MUNI MED

GRADE

Masaryk University
GRADE Workshop

MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare

Cochrane Czech Republic

JBI The Czech Republic (Middle European) Centre for Evidence-Based Healthcare

Masaryk University GRADE

Session 1: Introduction to GRADE

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

GRADE centers

McMaster University GRADE Center, Canada
Lanzhou University GRADE Center, China
Barcelona GRADE Center, Spain
Freiburg GRADE Center, Germany
American University of Beirut GRADE Center, Lebanon
Lazio Region-ASL Rome GRADE Center, Italy
JBI Adelaide GRADE Centre, Australia
Melbourne GRADE Centre, Australia
Nottingham Ningbo GRADE Center, China
Masaryk University GRADE Centre, Czechia
Krakow University GRADE Center, Poland
Minds Tokyo GRADE Center, Japan

GRADE networks

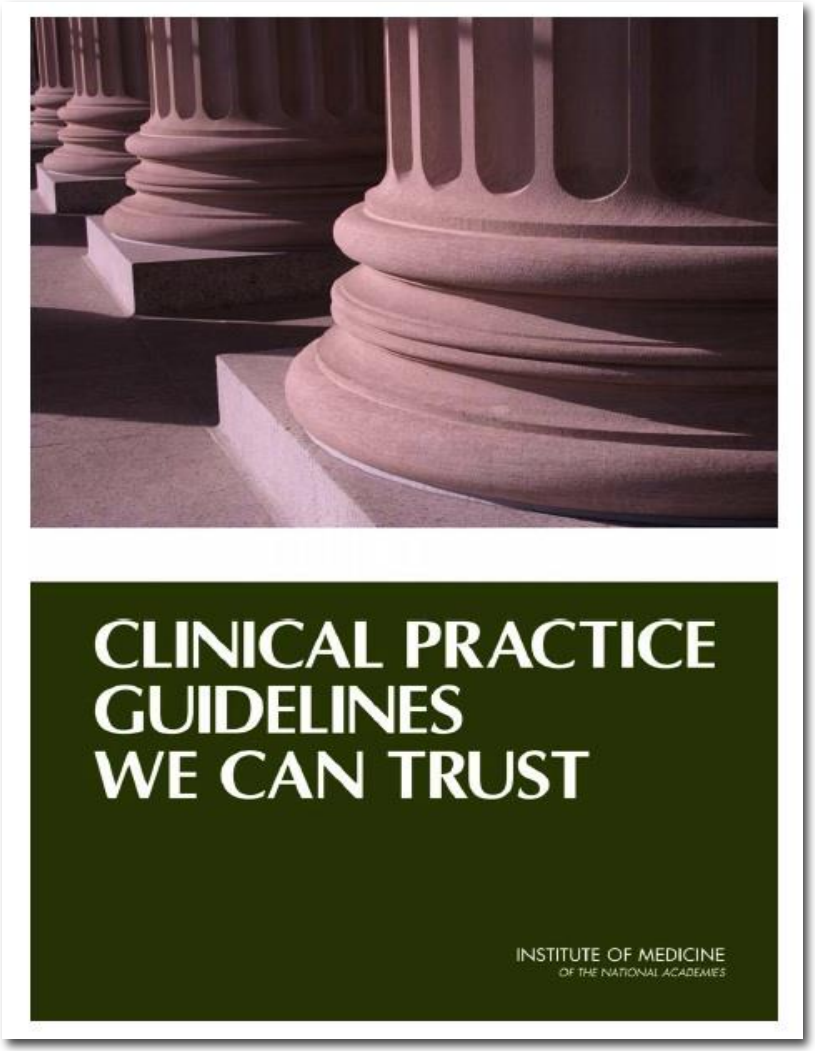
U.S. GRADE Network
Dutch GRADE Network
UK GRADE Network
South African GRADE Network

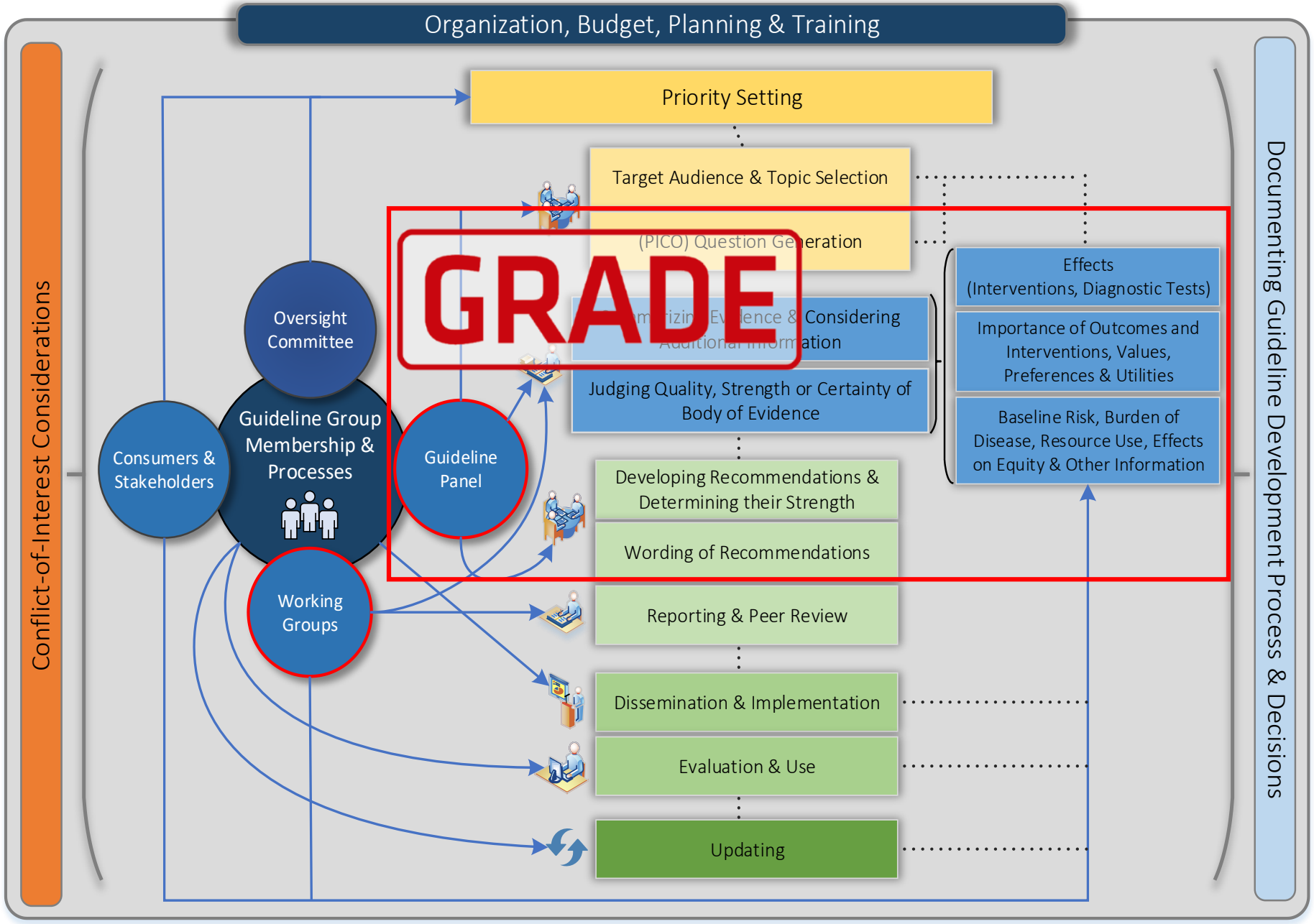
Groups and projects

DECIDE research project
Environmental health
Prognosis
Outcomes valuation
GRADE-CERQual
Diagnosis
Network meta-analysis
Observational studies
GRADE training and credentialing
Public health
Rare diseases
Evidence to decision
Equity
Algorithms and pathways
Modeling
Biosimilars
Animal studies
Complex interventions
GRADE NRS risk of bias integration

Institute of Medicine

- Be based on a **systematic review** of the existing evidence; **best available evidence**
- Be developed by a knowledgeable, multidisciplinary panel of experts and representatives from key affected groups; **diverse group**
- Consider important patient subgroups and patient preferences as appropriate; **patient values**
- Provide a clear explanation of the logical relationships between alternative care options and health outcomes, and provide **ratings** of both the **quality of evidence and the strength of recommendations**; and
- Be reconsidered and **revised as appropriate** when important **new evidence** warrants modifications of recommendations.
- Be based on an explicit and transparent process that **minimizes** distortions, biases, and **conflicts of interest**;





Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2014 Feb 18;186(3):E123-42.

Who are GRADE?

- Grading of Recommendations Assessment, Development and Evaluation (GRADE)
- International working group
- Endorsed by many EBHC organisations
- Website: http://www.gradeworkinggroup.org/

GRADE working group

After over 20 years of increasing confusion, GRADE developed a unifying, transparent and sensible system for grading the certainty of evidence and making decisions

- WHO, NICE, CDC, AHRQ, JBI, Cochrane, professional societies, academic institutions since 2000 – over 120 use GRADE
- For systematic reviews, HTA and guidelines
- International & diverse contributors (>800), regular workshops at GIN, Cochrane, WHO, JBI
- 2004/2008 BMJ series; 2011 JCE series: > 60,000 cites
- Various other publications (incl. GRADE Handbook)
- IT applications GRADEpro GDT

History of GRADE

- Began as an informal working group in 2000
- 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

Over 100 organisations From 19 countries

JBI

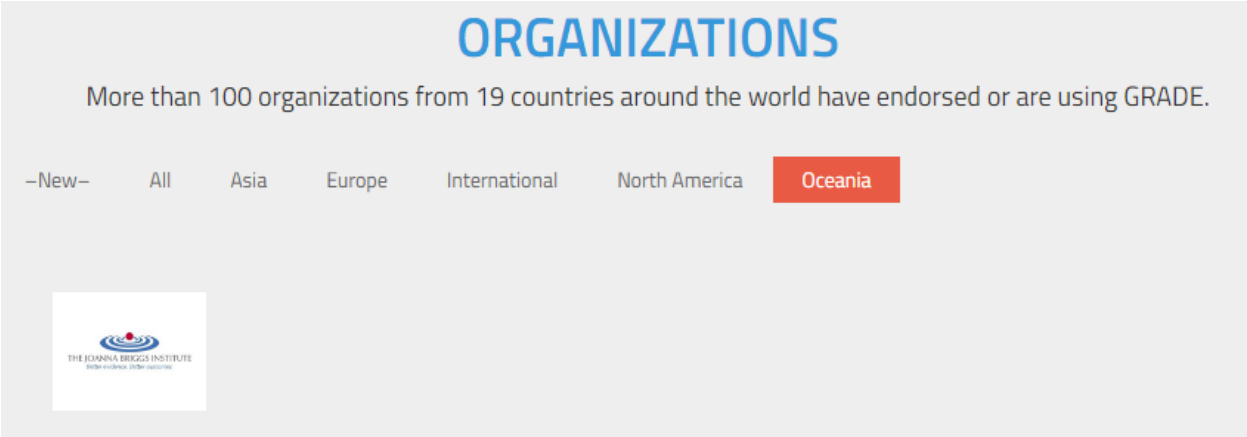
The Czech Republic (Middle European) Centre for Evidence-Based Healthcare

Masaryk University

GRADE

Systematic Reviewers

- JBI and Cochrane explicitly endorse the use of GRADE methods and require GRADE



Methodological Expectations of Cochrane Intervention Reviews (MECIR) 25

Assessing the quality of evidence and summarizing the findings

Standard		Rationale and Elaboration
C74	Assessing the quality of the body of evidence	Mandatory
Use the five GRADE considerations (risk of bias, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of the body of evidence for each outcome, and to draw conclusions about the quality of evidence within the text of the review.		GRADE is the most widely used approach for summarizing confidence in effects of interventions by outcome across studies. It is preferable to use the online GRADEpro tool, and to use it as described in the help system of the software. This should help to ensure that author teams are accessing the same information to inform their judgments. Ideally, two people working independently should assess the quality of the body of evidence and reach a consensus view on any downgrading decisions. The five GRADE considerations should be addressed irrespective of whether the review includes a 'Summary of findings' table. It is helpful to draw on this information in the Discussion, in the Authors' conclusions and to convey the certainty in the evidence in the Abstract and Plain language summary. See *Handbook* 12.2

In the Czech Republic – Guideline Developers

In the Czech Republic – Guideline Developers

Souhrn a klasifikace kvality vědeckých důkazů dle přístupu a metodiky GRADE

Přístup GRADE (*The Grading of Recommendations Assessment, Development and Evaluation*) umožňuje třídit vědecké důkazy, zhodnotit jejich kvalitu a tvořit doporučení (Atkins et al., 2004; Guyatt, Oxman, Kunz, Vist, et al., 2008; Guyatt, Oxman, Vist, et al., 2008).

To be trustworthy evidence informed guidelines will:

- - Be ADOLOPED (If they use GRADE methods)
- - Newly developed following Czech national methodology which is baased on the GRADE
- Be informed by well contucted systematic reviews
- Consider the body of evidence for each outcome (including the quality of that evidence) and other factors that infuence the process of making recommendations including benefits and harms, values and preferences, resource use and acceptability

a) Souhrn vědeckých důkazů dle přístupu GRADE

Závěrečným bodem systematického review či HTA dokumentující pouze vědecké důkazy je tzv. souhrn vědeckých důkazů *„Summary of evidence“*, hodnocení kvality každého každého výstupu *„quality rating for each outcome“* a odhad účinnosti *„estimate of effect“*. Pro tvůrce KDP a HTA, která obsahují doporučení pro tvůrce politik, představuje souhrn vědeckých

Zde předložený dílčí metodický postup byl vypracován jako část celkové metodiky s cílem vytvořit pracovní materiál pro tvůrce KDP. Během tvorby pilotních KDP lze předpokládat částečné změny či korekce metodiky finální verze metodiky projde finálním recenzím řízením.

Verze 2.0, duben 2018
Součást projektu: Klinické doporučené postupy
Autoři: PhDr. Miloslav Klugar, Ph.D., doc. PhDr. Andrea. Pokorná, Ph.D., Mgr. Jitka Klugarová, Ph.D., MUDr. Radim Ličeník, Ph.D., RNDr. Jan Mužík, Ph.D., Mgr. Dana Dolanová, Ph.D., RNDr. Martin Komenda, Ph.D., RNDr. Jakub Gregor Ph.D., doc. RNDr. Ladislav Dušek, Ph.D.

MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare
Cochrane Czech Republic
JBI

Session 2: 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

Clinical Practice guidelines & the origin of evidence appr

Effectiveness of intervention

The effectiveness of intervention was graded according to the quality of the evidence obtained, as follows:

I: Evidence obtained from at least one properly randomized controlled trial.

II-1: Evidence obtained from well designed cohort or case–control analytic studies, preferably from more than one centre or research group.

II-2: Evidence obtained from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin in the 1940s) could also be regarded as this type of evidence.

III: Opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees.

Classification of recommendations

On the basis of these considerations the task force made a clear recommendation for each condition as to whether it should be specifically considered in a periodic health examination. Recommendations were classified as follows:

A: There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination.

B: There is fair evidence to support the recommendation that the condition be specifically considered in a periodic health examination.

C: There is poor evidence regarding the inclusion of the condition in a periodic health examination, and recommendations may be made on other grounds.

D: There is fair evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.

E: There is good evidence to support the recommendation that the condition be excluded from con-

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’

Why GRADE?

Level	Adjustment	Description by Type of Questio
1	a	*SR (with homogeneity) of prospective cohort s*
	b	*Prospective cohort study with good follow-up*
	c	*All or none case-series*
2	a	*SR (with homogeneity) of 2b and better studie*
	b	*Retrospective cohort study, or poor follow-up*
	c	*Ecological studies*
3	a	*SR (with homogeneity) of 3b and better studie*
	b	*Non-consecutive cohort study, or very limited population*
4		*Case-series or superseded reference standards*
5		*Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"*

Adapted from: Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001), http://www.cebm.net/

Table I – Classification of the procedures

Class I
Conditions for which conclusive evidence exists, or in its absence, general consensus that the procedure is useful or effective, or both.

Class II
Conditions for which conflicting evidence or divergence of opinion, or both, exists in regard to usefulness/efficacy of the procedure.

Class IIA
Weight or evidence/opinion favoring usefulness/efficacy.

Class IIB
Usefulness/efficacy less well-established by evidence or opinion.

Class III
Conditions for which evidence or consensus, or both, exists that the procedure is not useful/efficient, and, in some cases, it may even be noxious.

Adapted from the criteria used in the guidelines of the American College of Cardiology/American Heart Association.

Table: Grade of recommendation and levels of evidence.

n	Levels of evidence	Types of study
	1	
	1a	Systematic review of homogeneous RCTs with good methodological quality
	1b	Individual RCTs with narrow confidence intervals
	1c	Uncontrolled studies (dramatic findings)
	2	
	2a	Systematic review of cohort studies (with homogeneity)
	2b	Individual cohort studies (including low quality RCTs, e.g. <80% follow-up)
	2c	Uncontrolled cohort studies/ecological studies
	3	
	3a	Systematic review of case control studies (with

III-2	Evidence obtained from comparative studies with concurrent controls and allocation not randomised (cohort studies), case–control studies, or interrupted time series without a control group.	B
III-3	Evidence obtained from comparative studies with historical control, two or more single-arm studies, or interrupted time series with a parallel control group.	C
IV	Evidence obtained from case series, either post-test or pre-test and post-test.	C
		D

Rich body of quality RCT

Limited body data or high-c non-RCT d

Limited evid

No evidence av

Table 2. American Heart Association Classification Of Recommendations And Level Of Evidence

Levels of Evidence	
Level A	Multiple populations evaluated. Data derived from multiple randomized controlled trials or meta-analyses.
Level B	Limited populations evaluated. Data derived from a single randomized trial or nonrandomized studies.
Level C	Very limited populations evaluated; only consensus opinion of experts, case studies, or standard of care

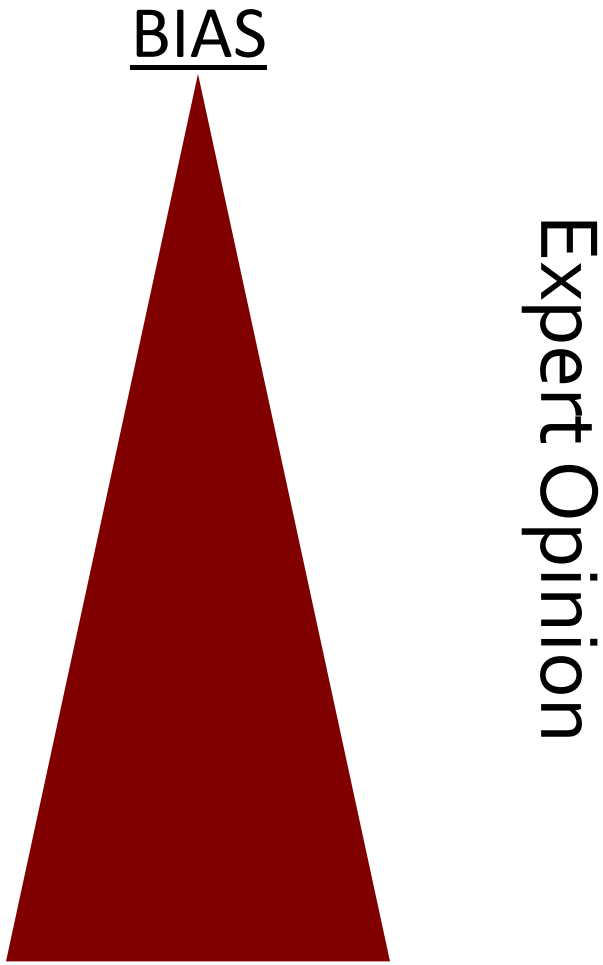
Classes of Evidence	
Class I	Benefit >>> Risk
Class IIa	Benefit >> Risk
Class IIb	Benefit ≥ Risk

Quality of evidence

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations

Expert Opinion



Schünemann & Bone, 2003

‘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’

Guyatt, Gordon, Victor Montori, Holger Schünemann, and Paul Glasziou. "When Can We Be Confident about Estimates of Treatment Effects?." *The Medical Roundtable General Medicine Edition* (2015).

Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group

David Atkins, Martin Eccles, Signe Flottorp, Gordon H Guyatt, David Henry, Suzanne Hill, Alessandro Liberati, Dianne O'Connell, Andrew D Oxman, Bob Phillips, Holger Schünemann, Tessa Tan-Torres Edejer, Gunn E Vist ✉, John W Williams Jr and The GRADE Working Group[3]

BMC Health Services Research 2004 **4**:38 | **DOI:** 10.1186/1472-6963-4-38 | © Atkins et al; licensee BioMed Central Ltd. 2004
Received: 23 January 2004 | **Accepted:** 22 December 2004 | **Published:** 22 December 2004

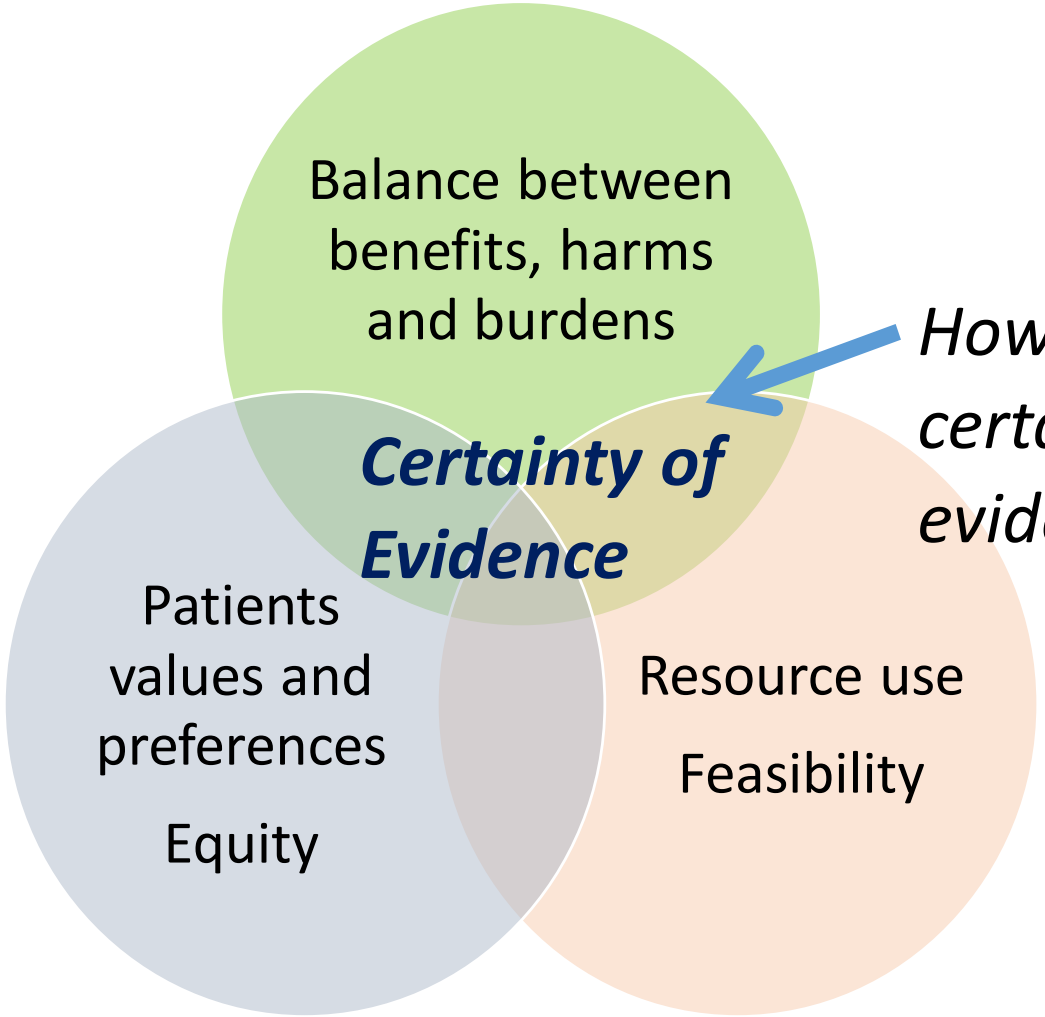
Open Peer Review reports

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



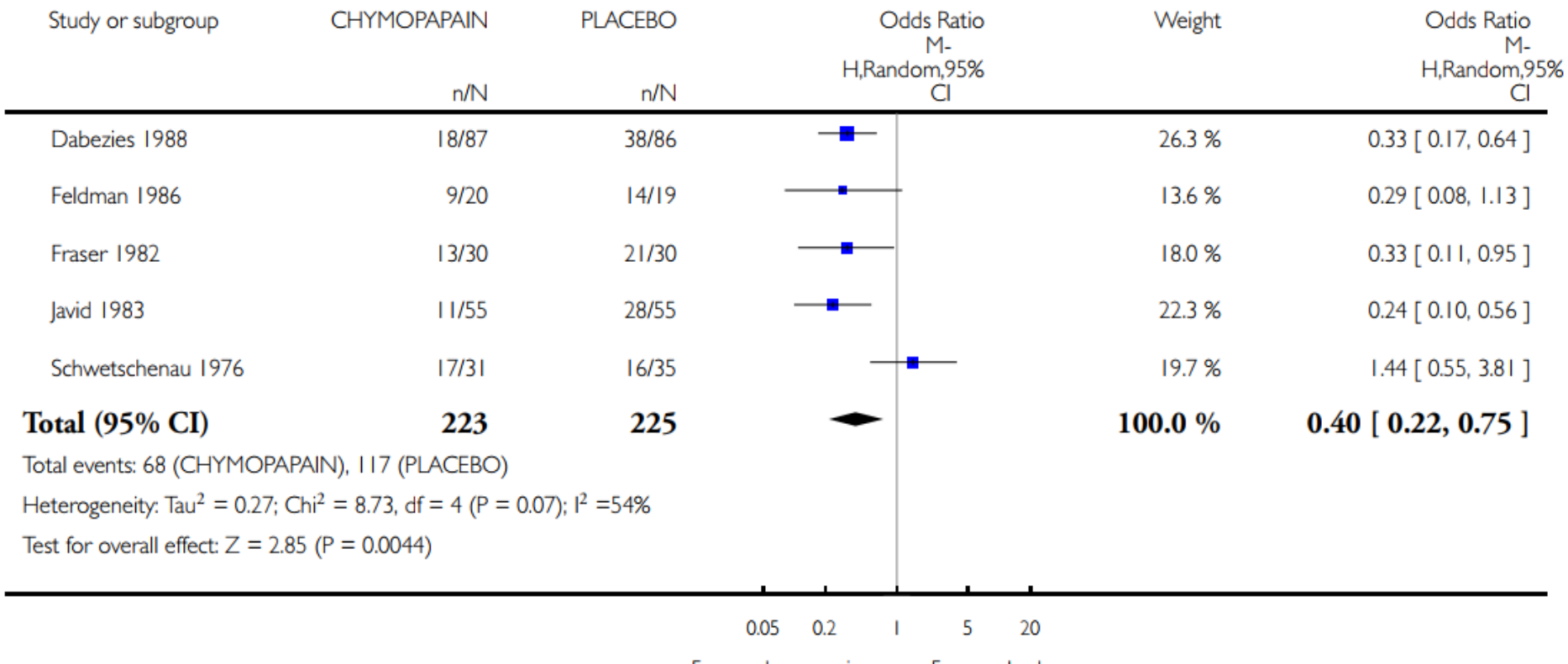
How do we determine certainty of the evidence?

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?

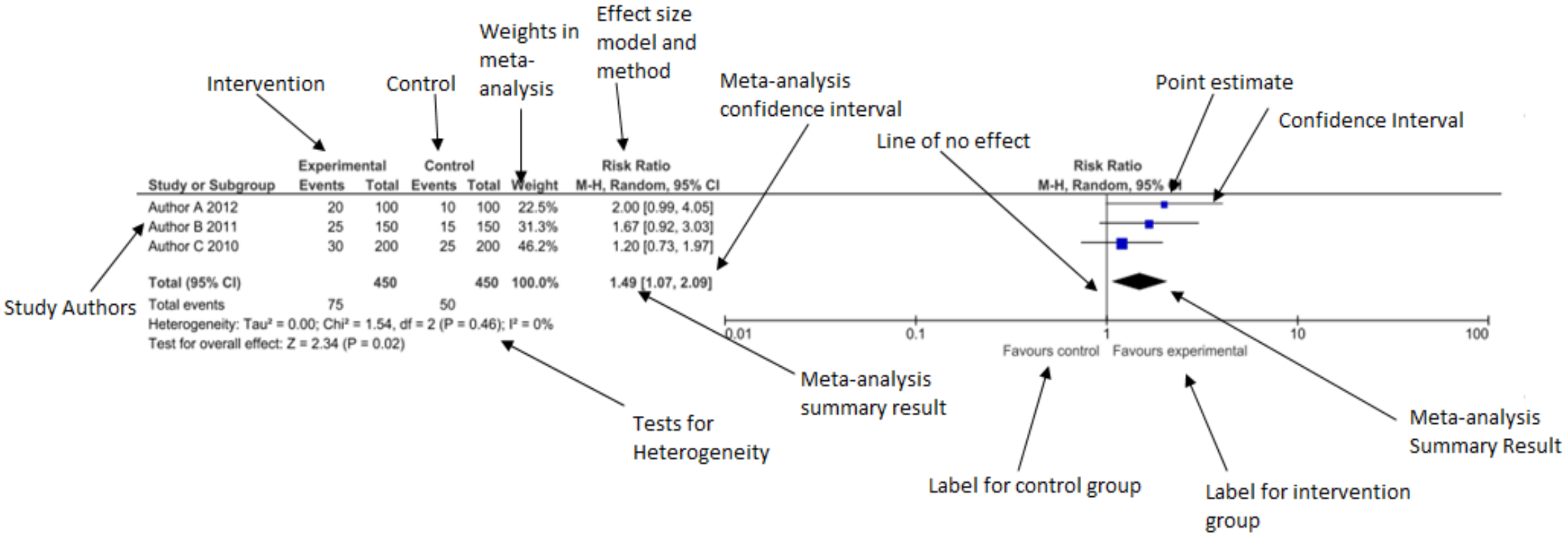
Activity 1: Example meta-analysis discussion

- From the example provided, what information would increase or decrease your confidence in these results?



Gibson JNA, Waddell G. Surgical interventions for lumbar disc prolapse. Cochrane Database of Systematic Reviews 2007, Issue 2. Art. No.: CD001350. DOI: 10.1002/14651858.CD001350.pub4.)

Meta-analysis forest plot "referesher"



Discussion results

- Increase:
- Heterogenita nižší
- Novější studie
- fulltexty studií – srovnatelnost intervence a populace, zdrav. prostředí
- Metodologie meta-analýzy
- Vyšší počet pacientů
- Více studií
- Efekt – velikost
- CI užší

- Decrease:
- Střety zájmů
- Kde publikováno, kým

GRADE

- Decrease
  - Limitations in study design and execution (risk of bias)
  - Indirectness (i.e applicability, generalisability, transferability etc)
  - Inconsistency (heterogeneity)
  - Imprecision (uncertainty)
  - Publication bias

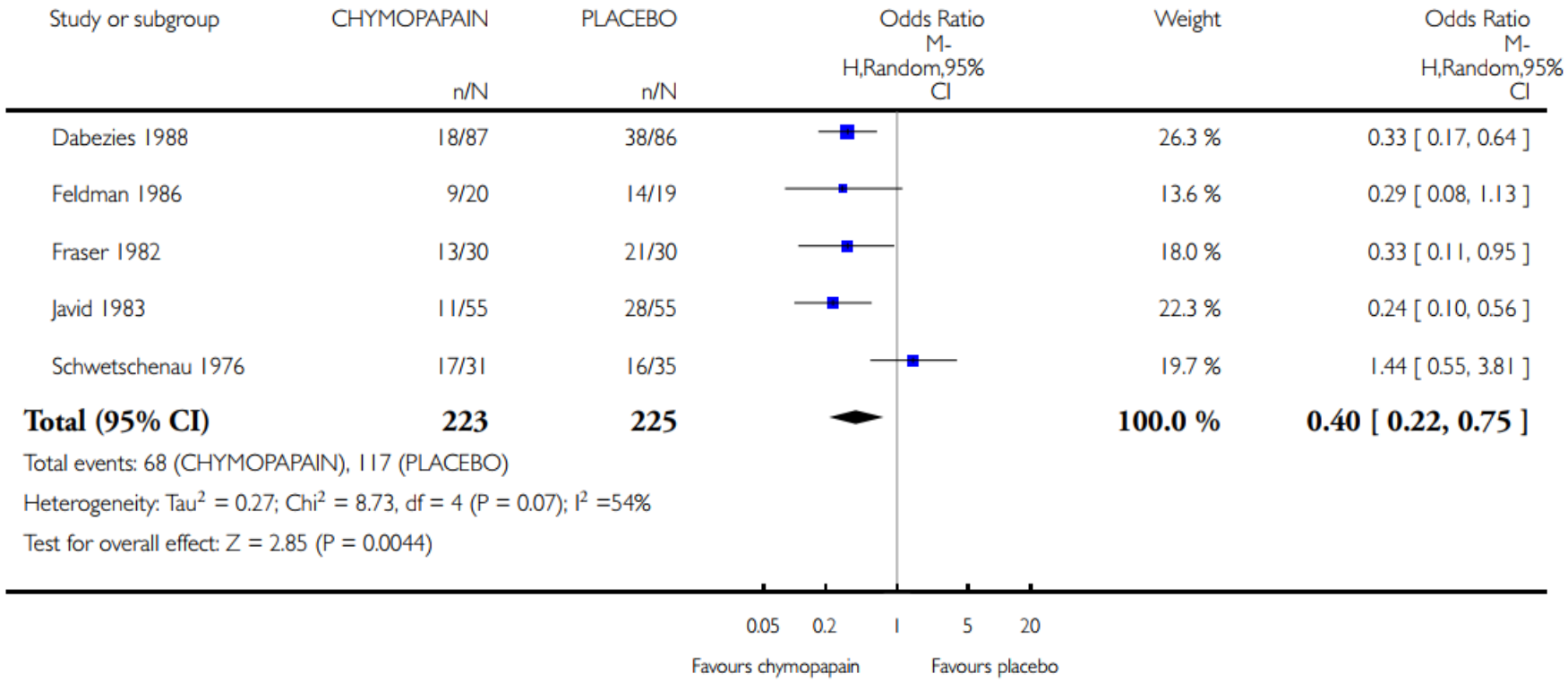
- Increase
  - Large, consistent, precise effect
  - All plausible biases underestimate the effect
  - Dose response effect

Session 3: Introduction to the GRADE approach

MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare
Cochrane Czech Republic
JBI
The Czech Republic (Middle European) Centre for Evidence-Based Healthcare
Masaryk University
GRADE

Key principle

- Important to communicate
  - Results
  - Our certainty in these results?



Certainty of evidence

How confident in the research?

- Are the research studies well done? Risk of bias
- Are the results consistent across studies ? Inconsistency
- How directly do the results relate to our question? Indirectness
- Is the effect size precise - due to random error? Imprecision
- Are these all of the studies that have been conducted? Pub. Bias
- Is there anything else that makes us particularly certain? Large effects, worst case scenario predictors still strong conclusions, exposure-effect relation

Magnitude of Effect (results)

Certainty/quality/ confidence in the evidence

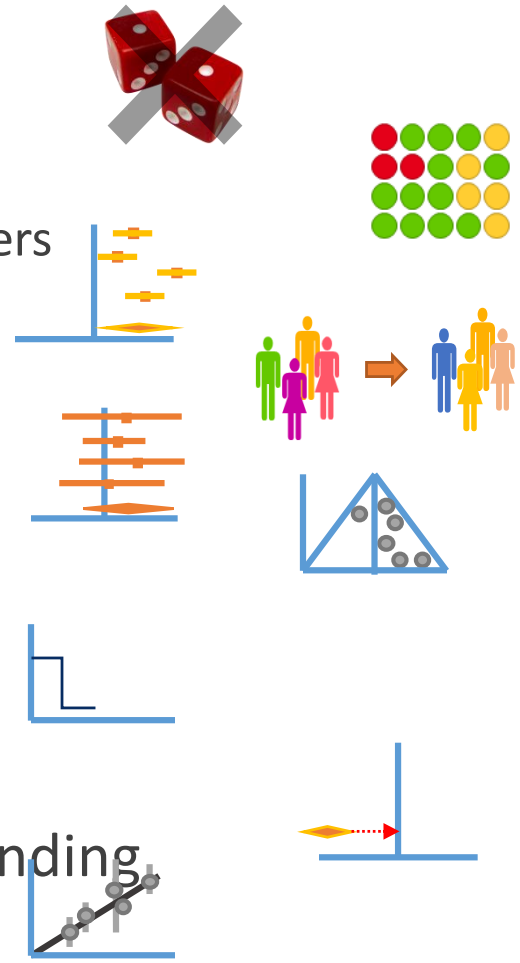
Figure 1. *Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street*

Determinants of certainty in a body of evidence: GRADE

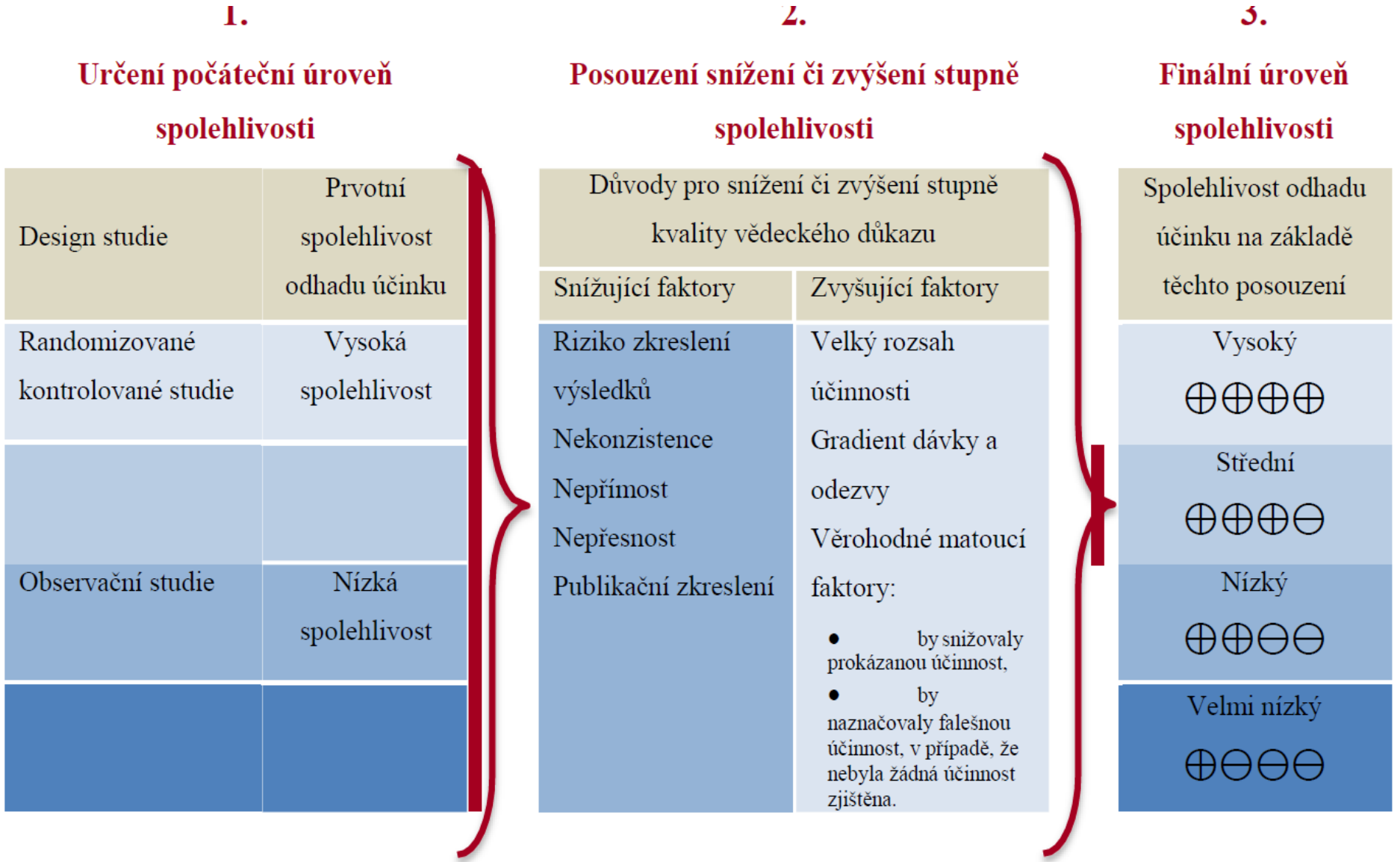
- **A body of evidence starts as: high | ⊕⊕⊕⊕**

- **5 factors that can lower certainty**
  1. Risk of bias criteria
     - Lack of randomization (observational studies) lowers confidence to low
  2. Inconsistency *(or heterogeneity)*
  3. Indirectness *(PICO and applicability)*
  4. Imprecision
  5. Publication bias

- **3 factors can increase certainty**
  1. large magnitude of effect
  2. opposing plausible residual bias or confounding
  3. dose-response gradient



Hodnocení jistoty důkazů



Lowering certainty in Studies

Table: GRADE's approach to rating certainty/quality of evidence (aka confidence in effect estimates)

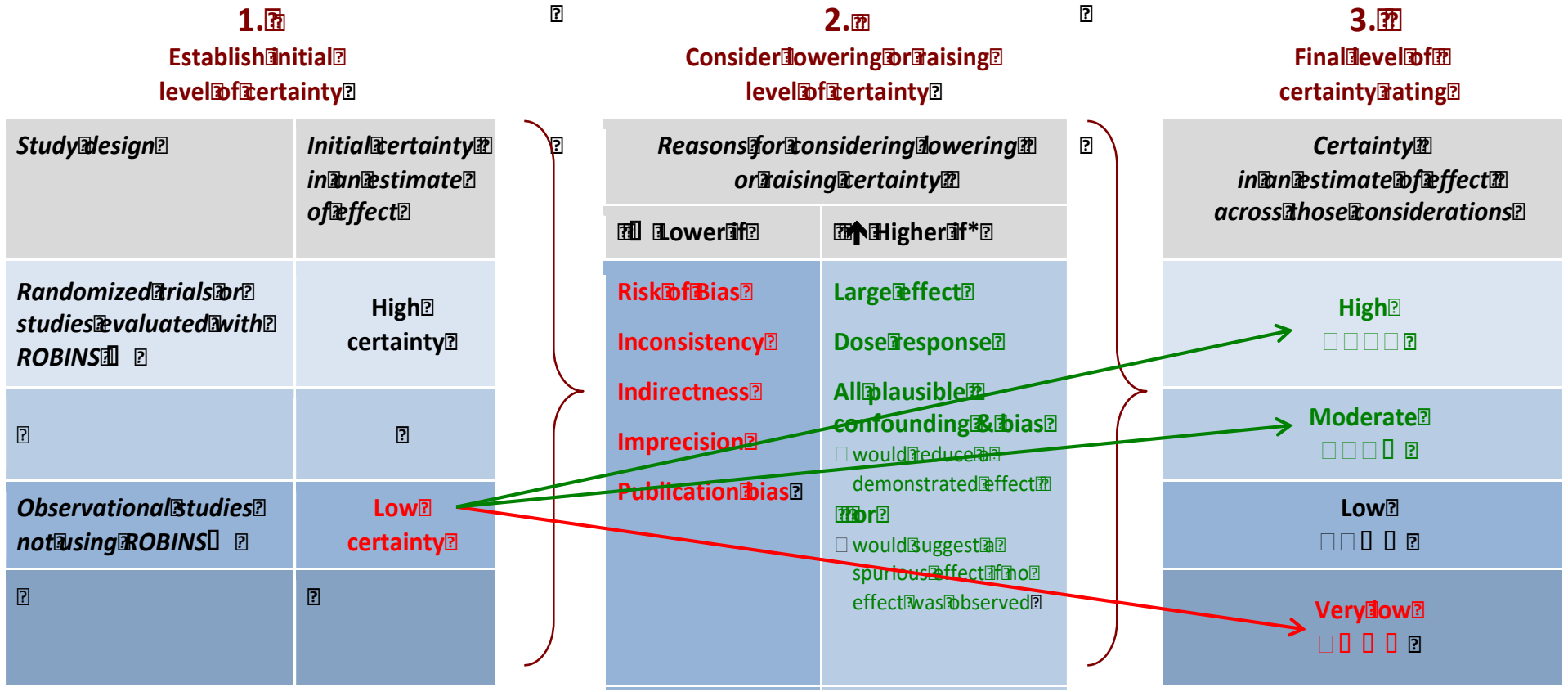
For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

* upgrading criteria are usually applicable to observational studies only.

Altering certainty of observational studies not assessed with ROBINS

Table: GRADE's approach to certainty/rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)



*upgrading criteria are usually applicable to observational studies only.

GRADEpro GDT

Formulate question

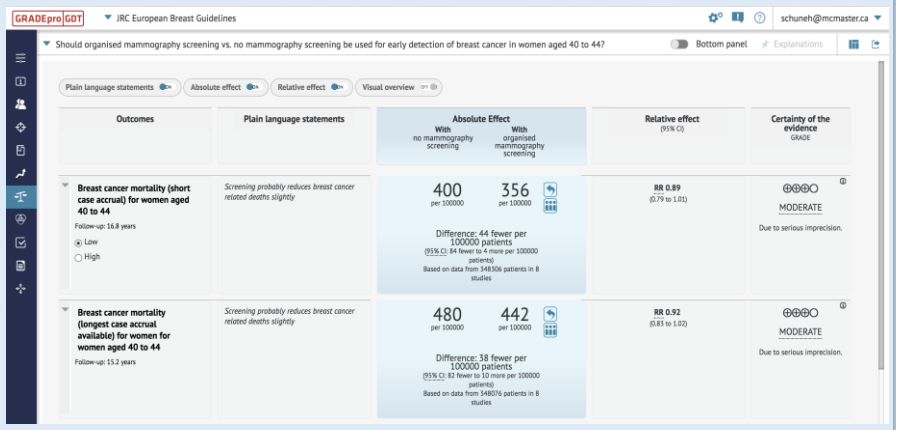
Assess single studies

Synthesize and Create **evidence profile or Summary of Findings** Table with GRADEpro

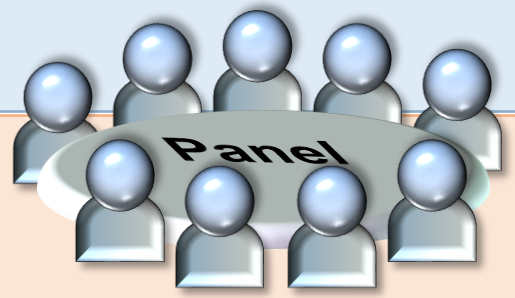
Rate certainty of evidence for each outcome

P
I
C
O

Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important



Evidence synthesis (systematic review/HTA)



Recommendation/Decision

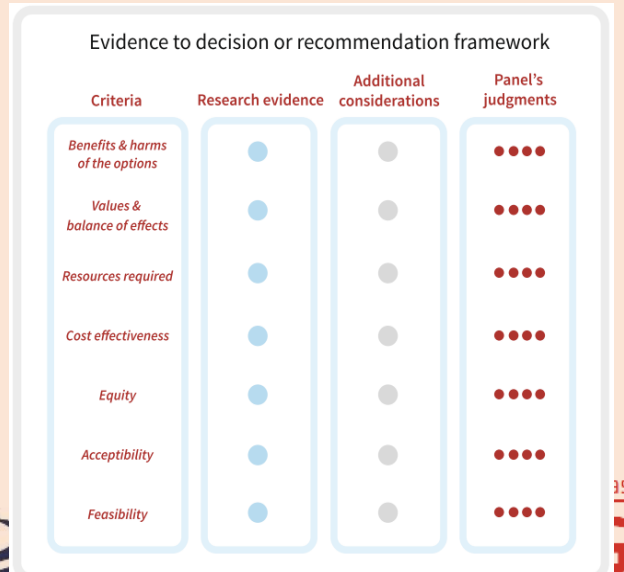
Guideline recommendation


AMERICAN GASTROENTEROLOGICAL ASSOCIATION

Colorectal Cancer Screening and Surveillance: Clinical Guidelines and Rationale–Update Based on New Evidence

Grade recommendations (Evidence to Recommendation)

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



MUNI MED

Cochrane Czech Republic

JBI

asaryk University

GRADE

Framing questions and selecting outcomes

- Use PICO for your SR or Guideline question/s
- Include a range of outcomes, addressing benefit and harms
  - SRs often miss harms, guideline panels need to consider all outcomes for decision making
- Should include all potential patient-important outcomes
  - Classify outcomes regarding importance for decision making:
    - Critical
    - Important but not critical
    - Of limited importance

rating scale:								
1	2	3	4	5	6	7	8	9
of **least** importance								of **most** importance
of limited importance for making a decision			important, but not critical for making a decision			Critical for making a decision		

Activity 2: Classifying outcomes

- Turn to your workbook and begin activity 2.

Have you thought about....?

- What would be important for someone making a decision?
- Have you considered benefits as well as harms?
- What outcomes are likely included in studies, and what may be missed?
- What outcomes should be included in a summary of findings table or evidence profile?

Activity 2: Example outcomes

- Outcomes
  - Mortality (all cause)
  - Infection (deep sternal or other)
  - Length of stay
  - Time on mechanical ventilation
  - Acute renal failure
  - Stroke
  - Hypoglycaemic episode
  - Health related quality of life
  - Weight gain
- Ranking

Activity 2: Example outcomes

- Outcomes
  - Mortality (all cause)
  - Symptomatic VTE
  - Major bleeding
  - Minor bleeding
  - Thrombocytopenia
  - Quality of life
- Ranking

GRADE is outcome-centric

- Previously, rankings were done on a study basis
- GRADE evaluations focus on the evidence relating to an outcome
- Different outcomes have different rankings

GRADE is outcome-centric

3 1a 4 1

Other systems

GRADE

GRADEpro GDT

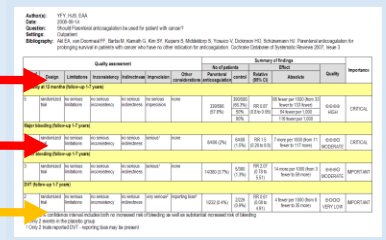
Formulate question

P
I/E
C
O

Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important

Assess single studies

Synthesize and Create
evidence profile or Summary of Findings
Table with GRADEpro



Rate certainty of evidence for each outcome

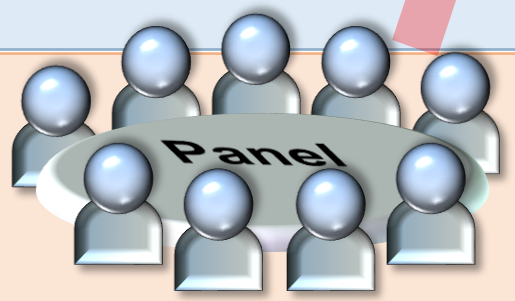
High
Moderate
Low
Very low

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

Evidence synthesis (systematic review/HTA)

Input?



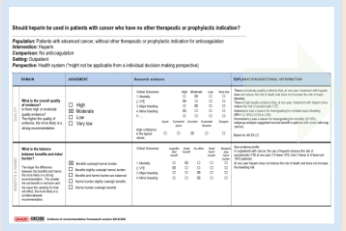
Recommendation/Decision

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

By balancing **consequences** (evidence to recommendations):
- ❑ **Certainty of evidence**
- ❑ **Values and preferences (utilities)**
- ❑ **Balance benefits/harms**
- ❑ **Resource use (cost)**

Grade overall Certainty of evidence across outcomes

EtD framework



Recommendation /Decision

Formulate Recommendations (↓↑ | ⊕…)
"The panel recommends that ….should..."
"The panel suggests that ….should..."
"The panel suggests to **not** ..."
"The panel recommends to **not**..."

"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."* JCE, 2011

Session 4: Determining quality (certainty) of the evidence

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

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

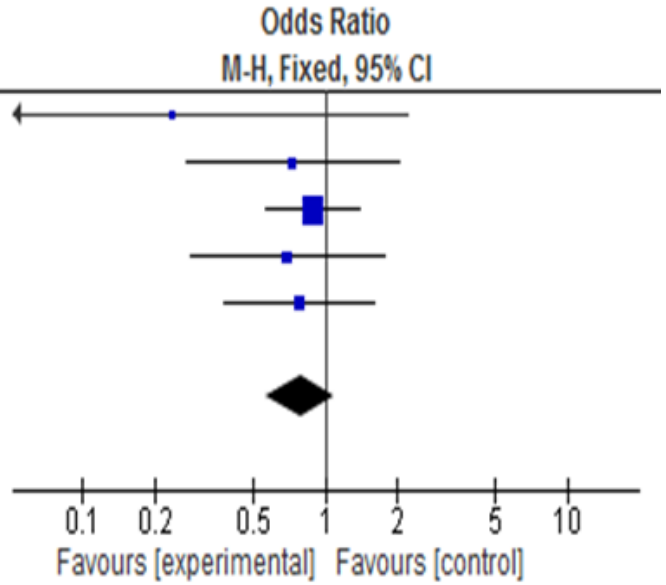
What is the effect?

Mortality

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Asida 2012	1	50	4	50	4.7%	0.23 [0.03, 2.18]
Desai 2012	7	91	10	98	10.7%	0.73 [0.27, 2.02]
Gandhi 2007	54	185	59	186	50.4%	0.89 [0.57, 1.38]
Lazar 2011	12	40	16	42	13.2%	0.70 [0.28, 1.75]
Rujirojindakul 2014	17	99	21	100	20.9%	0.78 [0.38, 1.59]
Total (95% CI)		**465**		**476**	**100.0%**	**0.79 [0.57, 1.09]**
Total events	91		110			

Heterogeneity: $Chi^2 = 1.50$, $df = 4$ ($P = 0.83$); $I^2 = 0\%$

Test for overall effect: $Z = 1.41$ ($P = 0.16$)



The odds for mortality are 0.79 (95% CI 0.57-1.09) of that in the experimental group compared to the control group

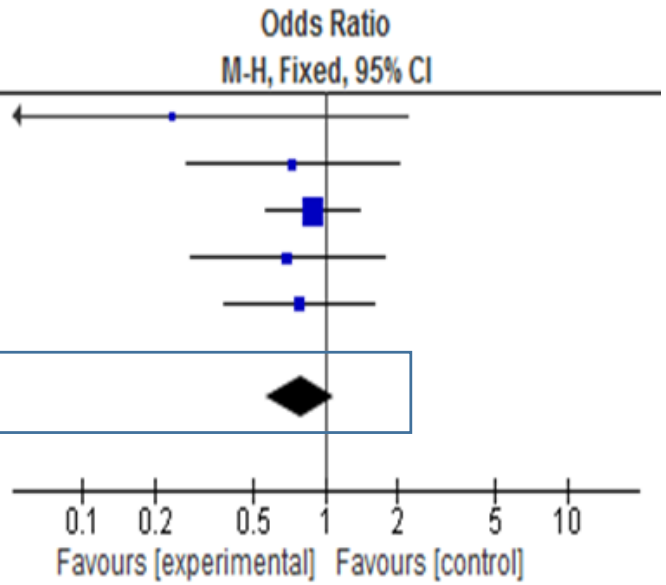
Misinterpretation of effects

Mortality

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Asida 2012	1	50	4	50	4.7%	0.23 [0.03, 2.18]
Desai 2012	7	91	10	98	10.7%	0.73 [0.27, 2.02]
Gandhi 2007	54	185	59	186	50.4%	0.89 [0.57, 1.38]
Lazar 2011	12	40	16	42	13.2%	0.70 [0.28, 1.75]
Rujirojindakul 2014	17	99	21	100	20.9%	0.78 [0.38, 1.59]
Total (95% CI)		**465**		**476**	**100.0%**	**0.79 [0.57, 1.09]**
Total events	91		110			

Heterogeneity: $Chi^2 = 1.50$, $df = 4$ ($P = 0.83$); $I^2 = 0\%$

Test for overall effect: $Z = 1.41$ ($P = 0.16$)



Contents Search - Search - Go

Front page
Handbook information
Part 1: Cochrane reviews
Part 2: General methods for Cochrane revi
5 Defining the review question and de
6 Searching for studies
7 Selecting studies and collecting data
8 Assessing risk of bias in included stu
Box 12.8.a: The Cochrane Applicab
12.9 References
Part 3: Special topics

Home > Part 2: General methods for Cochrane reviews > 12 Interpreting results and drawing conclusions > 12.4 Interpreting results of statistical analyses > 12.4.2 P values and statistical significance

12.4.2 P values and statistical significance

A P value is the probability of obtaining the observed effect (or larger) under a 'null hypothesis', which in the context of Cochrane reviews is either an assumption of 'no effect of the intervention' or 'no differences in the effect of intervention between studies' (no heterogeneity). Thus, a P value that is very small indicates that the observed effect is very unlikely to have arisen purely by chance, and therefore provides evidence against the null hypothesis. It has been common practise to interpret a P value by examining whether it is smaller than particular threshold values. In particular, P values less than 0.05

P values are commonly misinterpreted in two ways. First, a moderate or large P value (e.g. greater than 0.05) may be misinterpreted as evidence that "the intervention has no effect". There is an important difference between this statement and the correct interpretation that "there is not strong evidence that the intervention has an effect". To avoid such a misinterpretation, review authors should always examine the effect estimate and its 95% confidence interval, together with the P value. In small studies or small meta-analyses it is common for the range of effects contained in the confidence interval to include both no intervention effect and a substantial effect. Review authors are advised not to describe results as 'not statistically significant' or 'non-significant'.

an effect of a particular magnitude will be greater (the P value will be smaller) in a larger study than in a smaller study.

The second misinterpretation is to assume that a result with a small P value for the summary effect estimate implies that an intervention has an important benefit. Such a misinterpretation is more likely to occur in large studies, such as meta-analyses that accumulate data over dozens of studies and thousands of participants. The P value addresses the question of whether the intervention effect is precisely nil; it does not examine whether the effect is of a magnitude of importance to potential recipients of the intervention. In a large study, a small P value may represent the detection of a trivial effect. Again, inspection of the point estimate and confidence interval helps correct interpretations (see Section 12.4.1).

ASA News
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®
732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • www.twitter.com/AmstatNews

The statement's six principles, many of which address misconceptions and misuse of the *p*-value, are the following:

1. *P-values can indicate how incompatible the data are with a specified statistical model.*
2. *P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.*
3. *Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.*
4. *Proper inference requires full reporting and transparency.*
5. *A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.*
6. *By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.*

Key takeaways

1. The CI obtained provides a range of uncertainty
2. The point estimate and confidence interval provide information to assess the clinical usefulness of the intervention.
3. ‘Not statistically significant’ does not equal ‘no effect’
4. If review authors decide to present a P value with the results of a meta-analysis, they should report a precise P value, together with the 95% confidence interval**. (Cochrane Handbook)**

GRADEing the evidence

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

GRADE domains	Rating (circle one)	Footnotes (explain judgements)	Certainty of evidence (Circle one)
Risk of Bias	No serious (-1) very serious (-2)		⨁⨁⨁⨁ High □ ⨁⨁⨁◯ Moderate ⨁⨁◯◯ Low □ ⨁◯◯◯ Very Low
Inconsistency	No serious (-1) very serious (-2)		
Indirectness	No serious (-1) very serious (-2)		
Imprecision	No serious (-1) very serious (-2)		
Publication Bias	Undetected Strongly suspected (-1)		
Other (upgrading factors,	Large effect (+1 or +2) Dose response (+1)		

Session 5: Study limitations (Risk of bias)

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

Bias

- A bias is a systematic error, or deviation from the truth, in results or inferences (Higgins & Altman, 2008)
- Bias in research may lead to misleading estimates of effect
- Studies may be at risk of bias due to issues with the conceptualization, design, conduct or interpretation of the study
- There are many different types of bias that can arise in research

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)

Addressing Bias

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

Type of bias	Description	Relevant domains in Cochrane's 'Risk of bias' tool
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	•Sequence generation. •Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	•Blinding of participants and personnel. •Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	•Blinding of outcome assessment. •Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	•Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	•Selective outcome reporting
Other bias	Stopping trial early Invalid outcome measures Cluster or crossover trial issues	•Other types of bias

Overall Risk of Bias

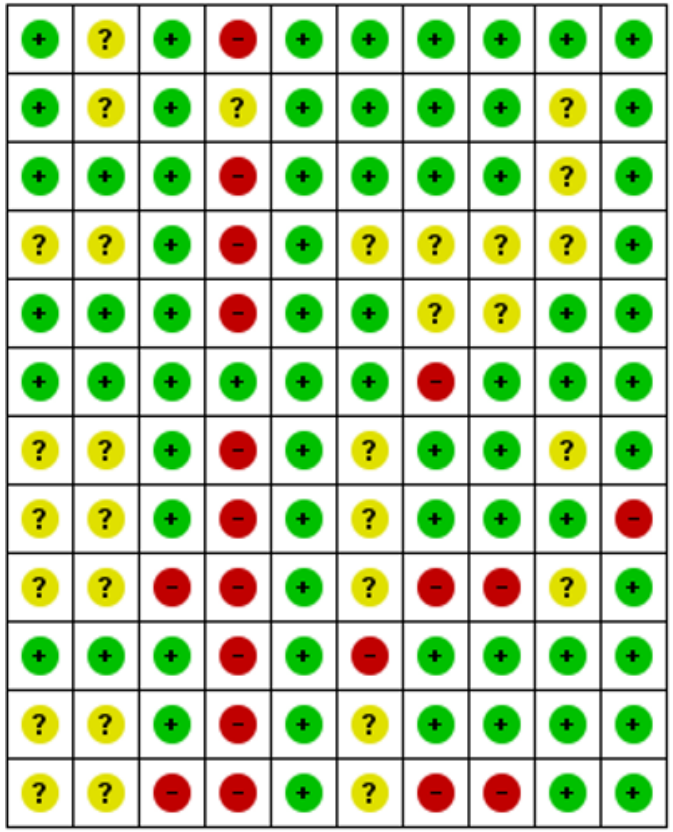
- Use the risk of bias assessment from all studies to determine overall risk of bias
- This can be difficult!

Activity 3

- Discuss with your partner the example in your workbook and come up with the answer. We will then discuss as a wider group.

So how should we do it?

- Can you simply count the number of green dots compared to yellow and red?
  - Rather than an average, consider judiciously the contribution of each study
- What about weighting?
  - Risk of bias of studies providing more weight to the analysis should be considered more
- Should trials with high risk of bias be excluded?
  - Potentially, although may be implications for imprecision



Key principles

- We suggest the following principles:
  - In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies, is warranted.
  - The judicious consideration requires evaluating the extent to which each trial contributes toward the estimate of magnitude of effect. This contribution will usually reflect study sample size and number of outcome events – larger trials with many events will contribute more, much larger trials with many more events will contribute much more.
  - One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.
  - The risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), we suggest rating down for at least one of the two.
  - Reviewers will face close-call situations. They should both acknowledge that they are in such a situation, make it explicit why they think this is the case, and make the reasons for their ultimate judgment apparent. **(GRADE Handbook)**

Final points

- You still need to assess risk of bias if only one study
- You still need to assess risk of bias if you cannot pool the results
- You still need to assess risk of bias is there is little information regarding the risk of bias

Session 6: Inconsistency

MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare
Cochrane Czech Republic
JBI
The Czech Republic (Middle European) Centre for Evidence-Based Healthcare
Masaryk University
GRADE

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 value (I2)

Standard chi-squared test (Cochran Q test)

- This tests the statistical hypothesis that the true treatments effects (the effect size parameters) are the same in all the primary studies included in meta-analysis
- If results of the test are statistically significant (P-value <0.1) the statistical hypothesis that the true treatments effects are the same in all the primary studies included in meta-analysis (the hypothesis of homogeneity) is rejected, therefore, it is considered that there is statistical heterogeneity

Standard chi-squared test (Cochran Q test)

- The statistical power of the test is in most cases very low due to the small number of studies; heterogeneity may be present even if the Q statistic is not statistically significant at conventional levels of significance such as 0.05. As such, a cut-off significance level of 0.10 rather than the usual 0.05 has been advocated
- With a small number of studies (< 20), the Q test should be interpreted very cautiously
- It is not appropriate to decide the meta-analysis model based only on the results of the Chi-squared statistical test (Q test) for heterogeneity

I square value (I^2)

- A statistic used for quantifying inconsistency in meta-analysis
- I^2 is a percentage and its value lies between 0% and 100%
- A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity . If I^2 = 0% this means that all variability in effect size estimates is due to sampling error within studies
- If I^2= 50% it means that half of the total variability among effect sizes is caused not by sampling error but by true heterogeneity between studies
- With a small number of studies (< 20), the I^2 test should be interpreted very cautiously

Interpreting I^2

- Generally in regards to heterogeneity:
- < 40% may be low
- 30-60% may be moderate
- 50-90% may be substantial
- 75-100% may be considerable

(GRADE Handbook)

Rule of thumb: less than 30% probably fine, above 30% needs to be investigated

Activity 4

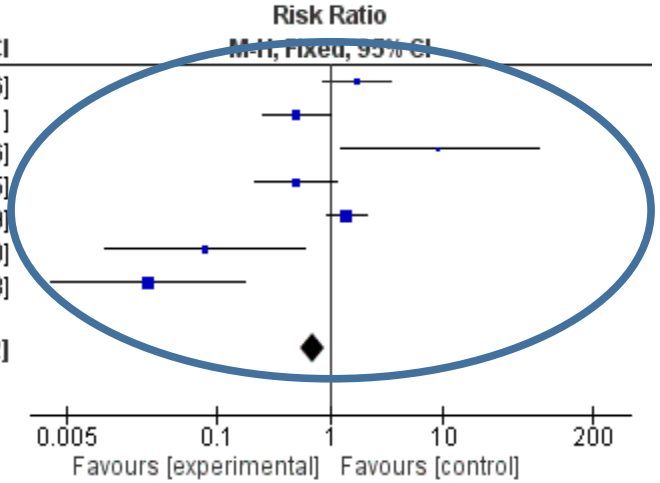
- Turn to your workbook and complete activity 4 with your partner. View the forest plots and decide whether or not you would rate down for inconsistency.

Example Forest Plot

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Asida 2012	20	450	13	510	8.6%	1.74 [0.88, 3.46]
Chan 2009	12	500	24	510	16.7%	0.51 [0.26, 1.01]
Desai 2012	8	22	1	25	0.7%	9.09 [1.23, 67.06]
Gandhi 2007	8	185	16	186	11.2%	0.50 [0.22, 1.15]
Hoedemaekers 2005	50	245	36	250	25.1%	1.42 [0.96, 2.09]
Lazar 2011	1	100	15	120	9.6%	0.08 [0.01, 0.60]
Rujirojindakul 2014	1	99	40	100	28.0%	0.03 [0.00, 0.18]
Total (95% CI)		**1601**		**1701**	**100.0%**	**0.72 [0.57, 0.92]**
Total events	100		145			

Heterogeneity: Chi² = 41.50, df = 6 (P < 0.00001); I² = 86%

Test for overall effect: Z = 2.62 (P = 0.009)

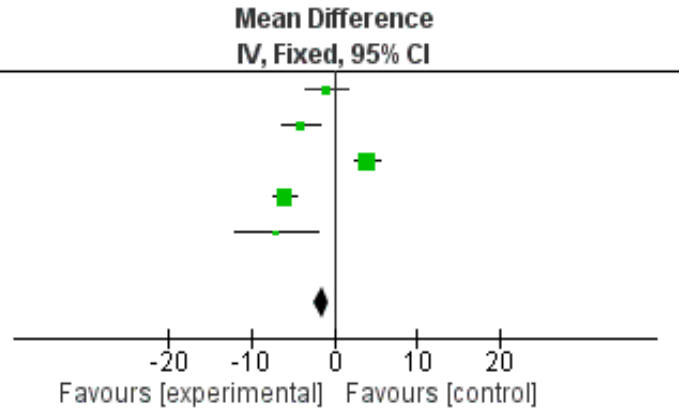


Forest Plot example: Continuous Data

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Azarfarin 2011	64	8	89	65	10	94	11.6%	-1.00 [-3.62, 1.62]
Gandhi 2007	72	12	150	76	9	150	13.8%	-4.00 [-6.40, -1.60]
Hoedemaekers 2005	58	5	110	54	6	108	36.8%	4.00 [2.53, 5.47]
Ingels 2006	84	4	79	90	5	64	34.8%	-6.00 [-7.51, -4.49]
Kirdemir 2008	68	16	58	75	12	63	3.1%	-7.00 [-12.07, -1.93]
Total (95% CI)			**486**			**479**	**100.0%**	**-1.50 [-2.39, -0.61]**

Heterogeneity: $Chi^2 = 96.92$, $df = 4$ ($P < 0.00001$); $I^2 = 96\%$

Test for overall effect: $Z = 3.29$ ($P = 0.0010$)



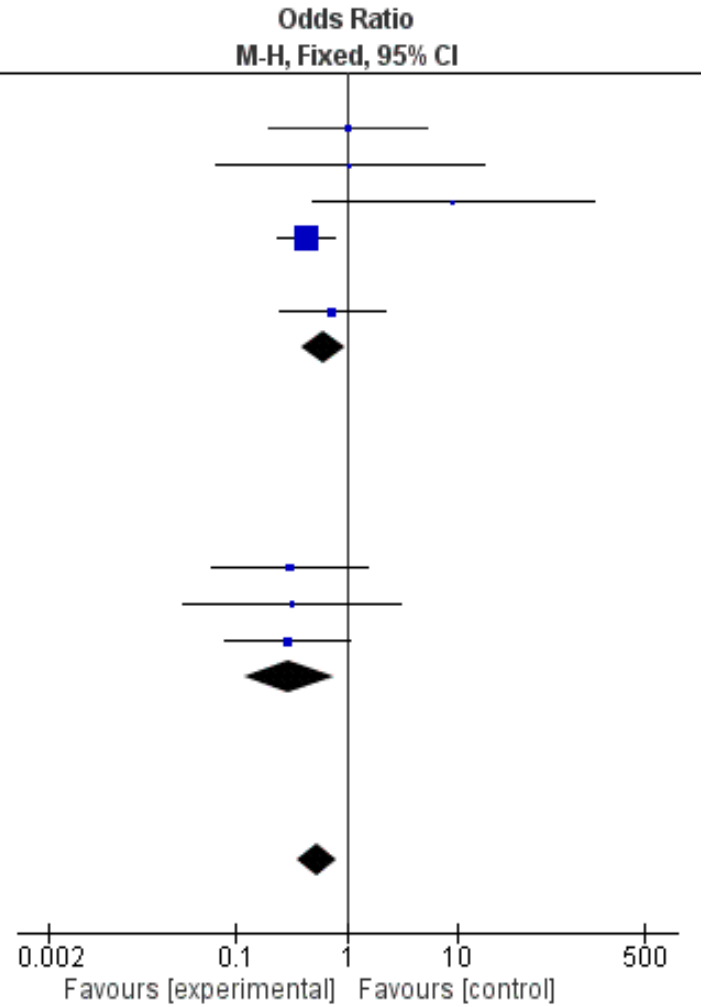
Note:

- As we define quality of evidence **for a guideline**, inconsistency is important only when it reduces confidence in results **in relation to a particular decision**. Even when inconsistency is large, it may not reduce confidence in results regarding a particular decision.
- Guideline developers may or may not consider this degree of variability important. Systematic review authors, much less in a position to judge whether the apparent high heterogeneity can be dismissed on the grounds that it is unimportant, are more likely to rate down for inconsistency.

Caution: subgroups

- Although the issue is controversial, we recommend that meta-analyses include formal tests of whether a priori hypotheses explain inconsistency between important subgroups
- If inconsistency can be explained by **differences in populations**, **interventions** or **outcomes**, review authors should offer different estimates across patient groups, interventions, or outcomes. Guideline panelists are then likely to offer different recommendations for different patient groups and interventions. If **study methods** provide a compelling explanation for differences in results between studies, then authors should consider focusing on effect estimates from studies with a lower risk of bias.

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
1.1.1 Nondiabetic patients						
Chan 2009	3	54	3	55	4.3%	1.02 [0.20, 5.29]
Desai 2012	1	91	1	98	1.5%	1.08 [0.07, 17.49]
Gandhi 2007	4	185	0	186	0.7%	9.25 [0.49, 173.00]
Ingels 2006	16	477	37	493	53.6%	0.43 [0.23, 0.78]
Lazar 2011	0	40	0	42		Not estimable
Rujirojindakul 2014	6	99	8	100	11.4%	0.74 [0.25, 2.22]
Subtotal (95% CI)		**946**		**974**	**71.5%**	**0.62 [0.39, 0.98]**
Total events	30		49			
Heterogeneity: Chi² = 5.34, df = 4 (P = 0.25); I² = 25%						
Test for overall effect: Z = 2.04 (P = 0.04)						
1.1.2 Diabetic patients						
Hoedemaekers 2005	2	60	6	59	8.9%	0.30 [0.06, 1.58]
Ingels 2006	1	110	3	108	4.6%	0.32 [0.03, 3.14]
Kirdemir 2008	3	200	10	201	15.0%	0.29 [0.08, 1.07]
Subtotal (95% CI)		**370**		**368**	**28.5%**	**0.30 [0.12, 0.76]**
Total events	6		19			
Heterogeneity: Chi² = 0.01, df = 2 (P = 1.00); I² = 0%						
Test for overall effect: Z = 2.53 (P = 0.01)						
Total (95% CI)		**1316**		**1342**	**100.0%**	**0.53 [0.35, 0.80]**
Total events	36		68			
Heterogeneity: Chi² = 6.79, df = 7 (P = 0.45); I² = 0%						
Test for overall effect: Z = 3.05 (P = 0.002)						
Test for subgroup differences: Chi² = 1.85, df = 1 (P = 0.17), I² = 46.0%						



Session 7: Imprecision

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

Imprecision

- Small sample size
- Small number of events
- Wide confidence intervals
  - uncertainty about magnitude of effect
- Optimal information size
- Different for SRs vs Guidelines
  - Guidelines contextualized for decision making and recommendations
  - SRs free of this context

Optimal Information Size

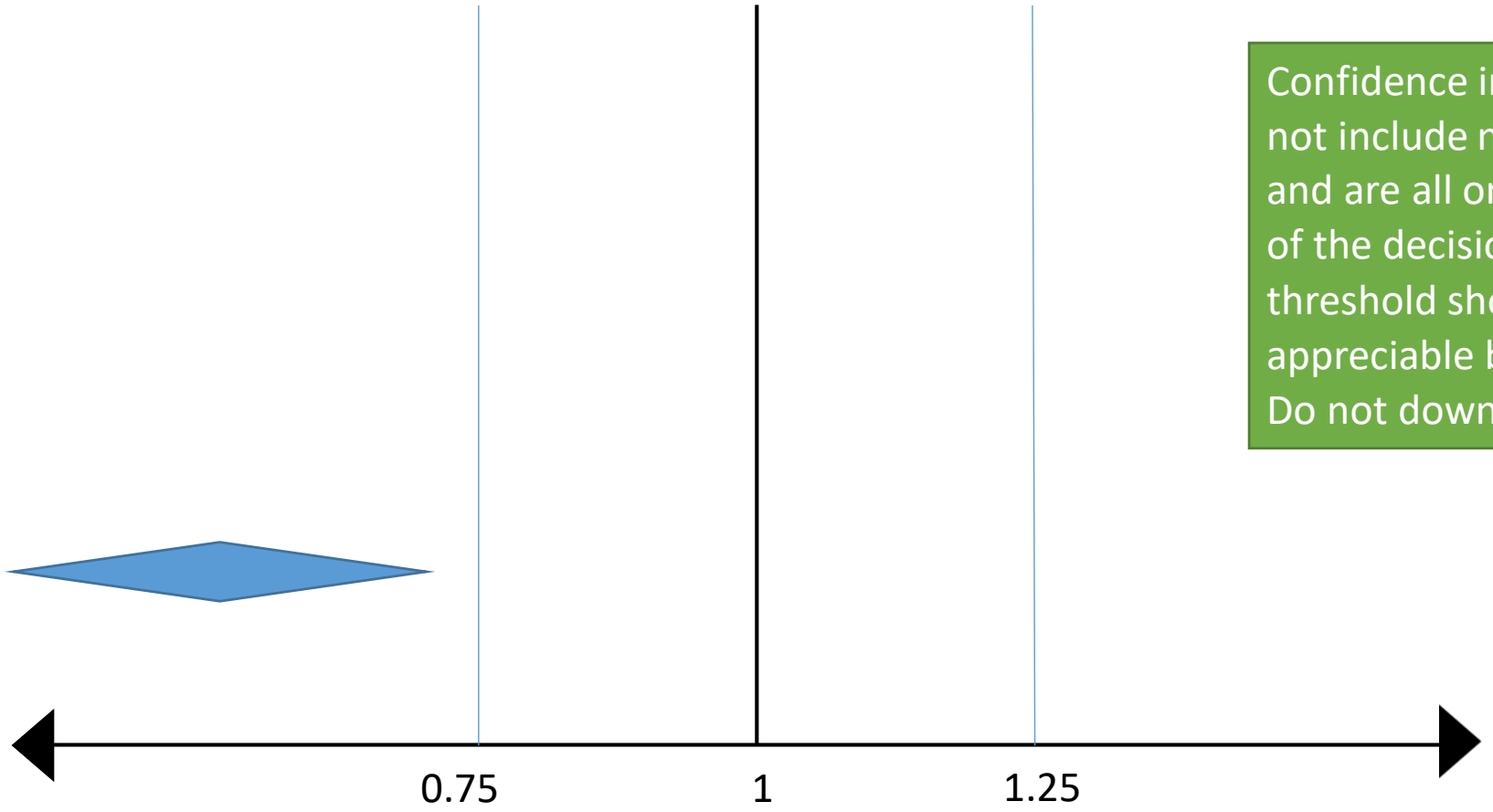
- If the **total number of patients** included in a systematic review is **less than** the number of patients generated by a **conventional sample size calculation** for a single adequately powered trial, consider **rating down** for imprecision.

Guyatt 2011

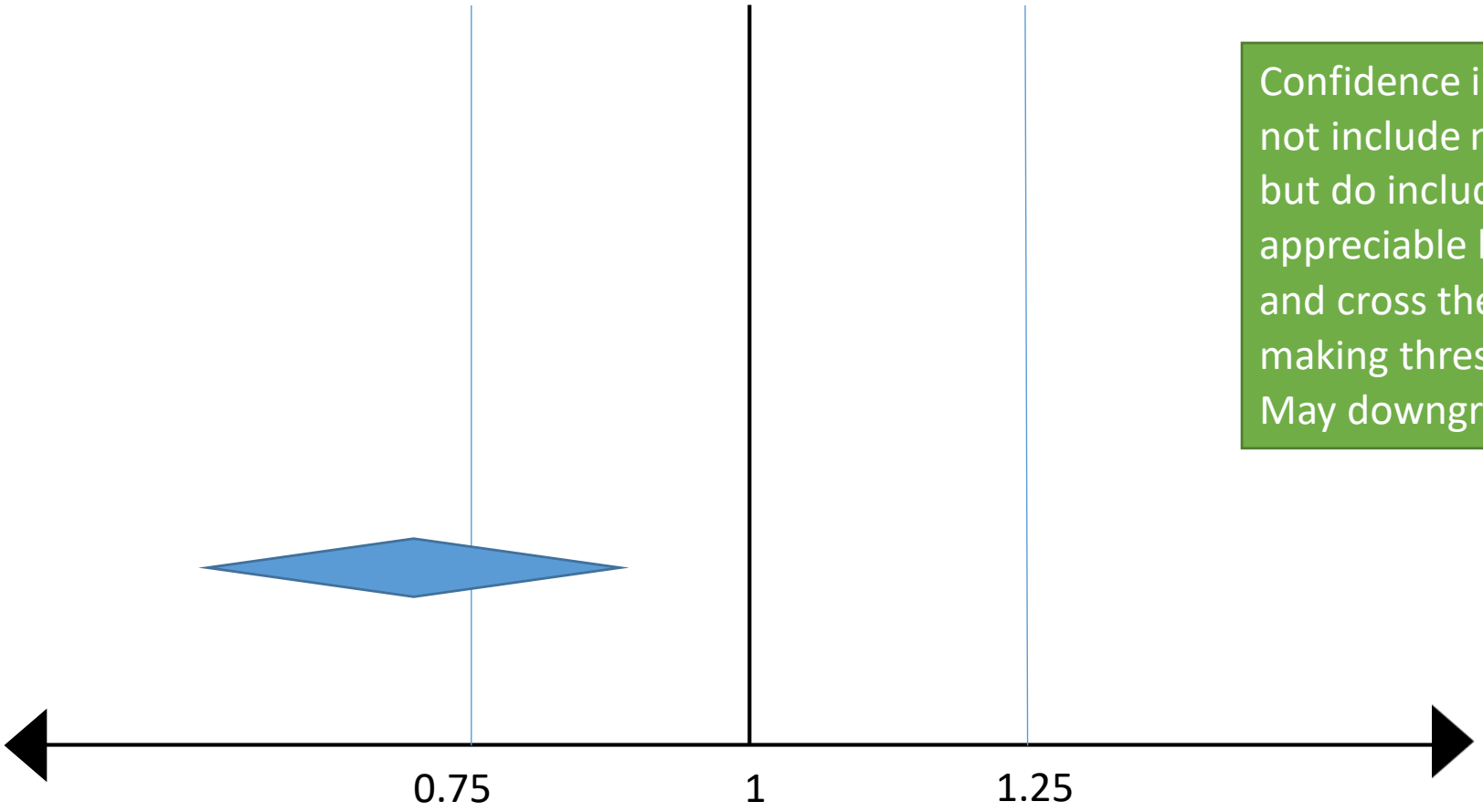
Total Number of Events	Relative Risk Reduction	Implications for meeting OIS threshold
100 or less	≤ 30%	Will almost never meet threshold whatever control event rate
200	30%	Will meet threshold for control event rates for ~ **25%** or greater
200	25%	Will meet threshold for control event rates for ~ **50%** or greater
200	20%	Will meet threshold only for control event rates for ~ **80%** or greater
300	≥ 30%	Will meet threshold
300	25%	Will meet threshold for control event rates ~ **25%** or greater
300	20%	Will meet threshold for control event rates ~ **60%** or greater
400 or more	≥ 25%	Will meet threshold for any control event rate
400 or more	20%	Will meet threshold for control event rates of ~ **40%** or greater

OIS rule of thumb:

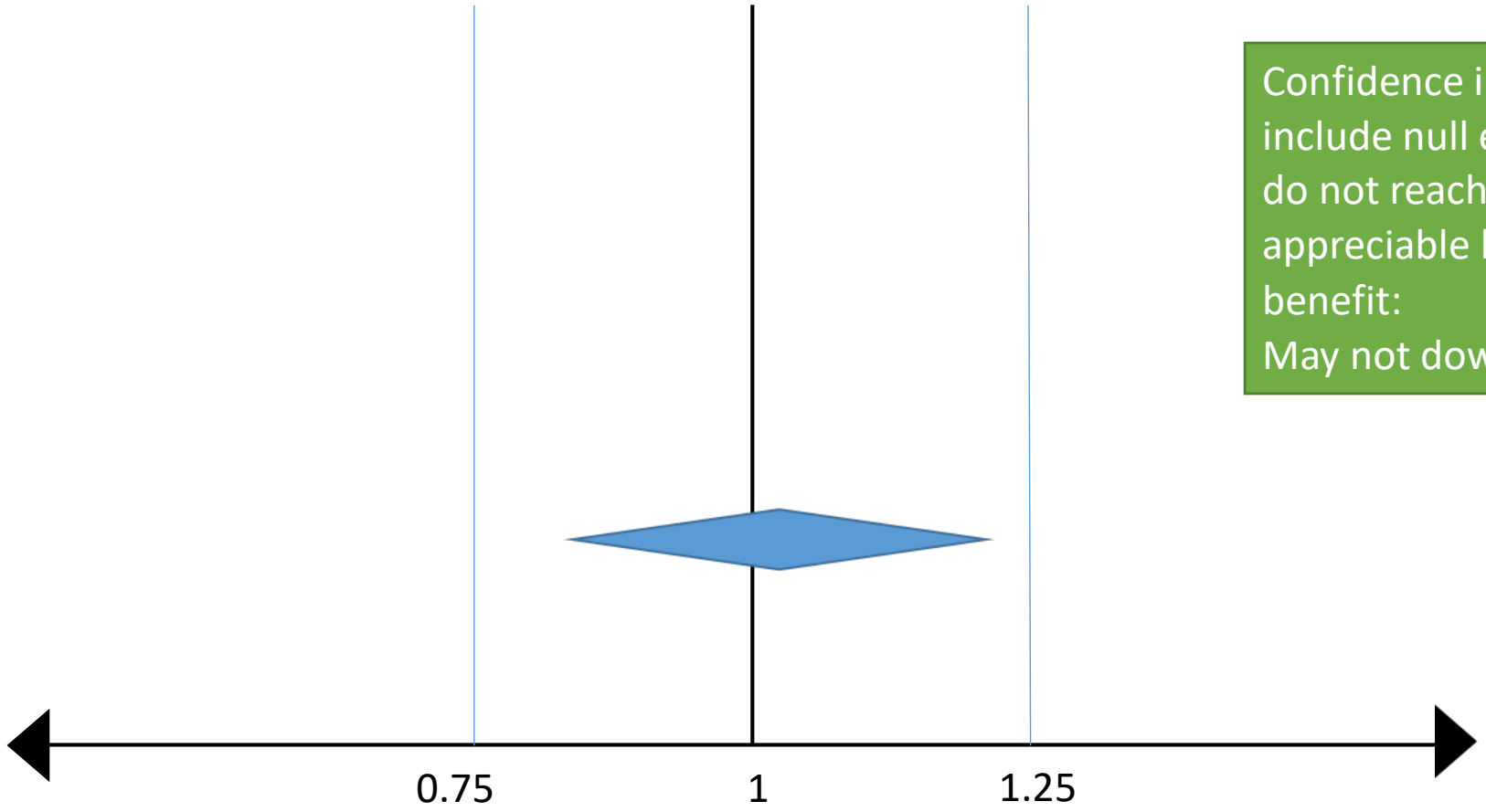
- dichotomous: 300 events
- continuous: 400 participants
- HOWEVER, carefully consider the OIS and event rate



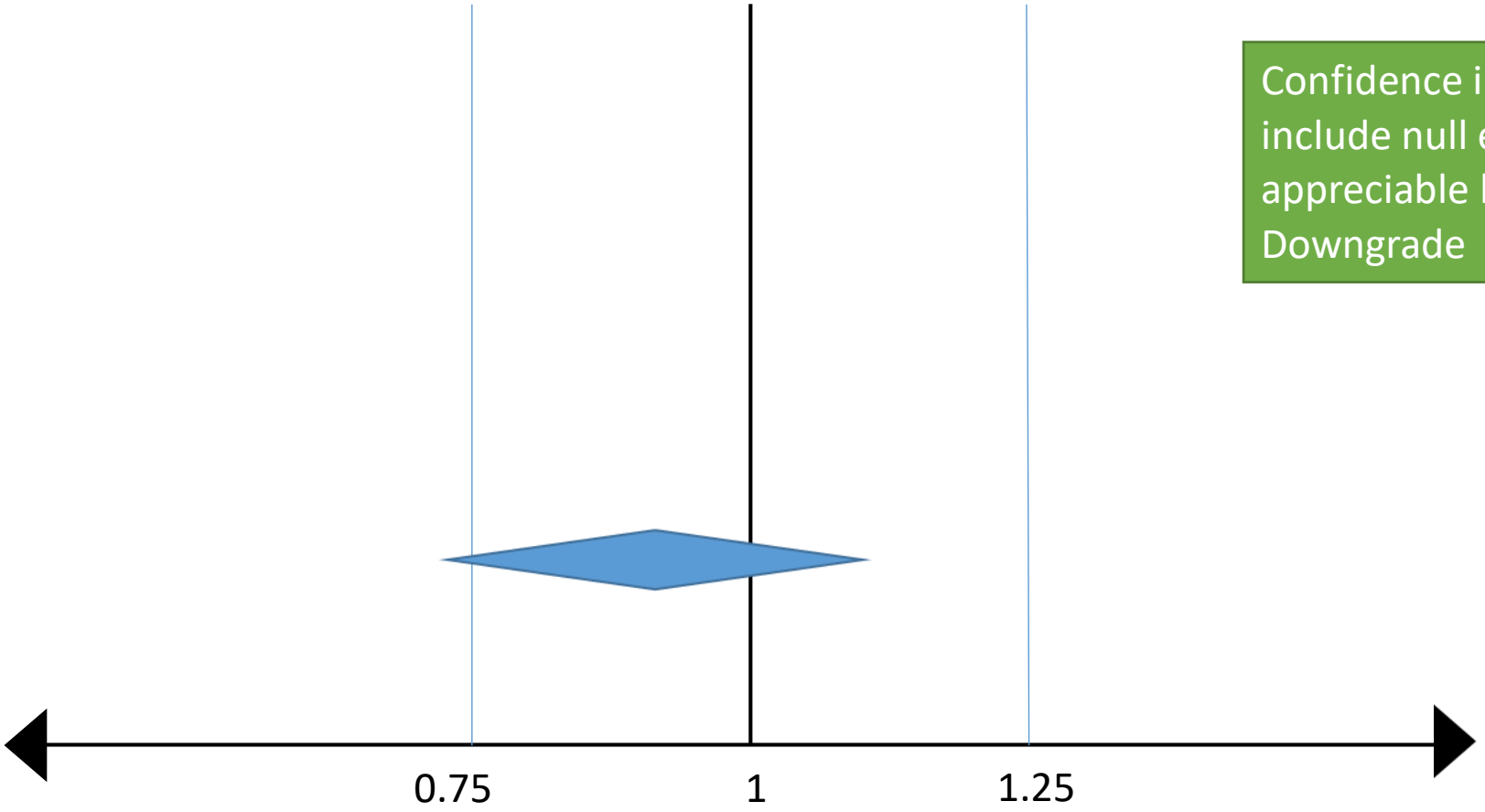
Confidence intervals do not include null effect, and are all on one side of the decision threshold showing appreciable benefit: Do not downgrade



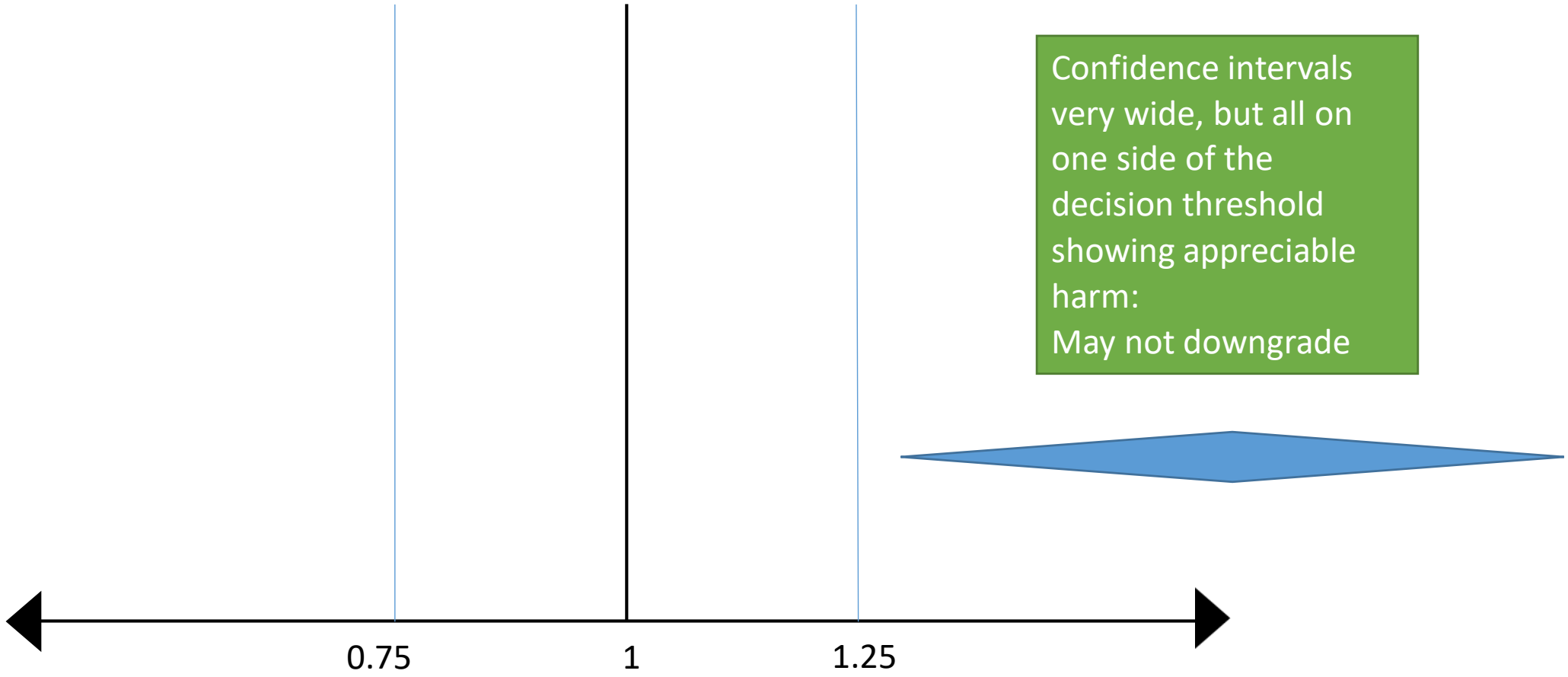
Confidence intervals do not include null effect, but do include appreciable benefit and cross the decision making threshold:
May downgrade



Confidence intervals do include null effect, but do not reach appreciable harm or benefit:
May not downgrade



Confidence intervals do include null effect, and appreciable benefit: Downgrade



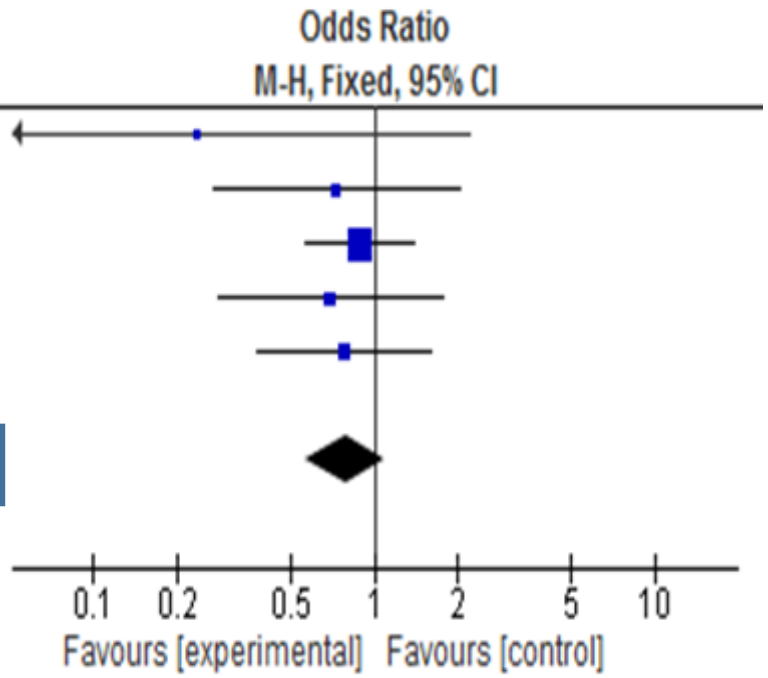
Confidence intervals very wide, but all on one side of the decision threshold showing appreciable harm:
May not downgrade

Activity 5: would you rate down?

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Asida 2012	1	50	4	50	4.7%	0.23 [0.03, 2.18]
Desai 2012	7	91	10	98	10.7%	0.73 [0.27, 2.02]
Gandhi 2007	54	185	59	186	50.4%	0.89 [0.57, 1.38]
Lazar 2011	12	40	16	42	13.2%	0.70 [0.28, 1.75]
Rujirojindakul 2014	17	99	21	100	20.9%	0.78 [0.38, 1.59]
Total (95% CI)		465		476	100.0%	0.79 [0.57, 1.09]
Total events	91		110			

Heterogeneity: Chi² = 1.50, df = 4 (P = 0.83); I² = 0%

Test for overall effect: Z = 1.41 (P = 0.16)

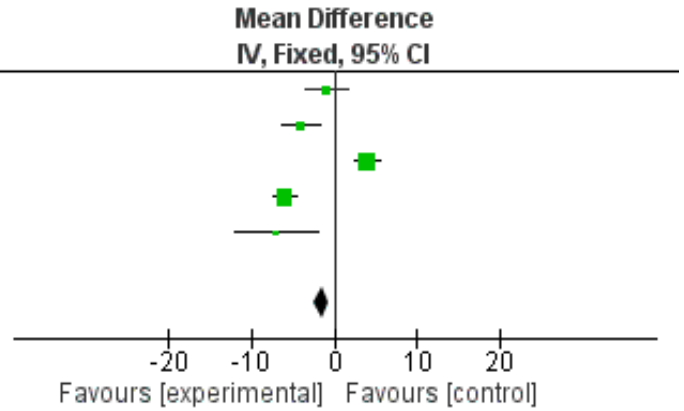


Forest Plot example: Continuous Data

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Azarfarin 2011	64	8	89	65	10	94	11.6%	-1.00 [-3.62, 1.62]
Gandhi 2007	72	12	150	76	9	150	13.8%	-4.00 [-6.40, -1.60]
Hoedemaekers 2005	58	5	110	54	6	108	36.8%	4.00 [2.53, 5.47]
Ingels 2006	84	4	79	90	5	64	34.8%	-6.00 [-7.51, -4.49]
Kirdemir 2008	68	16	58	75	12	63	3.1%	-7.00 [-12.07, -1.93]
Total (95% CI)			**486**			**479**	**100.0%**	**-1.50 [-2.39, -0.61]**

Heterogeneity: $Chi^2 = 96.92$, $df = 4$ ($P < 0.00001$); $I^2 = 96\%$

Test for overall effect: $Z = 3.29$ ($P = 0.0010$)



Session 8: Indirectness

MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare
Cochrane Czech Republic
JBI
The Czech Republic (Middle European) Centre for Evidence-Based Healthcare
Masaryk University
GRADE

Directness of Evidence (generalizability, transferability, external validity, applicability)

- Confidence is increased when we have direct evidence
- Ask: is the evidence applicable to our relevant question?
  - Population
  - Intervention
  - Comparisons
  - Outcome

Population

- Ask: Is the population included in these studies similar to those in my question?
  - Indirect evidence examples:
    - Evidence from high income countries compared to LMIC
    - All women as compared to pregnant women
    - Sick (or sicker) people compared to all people (mild vs severe)
    - Adults compared to children
  - May be addressed in subgroups where appropriate and possible
  - Can indicate different levels of risk for different groups
  - Can create different SoF tables for different groups, therefore won't need to downgrade

Interventions

- Ask: Is the population included in these studies similar to those in my question?
  - Older technology compared to newer technology
  - Co-interventions
  - Different doses, different delivery, different providers

Comparisons

- Are comparisons direct or indirect?
  - Interested in A vs B
    - A vs Control
    - B vs Control
  - May downgrade

Outcomes

- Make sure to:
  - Choose patient important outcomes
  - Avoid surrogate outcomes
- If surrogate outcomes used, is there a strong association between the surrogate and patient important outcome?

Note:

- Authors of systematic reviews should answer the health care question they asked and, thus, they will rate the directness of evidence they found. The considerations made by the authors of systematic reviews may be different than those of guideline panels that use the systematic reviews. The more clearly and explicitly the health care question was formulated the easier it will be for the users to understand systematic review authors' judgments.

Session 9: Publication bias

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

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

Publication Bias

- "**Publication bias** is the term for what occurs whenever the research that appears in the published literature is ***systematically unrepresentative*** of the ***population of completed studies***. Simply put, when the ***research that is readily available differs in its results from the results of all the research that has been done in an area***, readers and reviewers of that research are in danger of drawing the wrong conclusion about what that body of research shows." (Rothstein, Sutton and Borenstein 2005, p.1)

Rothstein HR, Sutton AJ, Borenstein M (Editors). Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments. Chichester: Wiley, 2005.

Publication Bias

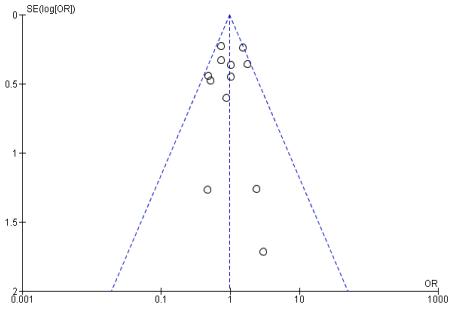
Potential information suppression mechanisms (causes of publication bias):

- **Language bias** ("selective inclusion of studies published in English");
- **Availability bias** ("selective inclusion of studies that are easily accessible to the researcher");
- **Cost bias** ("selective inclusion of studies that are available free or at low cost");
- **Familiarity bias** ("selective inclusion of studies only from one's own discipline";
- **Outcome bias** ("***selective reporting*** by the author of a primary study of some outcomes but not others, ***depending on the direction and statistical significance of the results***") (Rothstein, Sutton and Borenstein 2005, p.3)

Rothstein HR, Sutton AJ, Borenstein M (Editors). Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments. Chichester: Wiley, 2005.

Funnel Plot

- Funnel plots are a method of investigating the retrieved studies in a meta-analysis for publication bias
- A funnel plot is a scatter plot in which an effect estimate of each study is plotted against a measure of size or precision
- If no bias, expect symmetric and inverted funnel
- If bias, expect asymmetric or skewed shape
- Can also investigate small study effects



Funnel Plot

- A statistical test for funnel plot asymmetry investigates whether the association between effect estimate and measure of study size or precision is larger than what can be expected to have occurred by chance
- Egger test, Begg test, and Harbord test are the most popular statistical tests
- Due to low power a finding of no evidence of asymmetry does not serve to exclude bias
- Generally 10 studies are considered the minimum number to justify a funnel plot
- When there are less than 30 studies, the statistical power of all three tests is very low and results should be interpreted with caution

Figure 1

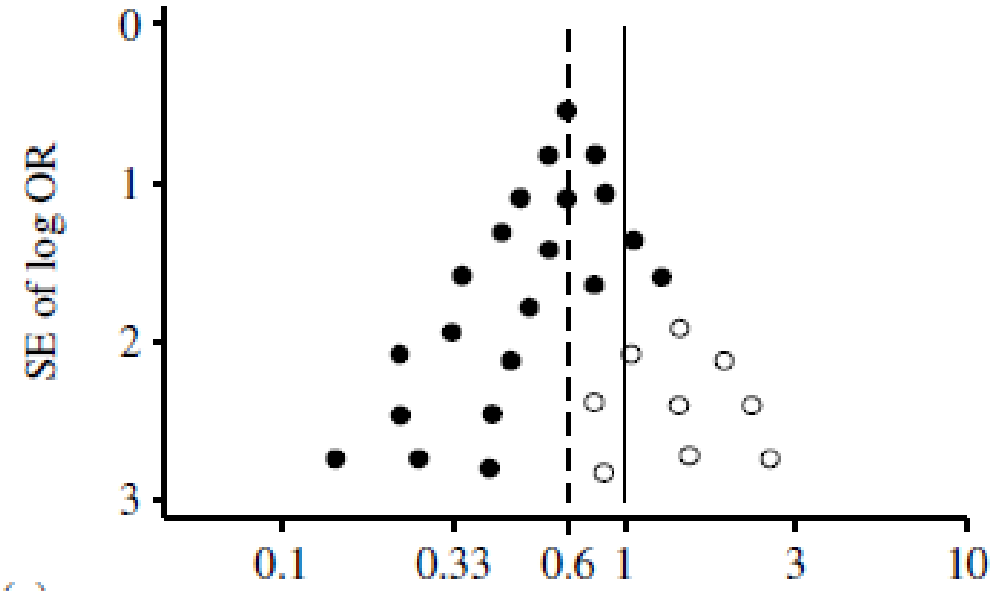
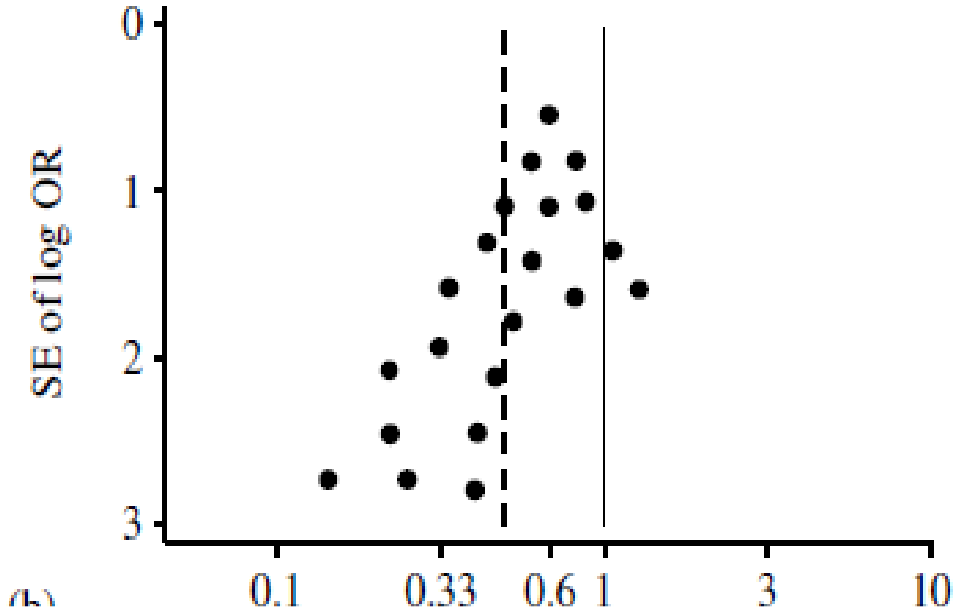


Figure 2



Taken from: Sterne et al 2005

What do we do?

"It is extremely difficult to be confident that publication bias is absent and almost as difficult to place a threshold on when to rate down quality of evidence due to the strong suspicion of publication bias. For this reason GRADE suggests rating down quality of evidence for publication bias by a maximum of one level." (GRADE Handbook)

Consider:

- study size (small studies vs. large studies)
- lag bias (early publication of positive results)
- search strategy (was it comprehensive?)
- asymmetry in funnel plot.

Session 10: 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.

(GRADE Handbook)

What can raise quality?

1. Large magnitude of an effect
2. Dose-response gradient
3. Effect of opposing plausible residual confounding

Large magnitude of an effect

- Large, consistent, precise effect
- Although observational studies may overestimate the effect, bias is unlikely to explain or contribute all the effect 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
  - New concept of E-value for small effects but substantialy population significant (eg. Air polution RR 1,06 for exposure higher than per 10 µg per cubic meter (increase in mortality per world population 45 mil)
  - May upgrade 1 level for large and 2 for very large

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)

Effect of opposing 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

Effect of opposing plausible residual confounding Example 1

- Example 1: When confounding is expected to reduce a demonstrated effect (Upgraded by One Level)
- A rigorous systematic review of observational studies including a total of 38 million patients demonstrated higher death rates in private for-profit versus private not-for-profit hospitals. It is likely, however, that patients in the not-for-profit hospitals were sicker than those in the for-profit hospitals. This would bias results against the not-for-profit hospitals. The second likely bias was the possibility that higher numbers of patients with excellent private insurance coverage could lead to a hospital having more resources and a spill-over effect that would benefit those without such coverage. Since for-profit hospitals are likely to admit a larger proportion of such well-insured patients than not-for-profit hospitals, the bias is once again against the not-for-profit hospitals. Because the plausible biases would all diminish the demonstrated intervention effect, one might consider the evidence from these observational studies as moderate rather than low quality.

Effect of opposing plausible residual confounding Example 2

- Example 2: When confounding is expected to increase the effect but no effect was observed (Upgraded by One Level)
- Consider the early reports associating MMR vaccination with autism. One would think that there would be over-reporting of autism in children given MMR vaccines. However, systematic reviews failed to prove any association between the two. Due to the negative results, despite the potential presence of confounders which would increase the likelihood of reporting of autism, no association was found. Therefore, we may upgrade the level of evidence by one level.

Session 11: Summary of findings tables and evidence profiles

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

Introduction to GRADEpro


GUIDELINE DEVELOPMENT
TOOL TUTORIAL
EVIDENCE TABLE GENERATION
Mac
GRADE
Centre | McMaster University

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 1
GRADE evidence profile: antibiotics for children with acute otitis media

Quality assessment						Summary of findings					
						Number of patients			Absolute risk		
No of studies (Design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Placebo	Antibiotics	Relative risk (95% CI)	Control risk[a]	Risk difference (95% CI)	Quality
Pain at 24h											
5 (RCT)	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	241/605	223/624	RR 0.9 (0.78–1.04)	367/1,000	Not Significant	⊕⊕⊕⊕ High
Pain at 2–7 d											
10 (RCT)	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	303/1,366	228/1,425	RR 0.72 (0.62–0.83)	257/1,000	72 fewer per 1,000 (44–98)	⊕⊕⊕⊕ High
Hearing, inferred from the surrogate outcome abnormal tympanometry—1 mo											
4 (RCT)	No serious limitations	No serious inconsistency	Serious indirectness (because of indirectness of outcome)	No serious imprecision	Undetected	168/460	153/467	RR 0.89 (0.75–1.07)	350/1,000	Not Significant	⊕⊕⊕○ Moderate
Hearing, inferred from the surrogate outcome abnormal tympanometry—3 mo											
3 (RCT)	No serious limitations	No serious inconsistency	Serious indirectness (because of indirectness of outcome)	No serious imprecision	Undetected	96/398	96/410	RR 0.97 (0.76–1.24)	234/1,000	Not Significant	⊕⊕⊕○ Moderate
Vomiting, diarrhea, or rash											
5 (RCT)	No serious limitations	Serious inconsistency (because of inconsistency in absolute effects)	No serious indirectness	No serious imprecision	Undetected	83/711	110/690	RR 1.38 (1.09–1.76)	113/1,000	43 more per 1,000 (10–86)	⊕⊕⊕○ Moderate

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

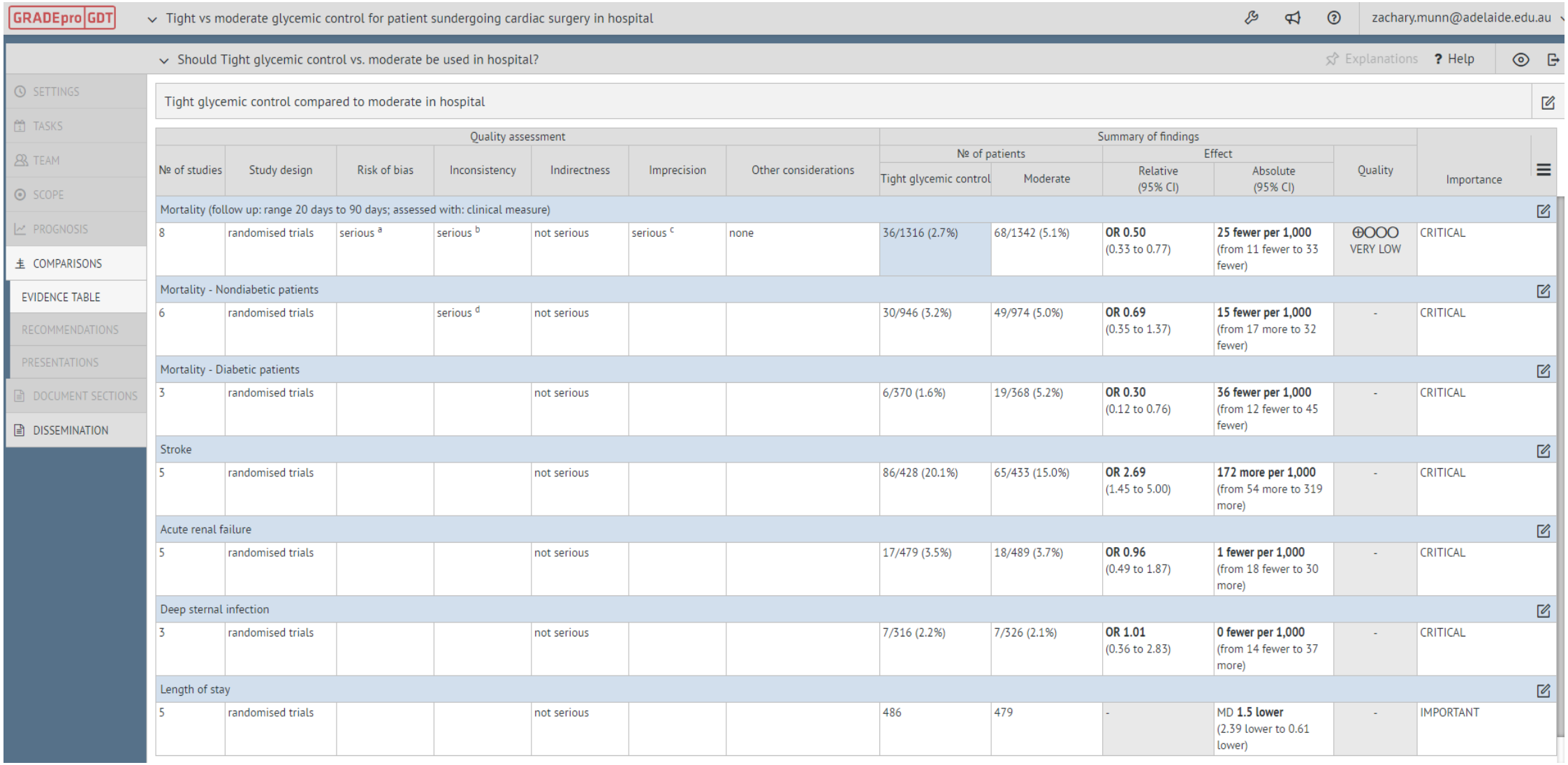
Probiotics as an adjunct to antibiotics for the prevention of pediatric antibiotic-associated diarrhea in children

Patient or population: children given antibiotics
Settings: inpatients and outpatients
Intervention: probiotics

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Probiotics				
Incidence of Diarrhea: Probiotic dose (equal to/ greater than) 5 billion CFU/day Follow-up: 10 days to 3 mo's	**223 per 1000**	**89 per 1000** (65 to 122)	**RR 0.4** (0.29 to 0.55)	1474 (7 studies)	⊕⊕◯◯ **low**[1,2]	Control group risk estimates come from control arm of meta-analysis, based on included trials. Relative effect based on available case analysis
Adverse Events Follow-up: 10 to 44 days	**18 per 1000**	**23 per 1000** (8 to 38)	See comment	1575 (11 studies)	⊕⊕◯◯ **low**[3,4]	Risks were calculated from pooled risk differences. Control group risk estimates come from control arm of the meta-analysis, based on included studies
Duration of Diarrhea Follow-up: 10 days to 3 mo's		The mean duration of diarrhea in the intervention groups was **0.6 lower** (1.18 to 0.02 lower)		897 (5 studies)	⊕⊕◯◯ **low**[5,6]	

Activity 6: Summary of Findings table

- Using the materials provided, logon to GRADEPro GDT and create a SoF table.
- Dropbox link:



Quality assessment							Summary of findings					
							№ of patients		Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tight glycemic control	Moderate	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (follow up: range 20 days to 90 days; assessed with: clinical measure)												
8	randomised trials	serious [a]	serious [b]	not serious	serious [c]	none	36/1316 (2.7%)	68/1342 (5.1%)	**OR 0.50** (0.33 to 0.77)	**25 fewer per 1,000** (from 11 fewer to 33 fewer)	⨁◯◯◯ VERY LOW	CRITICAL
Mortality - Nondiabetic patients												
6	randomised trials		serious [d]	not serious			30/946 (3.2%)	49/974 (5.0%)	**OR 0.69** (0.35 to 1.37)	**15 fewer per 1,000** (from 17 more to 32 fewer)	-	CRITICAL
Mortality - Diabetic patients												
3	randomised trials			not serious			6/370 (1.6%)	19/368 (5.2%)	**OR 0.30** (0.12 to 0.76)	**36 fewer per 1,000** (from 12 fewer to 45 fewer)	-	CRITICAL
Stroke												
5	randomised trials			not serious			86/428 (20.1%)	65/433 (15.0%)	**OR 2.69** (1.45 to 5.00)	**172 more per 1,000** (from 54 more to 319 more)	-	CRITICAL
Acute renal failure												
5	randomised trials			not serious			17/479 (3.5%)	18/489 (3.7%)	**OR 0.96** (0.49 to 1.87)	**1 fewer per 1,000** (from 18 fewer to 30 more)	-	CRITICAL
Deep sternal infection												
3	randomised trials			not serious			7/316 (2.2%)	7/326 (2.1%)	**OR 1.01** (0.36 to 2.83)	**0 fewer per 1,000** (from 14 fewer to 37 more)	-	CRITICAL
Length of stay												
5	randomised trials			not serious			486	479	-	MD **1.5 lower** (2.39 lower to 0.61 lower)	-	IMPORTANT

Activity 6:

- You can also:
  - Export table
  - View different presentation formats
  - View interactive summary of findings table

Session 12: Common questions regarding GRADE

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

Masaryk University
GRADE

What to do when you can't pool?

- Can report results from a single study
- Can report a range from multiple studies if can't pool
- Still need to consider all domains (inconsistency and imprecision included)

Length of intensive care unit (ICU) or hospital stay (follow-up: varied as a consequence of ICU/hospital stay)	See comment	ICU: 1060 (9) Hospital: 1259 (9)	⊕⊕⊕⊖ **moderate**	Length of intensive care unit stay: Mean differences between intensive and regular glucose control groups ranged from -1.7 days to 2.1 days Length of hospital stay: Mean differences between intensive and regular glucose control groups ranged from -8 days to 3.7 days

Murad, M. H., Mustafa, R. A., Schünemann, H. J., Sultan, S., & Santesso, N. (2017). Rating the certainty in evidence in the absence of a single estimate of effect. Evidence-Based Medicine, ebmed-2017.

GRADE in Public Health and Complex Interventions– concerns

- Assessment of evidence regarding public health can be challenging:
  1. Penalizing when RCTs are not available or even possible
  2. Observational studies not being equal, all start as 'low'
  3. Heterogeneity in these reviews
  4. Dependence on context – different findings expected in different settings
  5. No pooled effect size
  6. Likely to have low ratings

MH, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J, Klugar M, Choi L, Saz-Parkinson Z, Mustafa RA, Langendam MW, Crane O, Morgan RL, Rehfuess E, Johnston BC, Chong LY, Guyatt GH, Schünemann HJ, Katikireddi SV; GRADE Working Group. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: A concept paper from the GRADE Public Health Group. J Clin Epidemiol. 2021 Jan 18:S0895-4356(21)00003-2. doi: 10.1016/j.jclinepi.2021.01.001.

GRADE in public health - responses

1. Lower rankings should not be seen as a penalty – we know randomisation is of utmost importance – can reframe this as all studies equal and randomisation warrant 'upgrade'
2. Although they all start as low, assessment of methodological limitations will tease out differences in quality between observational studies. Also findings from observational studies can be upgraded
3. Inconsistency should be explored, investigated, and reasons identified
4. GRADE considers context explicitly in indirectness, addressing important contextual aspects for consideration in the summary of evidence
5. GRADE can be used when no pooled effect size
6. Recommendations rely not only on ranking. Still important to do this process to acknowledge any issues. If issues with interpretation, rankings can be reframed or different terminology used.

Schünemann, Holger, et al. "The GRADE approach and Bradford Hills criteria for causation." Journal of Epidemiology & Community Health 65.5 (2011): 392-395.

Considerations when ranking evidence

- While factors influencing the quality of evidence are additive – such that the reduction or increase in each individual factor is added together with the other factors to reduce or increase the quality of evidence for an outcome – grading the quality of evidence involves judgements which are not exclusive. Therefore, GRADE is not a quantitative system for grading the quality of evidence. Each factor for downgrading or upgrading reflects not discrete categories but a **continuum** within each category and among the categories. When the body of evidence is intermediate with respect to a particular factor, the decision about whether a study falls above or below the threshold for up- or downgrading the quality (by one or more factors) depends on judgment.

Considerations when ranking evidence

- For example, if there was some uncertainty about the three factors: study limitations, inconsistency, and imprecision, but not serious enough to downgrade each of them, one could reasonably make the case for downgrading, or for not doing so. A reviewer might in each category give the studies the benefit of the doubt and would interpret the evidence as high quality. Another reviewer, deciding to rate down the evidence by one level, would judge the evidence as moderate quality. Reviewers should grade the quality of the evidence by considering both the individual factors in the context of other judgments they made about the quality of evidence for the same outcome.

Considerations when ranking evidence

- In such a case, you should pick one or two categories of limitations which you would offer as reasons for downgrading and explain your choice in the footnote. You should also provide a footnote next to the other factor, you decided not to downgrade, explaining that there was some uncertainty, but you already downgraded for the other factor and further lowering the quality of evidence for this outcome would seem inappropriate. GRADE strongly encourages review and guideline authors to be explicit and transparent when they find themselves in these situations by acknowledging borderline decisions.
- Despite the limitations of breaking **continua** into categories, treating each criterion for rating quality up or down as discrete categories enhances transparency. Indeed, the great merit of GRADE is not that it ensures reproducible judgments but that it requires explicit judgment that is made transparent to users.

Is this the end...or next steps?

- The endpoint for systematic reviews and for HTA restricted to evidence reports is a summary of the evidence—the quality rating for each outcome and the estimate of effect. For guideline developers and HTA that provide advice to policymakers, a summary of the evidence represents a key milestone on the path to a recommendation.
- Guideline developers (but not systematic reviewers) then review all the information to make a final decision about which outcomes are critical and which are important and come to a final decision regarding the rating of overall quality of evidence, before considering making recommendations.

Session 13: Making Recommendations

MUNI MED
Institute of Biostatistics and Analyses
Czech National Centre
for Evidence-Based Healthcare

Cochrane
Czech Republic

JBI
The Czech Republic (Middle European)
Centre for Evidence-Based Healthcare

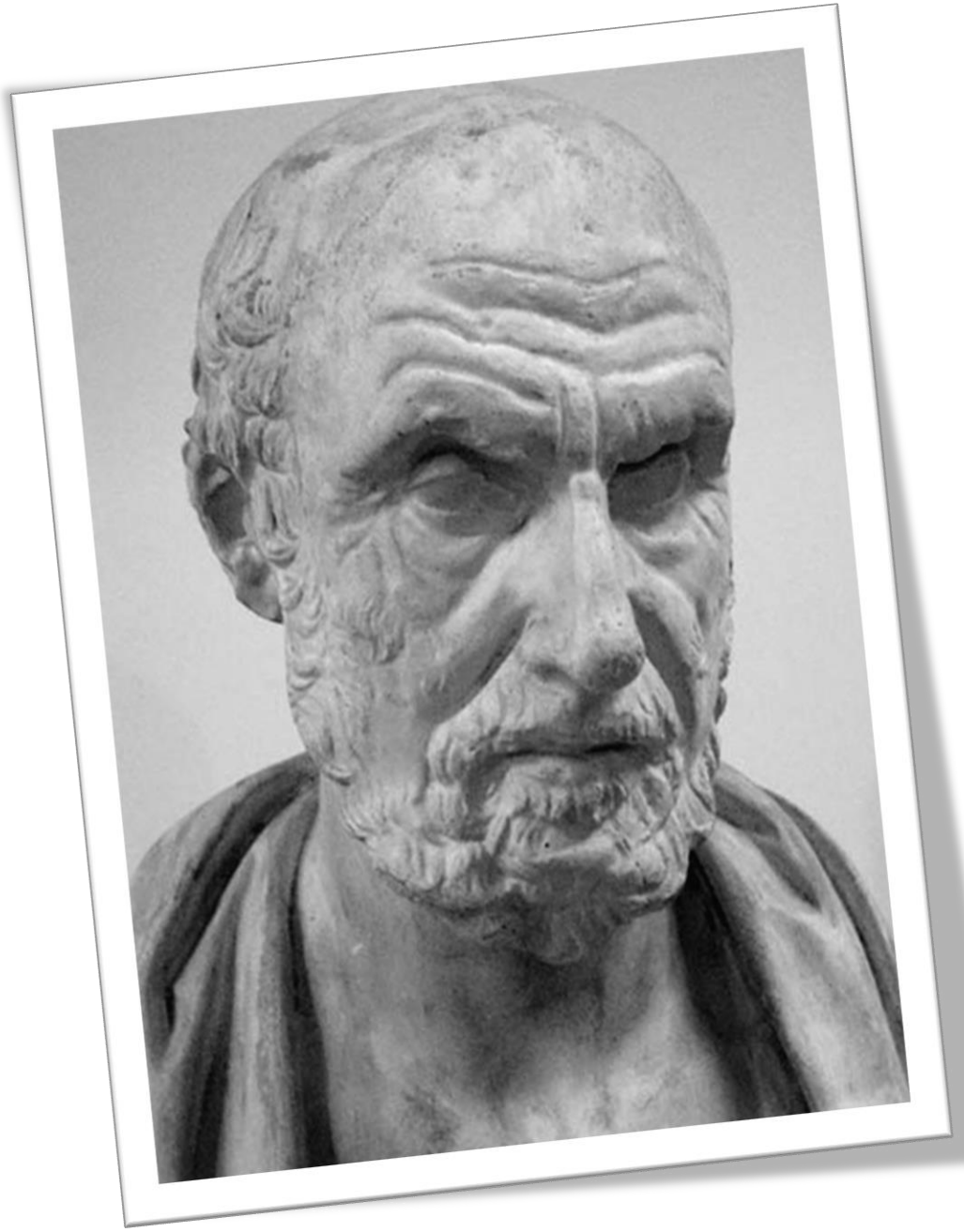
Masaryk University
GRADE

Overall certainty of evidence

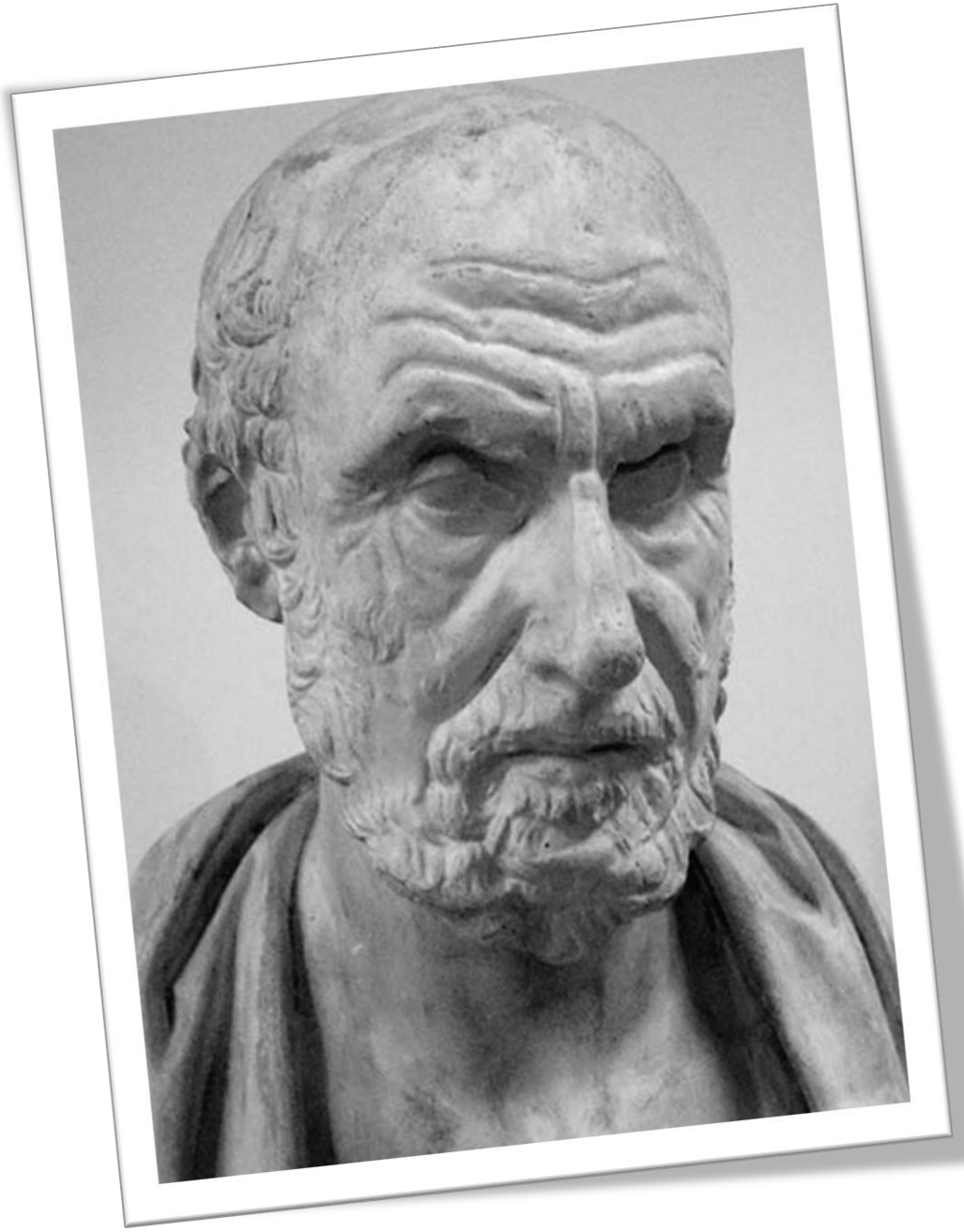
- Systematic review authors only rate the evidence for each outcome
- Guideline Panels need to determine the quality of evidence across outcomes to make a recommendation

Overall quality of evidence – Guideline panels

1. Consider **only** those outcomes that have been deemed **critical**.
2. If the quality of evidence is the **same** for all critical outcomes, then this becomes the overall quality of the evidence supporting the answer to the question.
3. If the quality of evidence **differs** across critical outcomes, it is logical that the overall confidence in effect estimates cannot be higher than the lowest confidence in effect estimates for any outcome that is critical for a decision. Therefore, the **lowest quality of evidence** for any of the critical outcomes determines the overall quality of evidence.
4. There is one exception to this rule that we will discuss.



do
no
harm



do
no net
harm

Benefits

Importance

Low co$t

Doable

Harms

Importance

High co$$$t

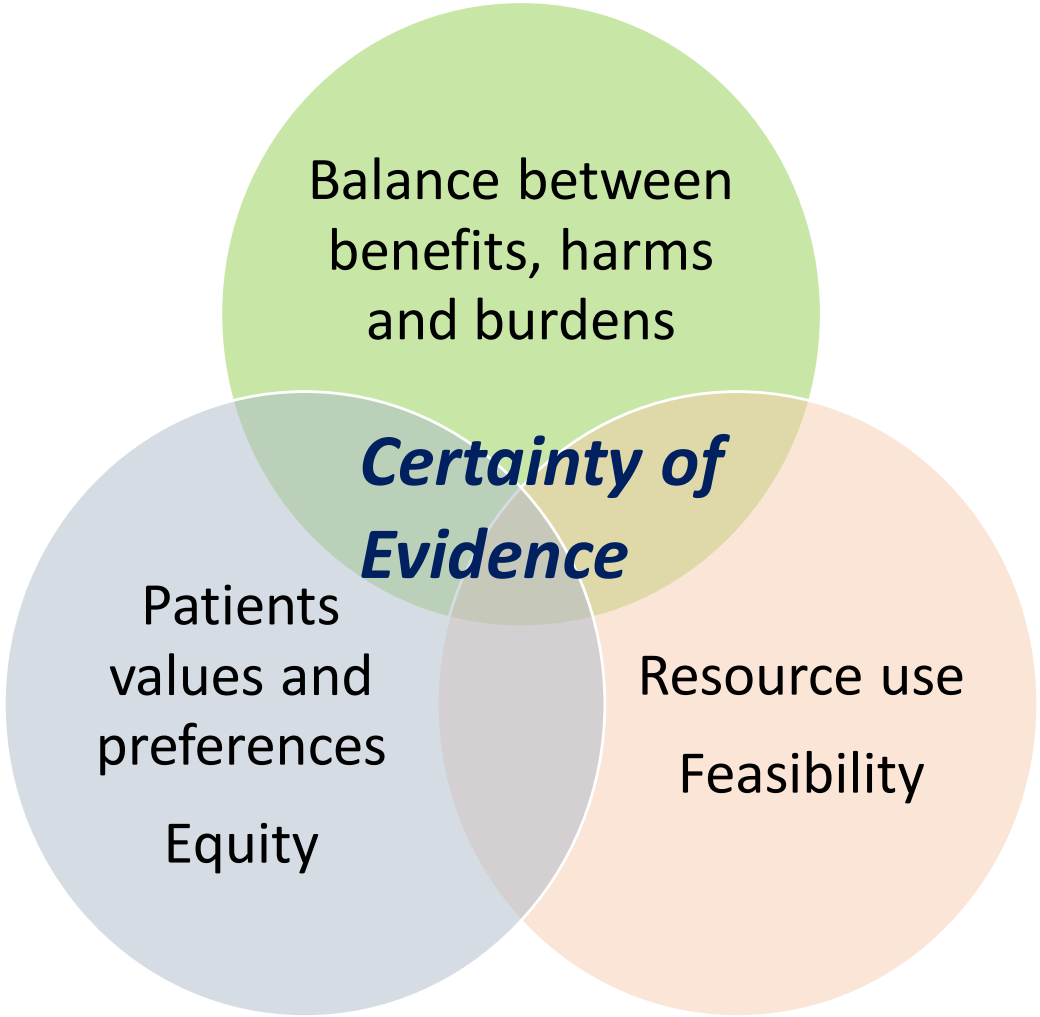
~~Doable~~

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

Forming recommendations with GRADE



Rating the importance of outcomes

- Need to understand that outcomes that are critical for decision making are identified
- Rating is done before, during and after the evidence review
- The rating may change in light of new information

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

Strong recommendation

- For patients: most individuals in this situation would want the recommended course of action, and only a small proportion would not.
- For clinicians: most individuals should follow the recommended course of action. Formal decision aids are not likely to be needed to help individual patients make decisions consistent with their values and preferences.
- For policy makers: the recommendation can be adopted as policy in most situations. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator.
- For researchers: the recommendation is supported by credible research or other convincing judgments that make additional research unlikely to alter the recommendation. On occasion, a strong recommendation is based on low or very low certainty in the evidence. In such instances, further research may provide important information that alters the recommendations.

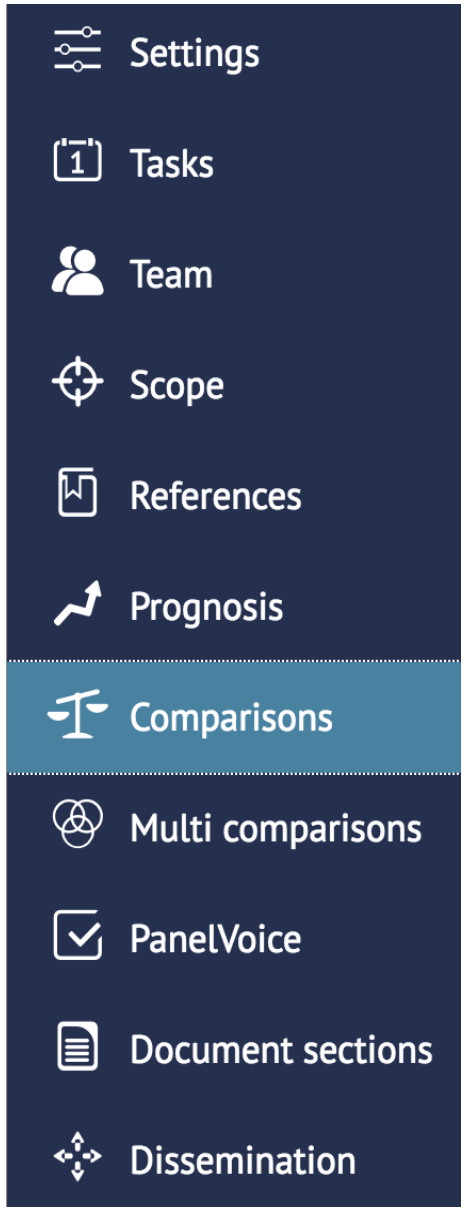
Conditional recommendation

- For patients: the majority of individuals in this situation would want the suggested course of action, but many would not. Decision aids may be useful in helping patients to make decisions consistent with their individual risks, values, and preferences.
- For clinicians: different choices will be appropriate for individual patients, and clinicians must help each patient arrive at a management decision consistent with the patient's values and preferences. Decision aids may be useful in helping individuals to make decisions consistent with their individual risks, values, and preferences.
- For policy makers: policy making will require substantial debate and involvement of various stakeholders. Performance measures about the suggested course of action should focus on whether an appropriate decision-making process is duly documented.
- For researchers: this recommendation is likely to be strengthened (for future updates or adaptation) by additional research.

Session 13: 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

GRADEpro GDT *interactive Evidence to Decision Frameworks*



- **Question**
  - Details
  - Subgroups
  - Background
- **Assessment**
  - Criteria
  - Judgements
  - Research evidence (HTA and Systematic Reviews)
  - Additional considerations
- **Conclusions**
  - Type of decision - recommendation
  - Justification
  - Implementation considerations - monitoring and evaluation
  - Research considerations

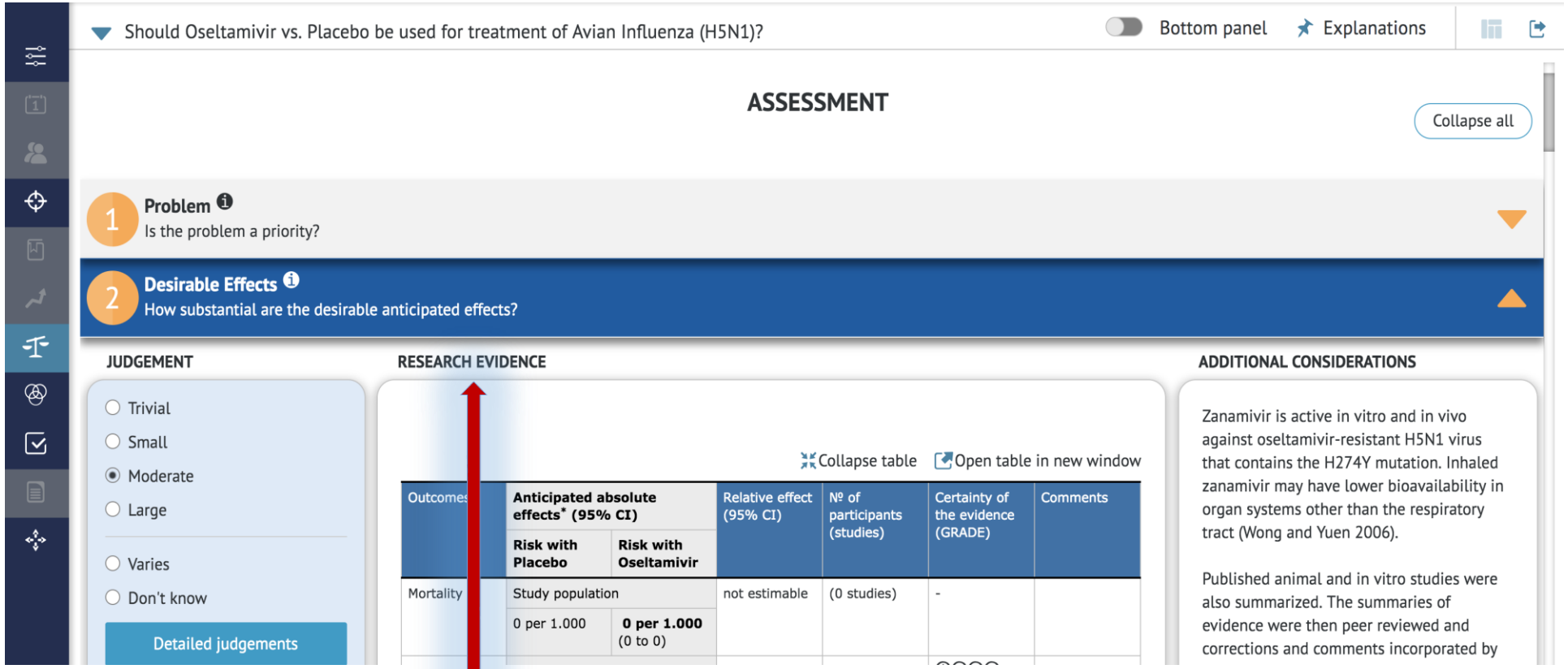
Presentation

- Guideline group meetings & informing coverage decisions
- Database of decision frameworks
- Interactive Decision Aids (iDeAs) apps

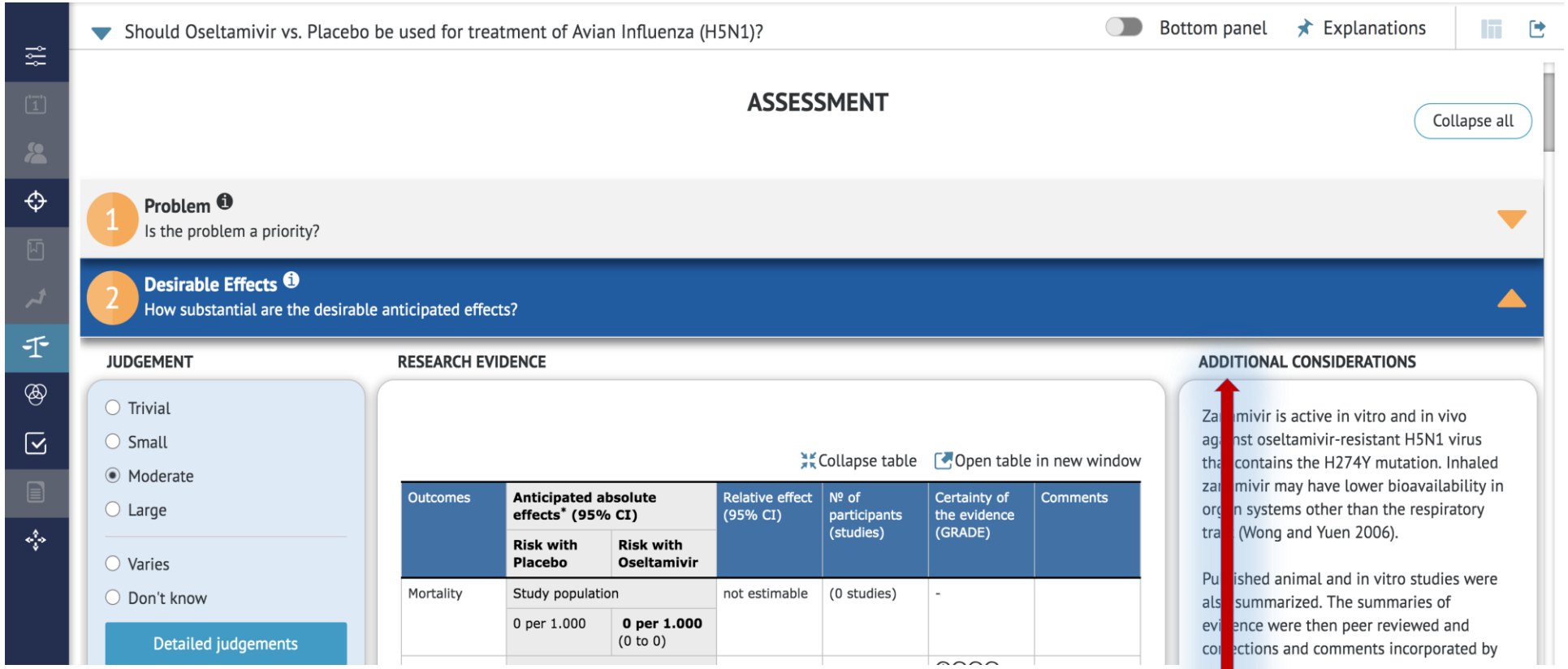
Decision making criteria

- Priority of problem
- Benefits and harms
- Certainty of evidence
- Values and Preferences
- Resources
- Equity
- Acceptability
- Feasibility

EtD frameworks

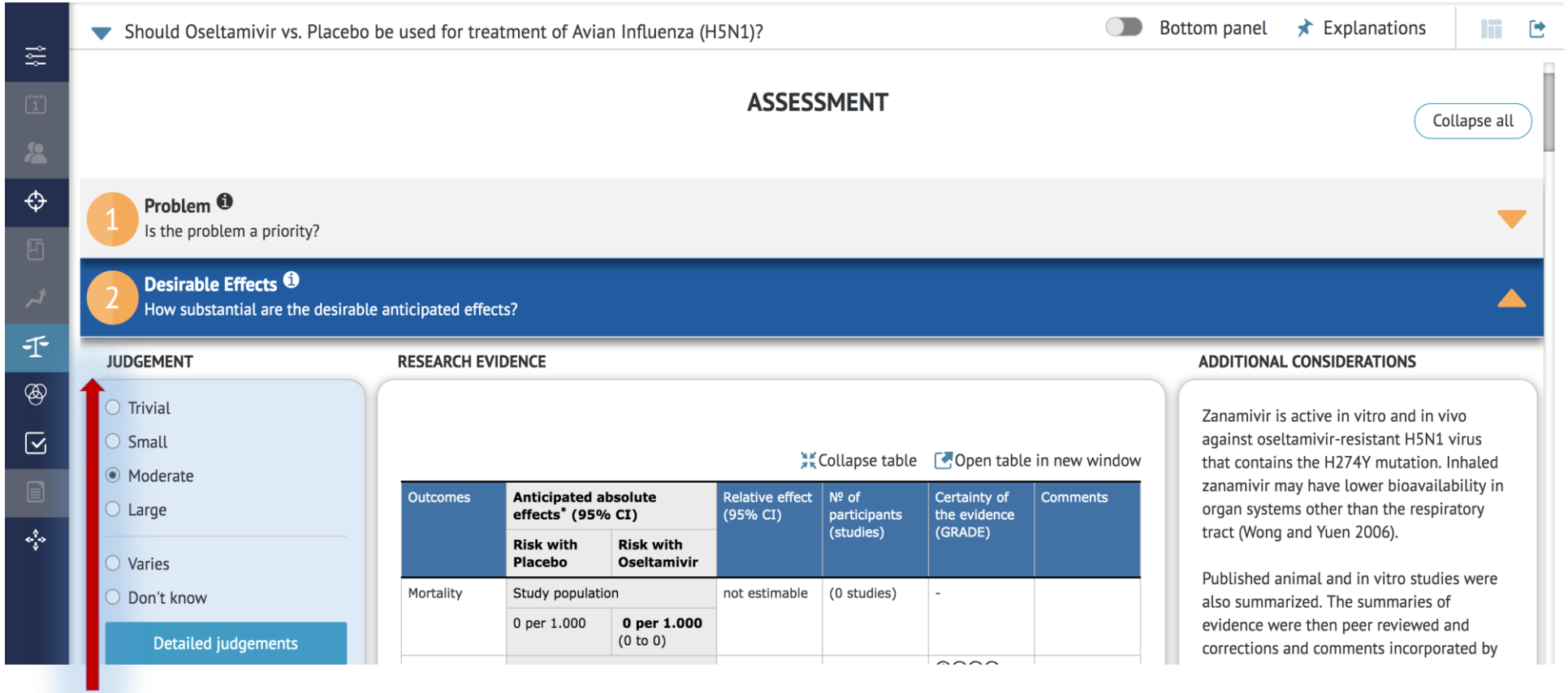


Discuss


Should Oseltamivir vs. Placebo be used for treatment of Avian Influenza (H5N1)?
Bottom panel
Explanations
ASSESSMENT
Collapse all
1 Problem
Is the problem a priority?
2 Desirable Effects
How substantial are the desirable anticipated effects?
JUDGEMENT
Trivial
Small
Moderate
Large
Varies
Don't know
Detailed judgements
RESEARCH EVIDENCE
Collapse table
Open table in new window
Outcomes
Anticipated absolute effects* (95% CI)
Risk with Placebo
Risk with Oseltamivir
Relative effect (95% CI)
№ of participants (studies)
Certainty of the evidence (GRADE)
Comments
Mortality
Study population
not estimable
(0 studies)
-
0 per 1.000
0 per 1.000 (0 to 0)
ADDITIONAL CONSIDERATIONS
Zanamivir is active in vitro and in vivo against oseltamivir-resistant H5N1 virus that contains the H274Y mutation. Inhaled zanamivir may have lower bioavailability in organ systems other than the respiratory tract (Wong and Yuen 2006).
Published animal and in vitro studies were also summarized. The summaries of evidence were then peer reviewed and corrections and comments incorporated by

Discuss

Add


Should Oseltamivir vs. Placebo be used for treatment of Avian Influenza (H5N1)?
Bottom panel
Explanations
ASSESSMENT
Collapse all
1 Problem
Is the problem a priority?
2 Desirable Effects
How substantial are the desirable anticipated effects?
JUDGEMENT
Trivial
Small
Moderate
Large
Varies
Don't know
Detailed judgements
RESEARCH EVIDENCE
Collapse table
Open table in new window
Outcomes
Anticipated absolute effects* (95% CI)
Risk with Placebo
Risk with Oseltamivir
Relative effect (95% CI)
№ of participants (studies)
Certainty of the evidence (GRADE)
Comments
Mortality
Study population
not estimable
(0 studies)
-
0 per 1.000
0 per 1.000 (0 to 0)
ADDITIONAL CONSIDERATIONS
Zanamivir is active in vitro and in vivo against oseltamivir-resistant H5N1 virus that contains the H274Y mutation. Inhaled zanamivir may have lower bioavailability in organ systems other than the respiratory tract (Wong and Yuen 2006).
Published animal and in vitro studies were also summarized. The summaries of evidence were then peer reviewed and corrections and comments incorporated by

Judge

No COI

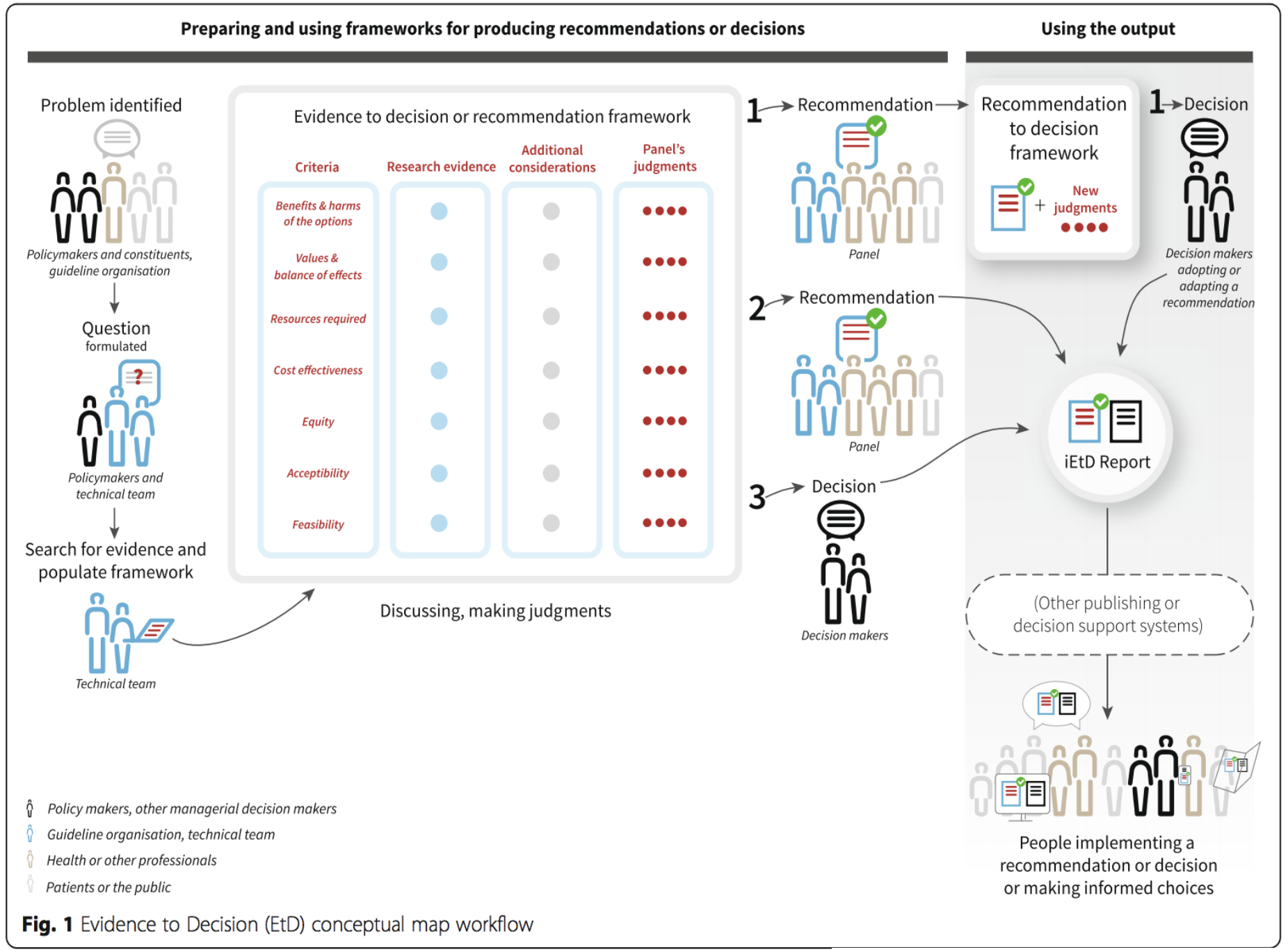
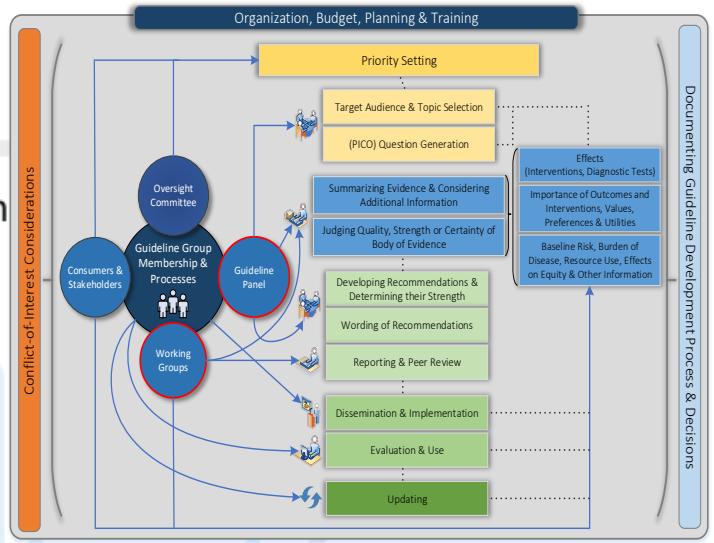


Fig. 1 Evidence to Decision (EtD) conceptual map workflow

Evidence sharing, updating & collaboration

Structured process for trustworthy guideline recs



Adapt, adopt, develop

Question

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

POPULATION:	hospital	**BACKGROUND:**
INTERVENTION:	Tight glycemic control	
COMPARISON:	placebo	
MAIN OUTCOMES:	Mortality; Mortality - Nondiabetic patients; Mortality - Diabetic patients; Atrial fibrillation; Stroke; Acute renal failure; Deep sternal infection; Length of stay;	
SETTING:		
PERSPECTIVE:		

Assessment

	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	**Is the problem a priority?** ○ No ○ Probably no ○ Probably yes ○ Yes ○ Varies ○ Don't know		
BLE EFFECTS	**How substantial are the desirable anticipated effects?** ○ Trivial ○ Small ○ Moderate		

DESIRA	○ Large ○ Varies ○ Don't know		
UNDESIRABLE EFFECTS	**How substantial are the undesirable anticipated effects?** ○ Large ○ Moderate ○ Small ○ Trivial ○ Varies ○ Don't know		
CERTAINTY OF EVIDENCE	**What is the overall certainty of the evidence of effects?** ○ Very low ○ Low ○ Moderate ○ High ○ No included studies		
VALUES	**Is there important uncertainty about or variability in how much people value the main outcomes?** ○ Important uncertainty or variability ○ Possibly important uncertainty or variability ○ Probably no important uncertainty or variability ○ No important uncertainty or variability		

BALANCE OF EFFECTS	**Does the balance between desirable and undesirable effects favor the intervention or the comparison?** ○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know		
ACCEPTABILITY	**Is the intervention acceptable to key stakeholders?** ○ No ○ Probably no ○ Probably yes ○ Yes ○ Varies ○ Don't know		
FEASIBILITY	**Is the intervention feasible to implement?** ○ No ○ Probably no ○ Probably yes ○ Yes ○ Varies ○ Don't know		

Summary of judgements

	JUDGEMENT							IMPLICATIONS
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know	
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know	
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know	
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies	
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability				
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	Don't know	
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know	
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know	

Conclusions

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

TYPE OF RECOMMENDATION	Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
	O	O	O	O	O
RECOMMENDATION					
JUSTIFICATION					
SUBGROUP CONSIDERATIONS					

Example Evidence to Decision Framework

https://goo.gl/PEfB72 for an example from a recent guideline

https://goo.gl/ztPoUP for patient material

American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Schünemann et al. Blood Adv. 2018 Nov 27;2(22):3198-3225

Example Evidence to Decision Framework

Example from the European Commission Initiative on the use of screening strategies for the detection of breast cancer:

https://ecibc.jrc.ec.europa.eu/recommendations/

https://ecibc.jrc.ec.europa.eu/recommendations/details/5bbf4f4e9a8bbc0fd5635575

NOTE: This is a controversial topic and the purpose here is not to debate the evidence, but see how the group came to their conclusion as an example EtD

European Commission. https://ecibc.jrc.ec.europa.eu/recommendations/

General Information

Should organised mammography screening vs. no mammography screening be used for early detection of breast cancer in women aged 45 to 49?

Background

Although mammography screening has both potential benefits and harms many countries have organised programmes for women aged 50 or older. However, there continues to be debate about recommendations for mammography screening, generally (Jorgensen 2009, Arie 2014), and particularly for women aged 40 to 49 (Petitti 2010).

Management of Conflicts of Interests (CoI): CoIs for all Guidelines Development Group (GDG) members were assessed and managed by the Joint Research Center (JRC) following an established procedure in line with the European Commission rules. GDG member participation in the development of the recommendations was according to CoI disclosure. Consequently, for this particular question, the following GDG members were recused from voting: Roberto d'Amico, Jan Danes, Axel Gräwingholt and Ruben van Engen.
For more information please visit: http://ecibc.jrc.ec.europa.eu/gdg-documents

Should organised mammography screening vs. no mammography screening be used for early detection of breast cancer in women aged 45 to 49?

Bottom panel | Explanations

Plain language statements ON | Absolute effect ON | Relative effect ON | Visual overview OFF

Outcomes	Plain language statements	Absolute Effect: With no mammography screening	Absolute Effect: With organised mammography screening	Relative effect (95% CI)	Certainty of the evidence GRADE
Breast cancer mortality (short case accrual) for women under 50 Follow-up: 16.8 years (•) Low ( ) High	*Empty summary*	400 per 100000	356 per 100000 Difference: 44 fewer per 100000 patients (95% CI: 84 fewer to 4 more per 100000 patients) Based on data from 348112 patients in 8 studies	**RR 0.89** (0.79 to 1.01)	⨁⨁⨁◯ MODERATE Due to serious imprecision.
Breast cancer mortality (longest case accrual available) for women under 50 Follow-up: 15.2 years	*Empty summary*	480 per 100000	437 per 100000 Difference: 43 fewer per 100000 patients (95% CI: 91 fewer to 19 more per 100000 patients) Based on data from 348076 patients in 8 studies	**RR 0.91** (0.81 to 1.04)	⨁⨁⨁◯ MODERATE Due to serious imprecision.
All-cause mortality Follow-up: 9.9 years					

1 Problem

Is the problem a priority?

JUDGEMENT

- ○ No
- ○ Probably no
- ○ Probably yes
- ◉ Yes
- ○ Varies
- ○ Don't know

Detailed judgements

RESEARCH EVIDENCE

Breast cancer is the second most common cancer in the world and, by far, the most frequent cancer among women with an estimated 1.67 million new cancer cases diagnosed in 2012—accounting for 25% of all cancers (GLOBOCAN 2012). Breast cancer ranks as the fifth leading cause of cancer death worldwide and it is the second leading cause of cancer-related death in developed regions (GLOBOCAN 2012). In the European Union, 367 090 women were diagnosed of breast cancer and 92 000 women died from the disease in 2012 (Ferlay 2013). Breast cancer ranks fourth among the top five cancers with the highest disease burden (Tsilidis 2016). Annual incidence of breast cancer in the EU among women aged 45 to 49 is 1.7 per 1 000 and mortality is 0.2 per 1 000 per year (GLOBOCAN 2012)

ADDITIONAL CONSIDERATIONS

2 Desirable Effects

How substantial are the desirable anticipated effects?

JUDGEMENT

- ○ Trivial
- ○ Small
- ◉ Moderate
- ○ Large
- ○ Varies
- ○ Don't know

Detailed judgements

RESEARCH EVIDENCE

Collapse table Open table in new window

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with no mammography screening	Risk difference with organised mammography screening
Breast cancer mortality (short case accrual) for women under 50 follow up: mean 16.8 years	348112 (8 RCTs)	⨁⨁⨁◯ MODERATE [a,b,c]	**RR 0.89** (0.79 to 1.01)	Low	
				400 per 100,000 [d]	**44 fewer per 100,000** (84 fewer to 4 more)
				High	

ADDITIONAL CONSIDERATIONS

These studies used an 'intention-to-treat' analysis thus, a per protocol approach would lead to larger absolute effects.

GDG members mentioned that modelling studies describing quality and duration of 'life gained' should be considered.

Long case accrual may dilute the effect of the intervention as for some trials it will include cases diagnosed after closure of the trial when both arms are receiving the same intervention. Therefore, we performed a

3 Undesirable Effects

How substantial are the undesirable anticipated effects?

JUDGEMENT

- ○ Large
- ◉ Moderate
- ○ Small
- ○ Trivial
- ○ Varies
- ○ Don't know

Detailed judgements

RESEARCH EVIDENCE

Collapse table | Open table in new window

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with no mammography screening	Risk difference with organised mammography screening
Breast cancer mortality (short case accrual) for women under 50 follow up: mean 16.8 years	348112 (8 RCTs)	⨁⨁⨁◯ MODERATE [a,b,c]	**RR 0.89** (0.79 to 1.01)	Low	
				400 per 100,000 [d]	**44 fewer per 100,000** (84 fewer to 4 more)
				High	

ADDITIONAL CONSIDERATIONS

Overdiagnosis and its magnitude are not greatly influenced by age at first screening.

Overdiagnosis estimates from both CNBSS1 and CNBSS2 may have been overestimated by subsequent screening in the population (both organised and opportunistic) after screening ceased in the CNBSS in 1988. Thus, while at 25 years of follow-up a non-statistically significant excess of all breast cancers was observed in the intervention arm of CNBSS trials (difference 2.6; 95%CI -0.8 to 5.9), the

4 Certainty of evidence

What is the overall certainty of the evidence of effects?

JUDGEMENT

- ○ Very low
- ○ Low
- ◉ Moderate
- ○ High
- ○ No included studies

Detailed judgements

RESEARCH EVIDENCE

The overall certainty (i.e. quality) of the evidence was considered moderate, as this was the lowest quality among the critical outcomes—namely, breast cancer mortality and overdiagnosis.

ADDITIONAL CONSIDERATIONS

5 Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT

- ○ Important uncertainty or variability
- ◉ Possibly important uncertainty or variability
- ○ Probably no important uncertainty or variability
- ○ No important uncertainty or variability

- ○ No known undesirable outcomes

Detailed judgements

RESEARCH EVIDENCE

A systematic review shows that participants place a low value on the psychosocial and physical effects of false-positive results and overdiagnosis (JRC Technical Report PICO 10-11, contract FWC443094012015; available upon request). Women generally consider these undesirable effects acceptable (low confidence in evidence). However, these findings are of limited value mainly given the significant concerns regarding the adequacy of the information provided to women, in order to make an informed decision about participation. Also, acceptability of false positive results is based on studies of participants who have already received a false positive result. Their preferences may differ from the general population. Another finding is that breast cancer screening represents a significant burden for some participants due to the associated psychological distress and inconvenience (moderate confidence in evidence).
Regarding breast cancer diagnosis, very limited data is available addressing people's views. One of the main themes identified in the literature is that people disvalue highly the anxiety caused by delays in receiving diagnostic results, or by a lack of understanding of the tests due to suboptimal communication with physicians (moderate confidence in evidence). Also, people have a higher overall preference towards more comfortable, brief diagnostic procedures (moderate confidence in evidence). (JRC Technical Report PICO 10-11, contract FWC443094012015; available upon request)

ADDITIONAL CONSIDERATIONS

6 Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT

- ○ Favors the comparison
- ○ Probably favors the comparison
- ○ Does not favor either the intervention or the comparison
- ◉ Probably favors the intervention
- ○ Favors the intervention

- ○ Varies

RESEARCH EVIDENCE

ADDITIONAL CONSIDERATIONS

GDG members agreed that first screening at age 45 had moderate desirable health effects and moderate undesirable health effects; however, consensus was not reached regarding the balance between these two.

Sixteen members voted that the balance probably favours the intervention; five members voted that the balance does not favour either the intervention or the comparison; and one voting member abstained.

7 Resources required

How large are the resource requirements (costs)?

JUDGEMENT

- ○ Large costs
- ◉ Moderate costs
- ○ Negligible costs and savings
- ○ Moderate savings
- ○ Large savings
- ○ Varies
- ○ Don't know

Detailed judgements

RESEARCH EVIDENCE

Differences in required resources for mammography screening versus no screening in women aged 40 to 49 in the studies analysed may be related to the inclusion or not of costs related to the screening process, diagnostic techniques, treatment and follow-up of diagnosed women (Madan, 2010 and Sankatsing, 2015).

Screening costs for a cohort of 10 000 women have been estimated to be £420 000 in the UK. The cost of diagnosis for positive results would be £70 000, and screening would lead to a saving of £17 000 in treatment costs (£480 per screen-detected cancer, calculated from the difference in treatment costs for the trial control and intervention arms), giving a net screening cost of £473 000 per 10 000 screened (using a 3.5% discount rate) (Madan et al. 2010)

Based on the results of Sankatsing et al. (2015), the total cost of breast cancer diagnosis, treatment and death in the absence of screening were estimated at €1 161 008 per 1 000 women, followed over their lifetime. The total cost of extended biennial screening in women aged 40 to 49 would increase to €306 590 per 1 000 women (using a 3.5% discount rate).

ADDITIONAL CONSIDERATIONS

Varies by screening interval and by country and by the presence of opportunistic screening.

GDG members judged the cost to be at least moderate.

However, substantial differences could be observed in European countries without population-based screening programmes or in those programmes with different screening policies.

8 Certainty of evidence of required resources

What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT

- ○ Very low
- ◉ Low
- ○ Moderate
- ○ High
- ○ No included studies

Detailed judgements

RESEARCH EVIDENCE

The certainty of the evidence of resource requirements is low due to the study design of the included studies which were modelling studies based on observational data. In addition, the following differences were observed: In Madan et al. (2010) model parameters were based on data from a triennial screening while data from Sankatsing et al. (2015) corresponded to biennial screening. The studies reported costs of screening, diagnosis, and treatment. Based on their data, total costs per extension of one round of triennial screening would be £47 per woman in the UK (2006 value) which is similar to the €61.3 per one round of biennial screening in the Netherlands (2014 value).

The formal assessment of the certainty in the evidence for cost and resources used was made using GRADE criteria and reported in the Evidence Profile (JRC Technical Report PICO 14-15, contract FWC443094012015; available upon request).

ADDITIONAL CONSIDERATIONS

Both studies assessed the extension of their current population-based screening programmes. As previously stated, substantial differences could be observed in European countries without population-based screening programmes or in those programmes with different screening policies.

9 Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT

- ○ Favors the comparison
- ○ Probably favors the comparison
- ○ Does not favor either the intervention or the comparison
- ○ Probably favors the intervention
- ○ Favors the intervention
- ◉ Varies
- ○ No included studies

RESEARCH EVIDENCE

Based on the evidence provided by Sankatsing et al. (2015), the extension of biennial mammography screening starting at age 40 appears to be cost-effective at a 'willingness-to-pay' of €20 000 per life year gained (LYG) with an incremental cost-effectiveness ratio (ICER) of €10 826 per LYG starting at age 40 instead of age 45.

On the contrary, based on the evidence provided by Madan et al. (2010), the extension of triennial mammography screening in women aged 47 to 49 does not appear to be cost-effective at a 'willingness-to-pay' of £20 000 per Quality Adjusted Life Years (QALY). The probability of being cost-effective at this threshold was low (29%). The ICER per QALY gained for triennial screening was £27 400.

ADDITIONAL CONSIDERATIONS

Differences in the cost-effectiveness results could be explained by the differences in setting, policy of the screening programmes, outcomes measures and type of technology used.

Whereas Sankatsing et al. (Sankatsing 2015) reported the ICER per LYG in the Netherlands, (Madan et al. 2010) reported the ICER per QALY in the UK. The negative effects of false-positive results in the UK significantly reduced QALYs.

Sankatsing et al. assessed digital

10 Equity

What would be the impact on health equity?

JUDGEMENT

- ○ Reduced
- ○ Probably reduced
- ○ Probably no impact
- ○ Probably increased
- ○ Increased
- ○ Varies
- ◉ Don't know

Detailed judgements

RESEARCH EVIDENCE

ADDITIONAL CONSIDERATIONS

A systematic review on this topic was not carried out. However, the utilisation of cancer screening services may largely depend on the availability of national public screening programmes; although European findings highlight that inequalities are larger in countries without population-based screening programmes (Palència, 2010).

11 Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT

- ○ No
- ○ Probably no
- ○ Probably yes
- ◉ Yes
- ○ Varies
- ○ Don't know

Detailed judgements

RESEARCH EVIDENCE

A systematic review (JRC Technical Report PICO 16-17, contract FWC443094032016; available upon request) found the following barriers associated with breast cancer screening: (*a*) lack of knowledge and misperceptions regarding preventive medicine and breast health (high confidence in evidence), (*b*) poor communication skills of healthcare providers (high confidence in evidence), (*c*) poor accessibility to breast screening, especially among women with disabilities (high confidence in evidence), (*d*) fear and stress related to the procedure and the possibility of cancer diagnosis (high confidence in evidence), (*e*) pain and discomfort during the procedure (moderate confidence in evidence), (*f*) embarrassment and shyness during the procedure (moderate confidence in evidence), (*g*) lack of support and encouragement from family members, caregivers and social network (moderate confidence in evidence), (*h*) lack of information regarding the available resources (low confidence in evidence) and (*i*) low prioritisation of breast cancer screening (low confidence in evidence).

ADDITIONAL CONSIDERATIONS

Some GDG members described that some professional groups may find a screening programme not acceptable due to their financial interests.

12 Feasibility

Is the intervention feasible to implement?

JUDGEMENT

- ○ No
- ○ Probably no
- ◉ Probably yes
- ○ Yes
- ○ Varies
- ○ Don't know

RESEARCH EVIDENCE

ADDITIONAL CONSIDERATIONS

A systematic review on this topic was not carried out. Some countries do not have screening programmes mainly due to lack of resources and also infrastructure.

Given that this recommendation would be additive to screening in older age groups (50 to 69), it was judged as being probably feasible to implement.

Should organised mammography screening vs. no mammography screening be used for early detection of breast cancer in women aged 45 to 49?

Bottom panel | Explanations

CRITERIA	SUMMARY OF JUDGEMENTS						IMPORTANCE FOR DECISION
PROBLEM	No	Probably no	Probably yes	**Yes**	Varies	Don't know	
DESIRABLE EFFECTS	Trivial	Small	**Moderate**	Large	Varies	Don't know	
UNDESIRABLE EFFECTS	Large	**Moderate**	Small	Trivial	Varies	Don't know	
CERTAINTY OF EVIDENCE	Very low	Low	**Moderate**	High	No included studies		
VALUES	Important uncertainty or variability	**Possibly important uncertainty or variability**	Probably no important uncertainty or variability	No important uncertainty or variability	No known undesirable outcomes		
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	**Moderate costs**	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	**Low**	Moderate	High	No included studies		
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	**Yes**	Varies	Don't know	
FEASIBILITY	No	Probably no	**Probably yes**	Yes	Varies	Don't know	

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	◉	○

Should organised mammography screening vs. no mammography screening be used for early detection of breast cancer in women aged 45 to

Presentation for Clinicians

RECOMMENDATION

WE SUGGEST THE INTERVENTION

For asymptomatic women aged 45 to 49 with an average risk of breast cancer, the ECIBC's Guideline Development Group (GDG) suggests mammography screening over no mammography screening, in the context of an organised screening programme (conditional recommendation, moderate certainty in the evidence).

Background | Subgroup considerations | Justification | Detailed justification | Summary of findings

Background

Although mammography screening has both potential benefits and harms many countries have organised programmes for women aged 50 or older. However, there continues to be debate about recommendations for mammography screening, generally (Jorgensen 2009, Arie 2014), and particularly for women aged 40 to 49 (Petitti 2010).

Activity 7 and Activity 8

- Consult your handbook

For further information

RESEARCH METHODS AND REPORTING

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

For numbered affiliations see end of article.

Correspondence to:
A D Oxman oxman@online.no

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;353:i2016
http://dx.doi.org/10.1136/bmj.i2016

Introduction

Healthcare decision making is complex. Decision-making processes and the factors (criteria) that decision makers should consider vary for different types of decisions, including clinical recommendations, coverage decisions, and health system or public health recommendations or decisions.[1-4] However, some criteria are relevant for all of these decisions, including the anticipated effects of the options being considered, the certainty of the evidence for those effects (also referred to as quality of evidence or confidence in effect estimates), and the costs and feasibility of the options. Decision makers must make judgments about each relevant factor, informed by the best evidence that is available to them.

Often, the processes that decision makers use, the criteria that they consider and the evidence that they use to reach their judgments are unclear.[5-8] They may omit important criteria, give undue weight to some criteria, or not use the best available evidence. Systematic and transparent systems for decision making can help to ensure that all important criteria are considered and that the best available research evidence informs decisions.

Clinicians depend on clinical practice guidelines. Rigorously developed guidelines synthesise the available relevant research, facilitating the translation of evidence into recommendations for clinical practice.[9] However, the quality of guidelines is often suboptimal.[10 11] If guidelines are not developed systematically and transparently, clinicians are not able to decide whether to rely on them or to explore disagreements when faced with conflicting recommendations.[12]

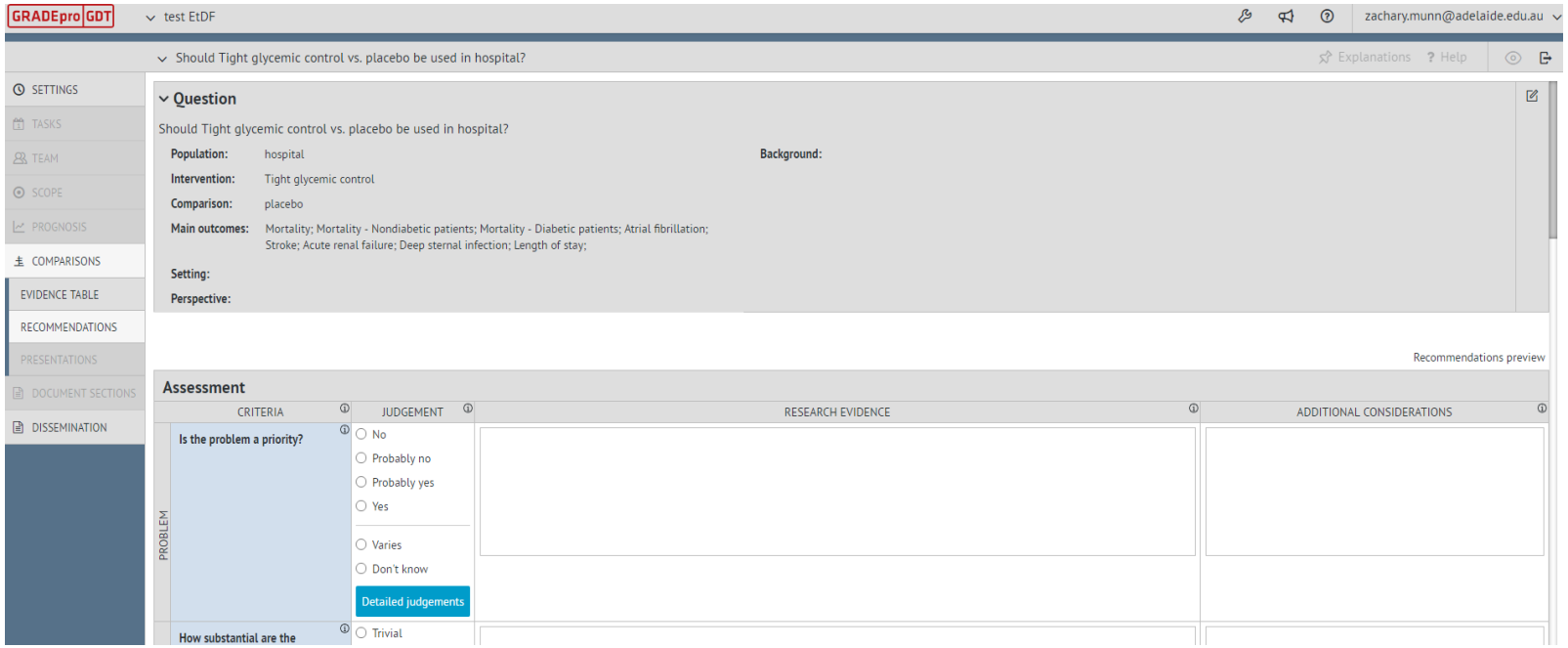
The GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group has previously developed and refined a system to assess the certainty of evidence of effects and strength of recommendations.[13-15] More than 100 organisations globally, including the World Health Organization, the Cochrane Collaboration, and the National Institute for Health and Care Excellence (NICE) now use or have adopted the principles of the GRADE system. Recently, through the DECIDE (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence) project (http://www.decide-collaboration.eu),[16] funded by the European Union, the GRADE Working Group has developed the Evidence to Decision (EtD) frameworks to support the process of moving from evidence to decisions. We have developed EtD frameworks for making clinical recommendations, coverage decisions, and health system or public health recommendations and decisions. The frameworks build on the GRADE approach to assessing the strength of recommendations.[17-19]

We developed EtD frameworks using an iterative process that is described in the project protocol.[16] The starting point for EtD frameworks was the GRADE Working Group's approach for moving from evidence to clinical recommendations.[17-19] We iteratively developed the frameworks based on reviews of relevant literature,[1-4] brainstorming, feedback from stakeholders,[20] application of EtD frameworks to a variety of recommendations and decisions, and user testing. We strove for consistency across EtD frameworks for different types of decisions, but, because of differences in the nature of the decisions, there are some differences in the frameworks. In appendix 1, we have provided a glossary of terms used in EtD frameworks, including certainty of the evidence, decisions, recommendations, and strength of recommendations.

This series of two articles describing the EtD frameworks is targeted at guideline developers and users of guidelines. This first article introduces the frameworks. It describes their purpose, development, and structure. It also describes how different organisations can adapt the frameworks to their own contexts and decision-making processes. The second article presents the framework for clinical recommendations.[21]

SUMMARY POINTS

- Clinicians, guideline developers, and policymakers sometimes neglect important criteria, give undue weight to criteria, and do not use the best available evidence to inform their judgments
- Explicit and transparent systems for decision making can help to ensure that all important criteria are considered and that decisions are informed by the best available research evidence
- The purpose of Evidence to Decision (EtD) frameworks is to help people use evidence in a structured and transparent way to inform decisions in the context of clinical recommendations, coverage decisions, and health system or public health recommendations and decisions
- EtD frameworks have a common structure that includes formulation of the question, an assessment of the evidence, and drawing conclusions, though there are some differences between frameworks for each type of decision
- EtD frameworks inform users about the judgments that were made and the evidence supporting those judgments by making the basis for decisions transparent to target audiences
- EtD frameworks also facilitate dissemination of recommendations and enable decision makers in other jurisdictions to adopt recommendations or decisions, or adapt them to their context



Cochrane Czech Republic

JBI

The Czech Republic (Middle European) Centre for Evidence-Based Healthcare

Masaryk University

GRADE

Conclusion

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
- Refer to workbook for additional resources

GRADE project groups

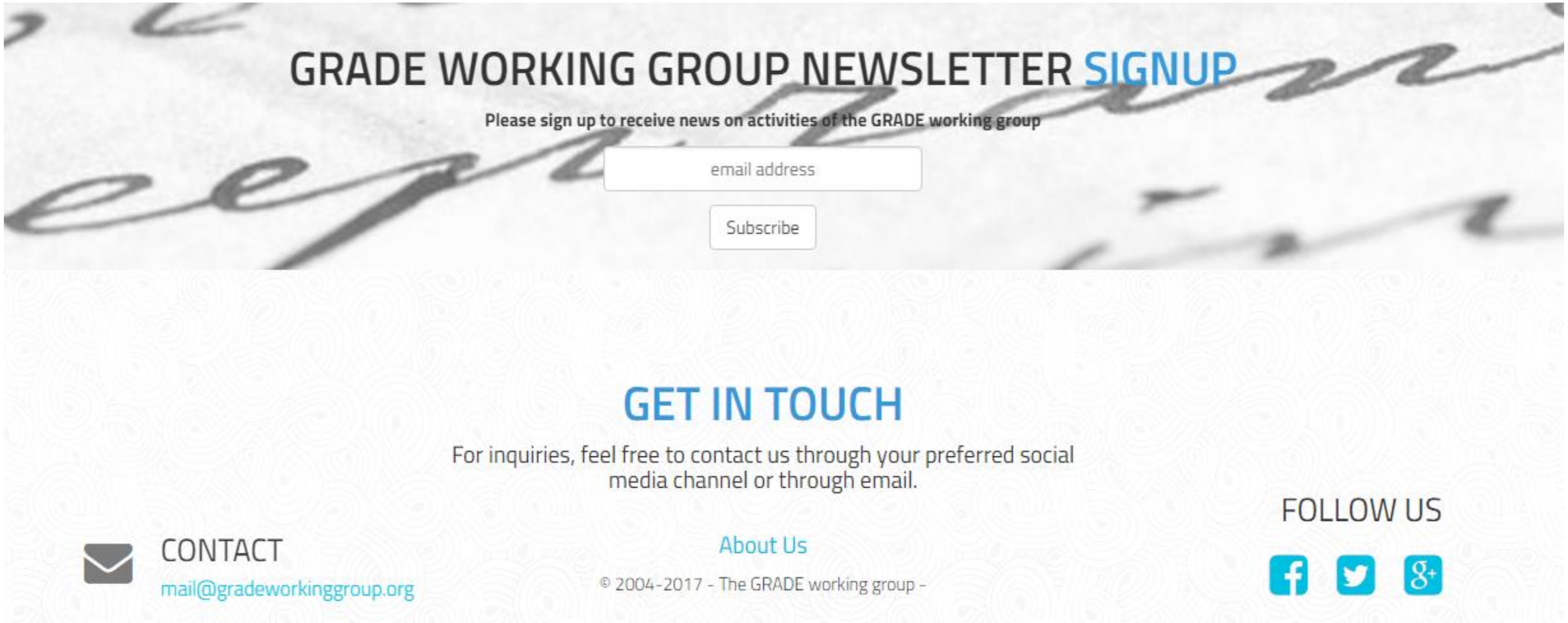
1. Environmental and Occupational Health	19. Complex interventions
2. Prognosis	20. GRADE Dispute
3. Outcomes valuation	21. NRS Risk of bias
4. Technology	22. Certainty in evidence
5. GRADE-CERQual	23. Philosophy of GRADE
6. Diagnosis	24. Modelling
7. Network Meta-analysis	25. Genetic Epidemiology
9. Training and Credentialing	26. Performance measurement/quality improvement (QI)
10. Public Health	27. Standardised wording of results and interpretation
21. Rare diseases	28. Overview of rerviews
12. Communication	29. Implementation of guidelines
13. Evidence to Decisions	30. Time-to-event outcomes
14. Equity	31. Stakeholders involvment
15. Algorithms and pathways	32. Rapid guidelines
17. Biosimilars	
18. GRADE for animal studies	

Final questions?

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

Get involved!



MUNI MED Institute of Biostatistics and Analyses
Czech National Centre for Evidence-Based Healthcare

Cochrane Czech Republic

JBI The Czech Republic (Middle European) Centre for Evidence-Based Healthcare

Masaryk University GRADE