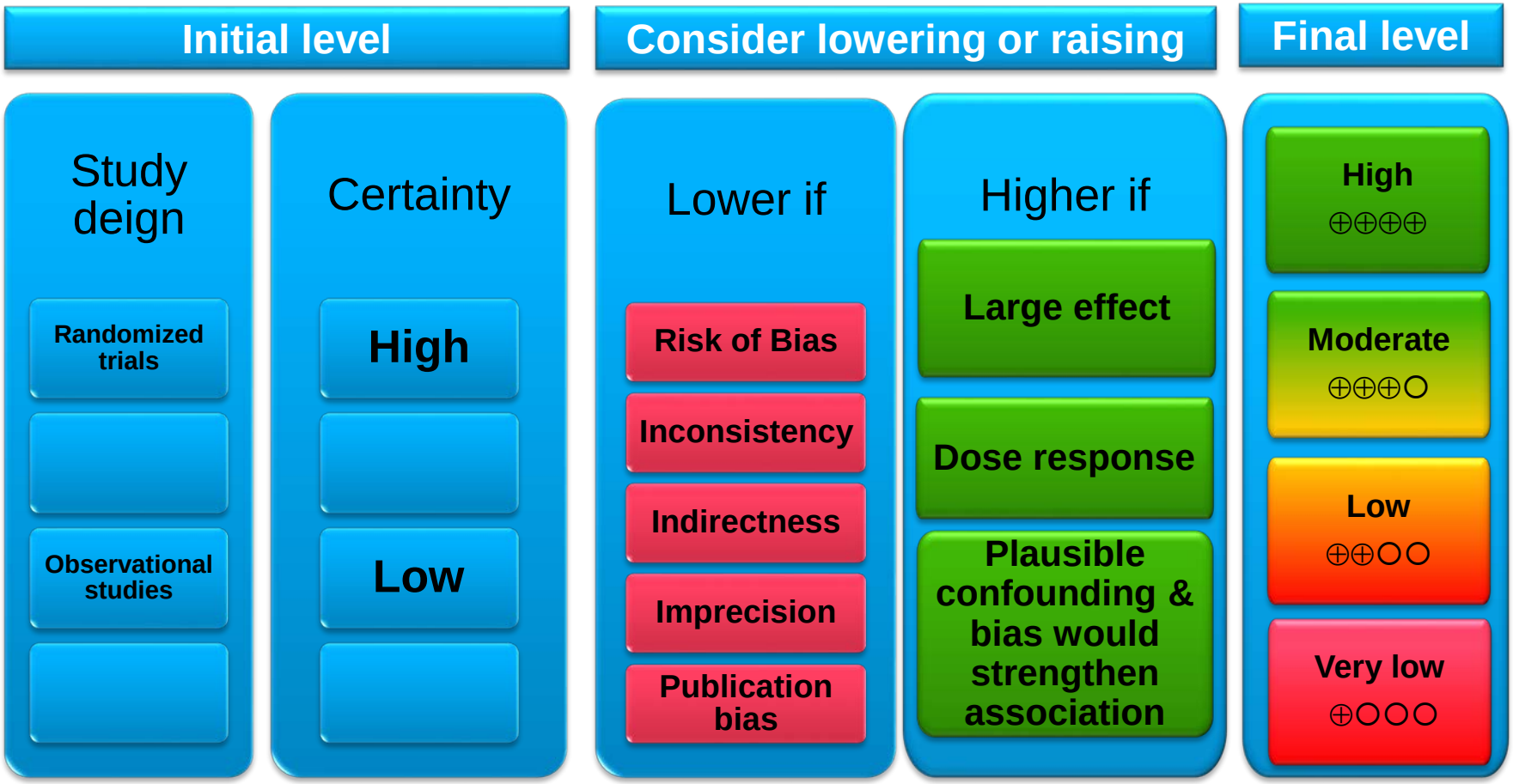
THE U.S. GRADE NETWORK

Scenario #3
Consider rating up when residual confounding or bias strengthens the association

- Well-done observational studies attempt to adjust analysis for known prognostic confounders
- Most of the time however, we cannot control for everything and residual confounding remains
- If this residual confounding is in the direction that strengthen the association, we may have higher confidence of the association
- "despite"

THE U.S. GRADE NETWORK

Level of certainty (quality, confidence, strength)



1

Other than the certainty of evidence, what factors should we consider?

2. Balance of benefits and harms
3. Patients' values
4. Costs and resources
5. Acceptability
6. Feasibility
7. Impact on equity

2. Balance of benefits and harms

- Babies born <32 weeks
- Lots of oxygen vs a little oxygen
- Quality of evidence: Amazing! The best you have ever seen
- Results: lots of oxygen ➔ lower mortality and higher risk of blindness
- Should the recommendation for oxygen amount be strong? Remember the evidence is amazing.

Should heparin be given to patients in ICU?

Outcome	effect
DVT/PE	6 fewer per 1,000 (2 fewer -10 fewer)
Major bleeding	3 more per 1,000 (1 more-5 more)
mortality	1 more per 1,000 (2 fewer to 4 more)

Desirable effects

Undesirable effects

Less strong recommendation

More strong recommendation

In your head:

$B_1 + B_2 - H_1 - H_2 = \text{Net benefit}$

$B_1 \times V_1 + B_2 \times V_2 - H_1 \times V_1 - H_2 \times V_2 = \text{Net benefit}$

3. Values

- Studies have shown that PCN reduces mortality from pneumococcal pneumonia by 20% (⊕⊕⊕O)
- Any reasons to not offer PCN to a patient you have just admitted with this condition?
  - *29 year old mother of 2, legal secretary wo became ill with cough and fever last night*
  - *91 year old nursing home resident who has been in vegetative state for 2 years, no visitors for the last 8 months*

What are values and preferences?

- Utilities and disutilities associated with a particular health state
- “a broad term that includes patient perspectives, beliefs, expectations, and goals for health and life, …”

4. Costs and resources

- Example:
  - Aspirin for secondary prevention of heart disease
  - Hepatitis C (sofosbuvir, simeprevir, $1,000/pill)

Other considerations (not always relevant)

5. Acceptability: circumcision in Africa to reduce HIV transmission

6. Feasibility: Proton beam radiotherapy in rural Oklahoma

6. Impact on equity: transplant vs chronic transfusion therapy to prevent stroke in children with sickle cell disease

thebmj | BMJ 2016;353:i2016 | doi: 10.1136/bmj.i2016

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines

Pablo Alonso-Coello,[1,2] Andrew D Oxman,[3] Jenny Moberg,[3] Romina Brignardello-Petersen,[2,4] Elie A Akl,[2,5] Marina Davoli,[6] Shaun Treweek,[7] Reem A Mustafa,[2,8] Per O Vandvik,[3] Joerg Meerpohl,[9] Gordon H Guyatt,[2,10] Holger J Schünemann,[2,10] the GRADE Working Group

Implications of a *strong* recommendation

- Population: Most people in this situation would want the recommended course of action and only a small proportion would not
- Health care workers: Most people should receive the recommended course of action
- Policy makers: The recommendation can be adapted as a policy in most situations

Implications of a *conditional* recommendation

- Population: The majority of people in this situation would want the recommended course of action, but many would not
- Health care workers: Be prepared to help people to make a decision that is consistent with their own values/decision aids and shared decision making
- Policy makers: There is a need for substantial debate and involvement of stakeholders

Summary

- GRADE is a framework for decisionmaking
- The two parts of GRADE:
  - Judging certainty in research evidence
  - Moving from evidence to action

