Holger Schünemann

Chair and Professor

Department of Health Research Methods, Evidence & Impact*

Professor of Medicine | McMaster University

Director, Cochrane Canada

***formerly „Clinical Epidemiology and Biostatistics“**

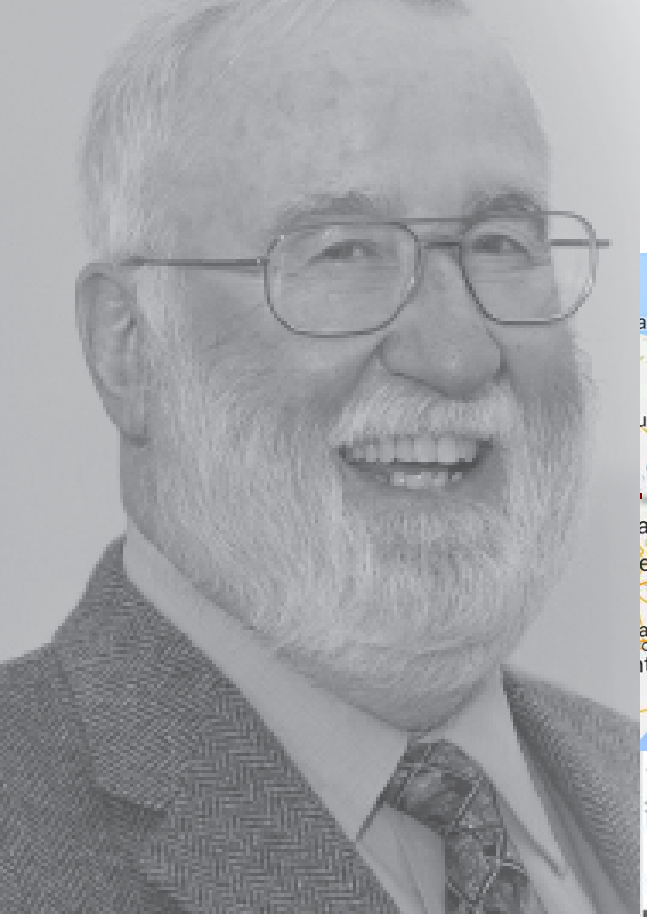
@schunemann_mac

Assessing the certainty of evidence for informed decisions - beyond interventions: GRADE and IQWiG approaches

GRADE working group

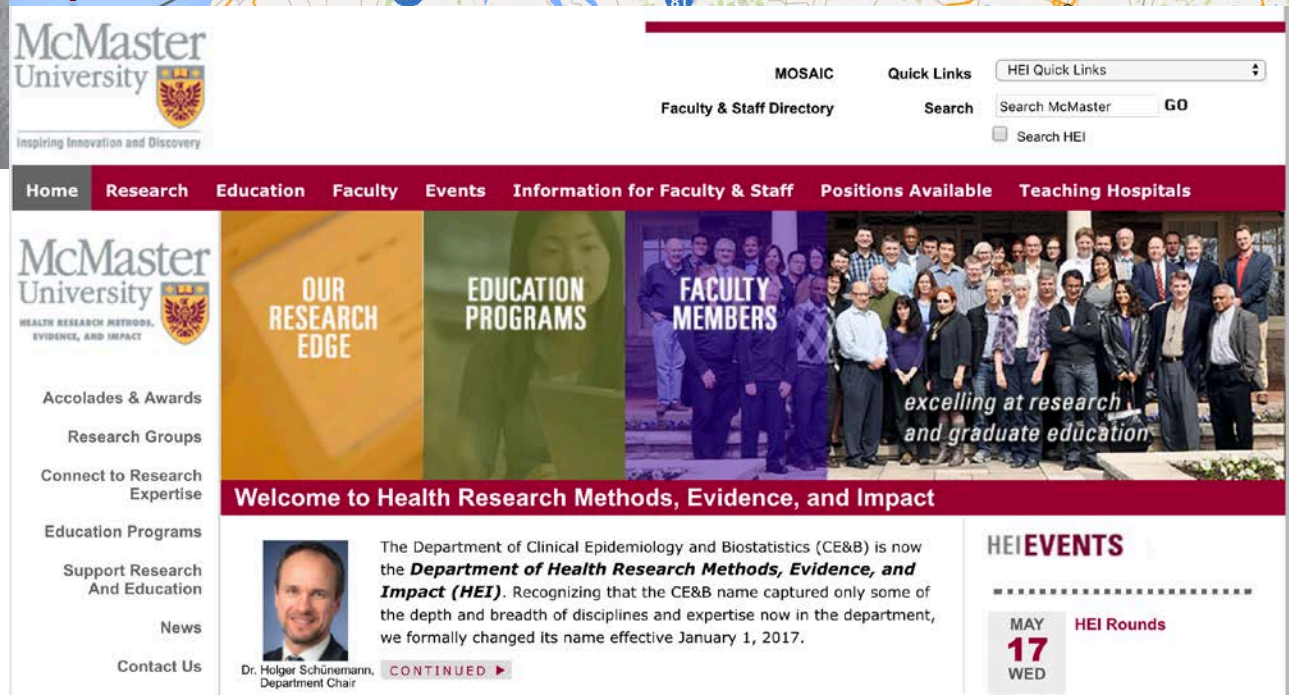
Cochrane Canada

McMaster University
HEALTH RESEARCH METHODS, EVIDENCE, AND IMPACT



“Birthplace of evidence-based medicine and problem based learning”

1967 - http://www.fhs.mcmaster.ca/ceb/



Disclosures

Cochrane Canada - Director

GRADE working group - Co-chair

IQWiG Scientific Board – past member

No direct financial COI

Views expressed my own

IQWiG

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Institute for Quality and Efficiency in Health Care

General Methods[a]

Version 4.2 of 22 April 2015

Today

1. Intro to GRADE
2. Overview of what evidence is needed to make informed, evidence-based health decisions
3. How IQWiG deals with assessing the evidence and how this compares with GRADE
4. When modeling is required and how certain we can be in modeled evidence

In the context of “what’s next”

The origin of evidence appraisal (systems)

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:

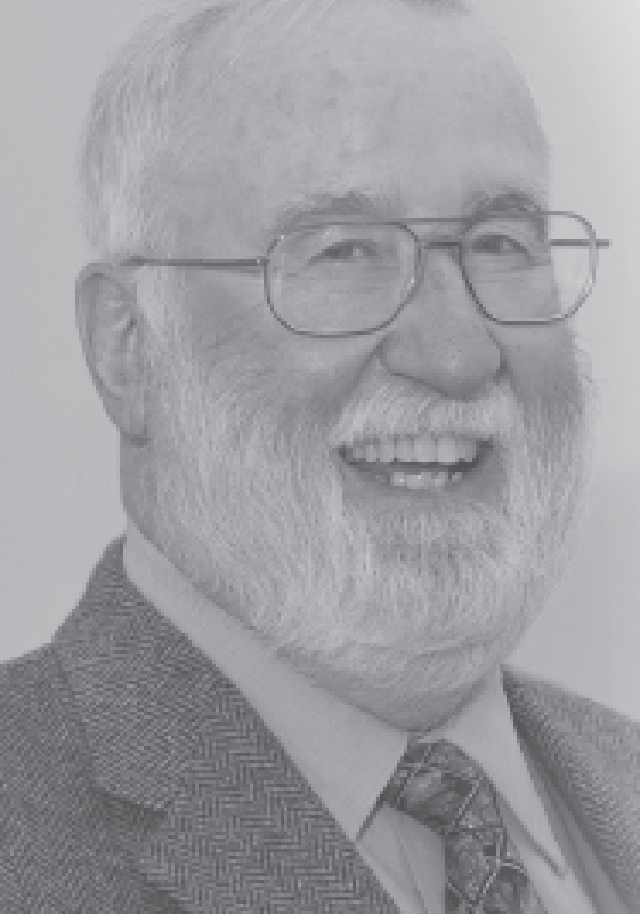
A: There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination.

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 consideration in a periodic health examination.



Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents

D. L. Sackett M.D.

Introduction

What rules of evidence ought to apply when expert committees meet to generate recommendations for the clinical management of patients? Should only the thoroughly validated results of randomized clinical trials be admissible to avoid or minimize the application of useless or harmful therapy? Or, to maximize the potential benefits to patients (including those possible from unproved remedies), ought a synthesis of the experiences of seasoned clinicians form the basis for such recommendations?

Ample precedent exists for the latter approach even when attempts are made to replace it.[1] However, for the following three reasons, the nonexperimental evidence that forms the recalled experiences of seasoned clinicians will tend to overestimate efficacy:

1. Favorable treatment responses are more likely to be recognized and remembered by clinicians when their patients comply with treatments and keep follow-up appointments. However, there are already five documented instances in which compliant patients in the *placebo* groups of randomized trials exhibited far more favorable outcomes (including survival) than their noncompliant companions.[2-6] Because high compliance is therefore a marker for better outcomes, even when treatment is useless, our uncontrolled clinical experiences often will cause us to conclude that compliant patients must have been receiving efficacious therapy.

agents in an effort to halt the progression and complications of thromboembolism. For many of the disorders under consideration here, randomized control trials have never been (and, arguably, never could be) carried out, and the only information base for generating some of the recommendations comes from uncontrolled clinical observations.

What this does mean, however, is that it is important, whenever possible, to base firm recommendations (and especially those involving risk to patients) on the results of rigorously controlled investigations and to be much more circumspect when recommendations rest only on the results of uncontrolled clinical observations. This approach was adopted by the conference participants and led to the definition and adoption of both Levels of Evidence and Grades of Recommendations.

Levels of Evidence

The participants in this undertaking, when summarizing what was known about the causes, clinical course, and management of a given clinical entity, specified the level of evidence that was being used in each case, according to the following classification:

Level I: Randomized trials with low false-positive (α) and low false-negative (β) errors (high power)

By "low false-positive (α) error" is meant a "positive" trial that demonstrated a statistically significant benefit from experimental treatment. For example, there have now been two randomized trials in which aspirin produced very large, statistically significant reductions in the risk of stroke and death among patients with transient ischemic attacks.

By "low false-negative (β) error (high power)" is meant a "negative" trial that demonstrated no effect of therapy, yet was large enough to exclude the possibility of a clinically important benefit (*ie*, had very narrow 95% confidence limits

Chest 1986

GRADE working group

After 30 years of increasing confusion, GRADE developed a unifying, transparent and sensible system for grading the certainty of evidence and making decisions

- WHO, NICE, CADTH, CDC, AHRQ, professional societies, academic institutions
- For systematic reviews, HTA and guidelines
- International & diverse contributors (>500)
- 2008 BMJ series; 2011 JCE series – over 30,000 cites
- Various other publications (incl. GRADE Handbook)
- IT applications: GRADEpro | GDT

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, AJRCCM 2006, Chest 2006, BMJ 2008, JCE 2011-2017

Over 100 organizations adopted or use GRADE

Open membership – free: www.gradeworkinggroup.org

Breast cancer services quality

Holger

ecibc.jrc.ec.europa.eu/home

Apps · Google URL Shortener · Tngs · IBM · EST · Soc · ASH · CE&B · Bank · Airlines · GRADE · Chrane · Bike · Mac · F'dtions · Other Bookmarks

European Commission

EUROPEAN COMMISSION INITIATIVE ON BREAST CANCER

European Commission > EU Science Hub > ECIBC > Home

Home | Who we are | Quality Assurance | European Guidelines | News & Events | Publications | Contribute! | ECIBC for You

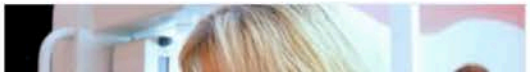
EC Initiative on Breast Cancer (ECIBC)

Guidelines and Quality Assurance scheme for Breast Cancer

The European Commission, in response to the Council of the European Union's conclusions on reducing the burden of cancer, initiated a ground-breaking

Report of a European Survey on the Implementation of Breast Units

Breast units implementation? À la carte, survey says. Breast units patchy panorama confirms ECIBC evidence-based approach is needed.



European Breast Guidelines

ECIBC recommendations for breast cancer screening and diagnosis.

EUROPEAN COMMISSION INITIATIVE ON BREAST CANCER

European Commission

European Commission > EU Science Hub > ECIBC > Recommendations

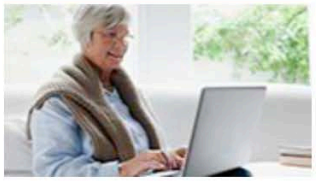
Home | Recommendations

Recommendations on Breast Cance

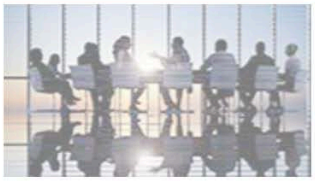
Read me

General Information

I'm a patient/individual



I'm a policy maker



If you are aged 40 to 44, should you attend an organised mammography screening programme?

Recommendation | Justification | Considerations | Assessment | Bibliography

Recommendation

The ECIBC guidelines suggests not providing mammography screening to women between 40 and 44 years old who are at average risk of breast cancer and do not have symptoms.

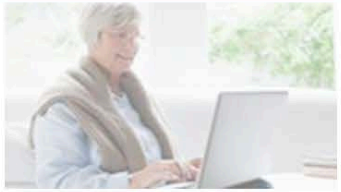
Recommendation strength

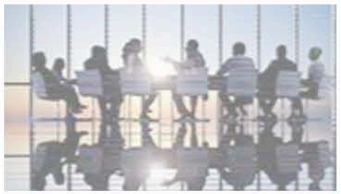
Conditional recommendation against the intervention*

Recommendations on Breast Cancer Screening

Read me | I'm a patient/individual | I'm a professional | I'm a policy maker

General Information





Should organised mammography screening vs. no mammography screening be used for early detection of breast cancer in women aged 40 to 44?

Recommendation | Justification | Considerations | Assessment | Bibliography

Recommendation

For asymptomatic women aged 40 to 44 with an average risk of breast cancer, the ECIBC's Guidelines Development Group (GDG) suggests not implementing mammography screening (conditional recommendation, moderate certainty in the evidence).

Recommendation strength

Conditional recommendation against the intervention*

2 aspects

Certainty of evidence

- Evidence assessed transparently across all certainty domains
- Confidence in an estimate?
- Starts with single research studies
- Ends with a body of evidence by outcome
  - High, moderate, low, very low certainty

Recommendations/Decisions

- Involves making judgments and decisions transparent
- Evidence to Decision (EtD) frameworks
- Comprehensive list of criteria that influence a decision or recommendation
- Clearly developed & formulated action message
  - Strong or conditional for or against an option

Certainty of evidence?

How confident in the research?

GRADE

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

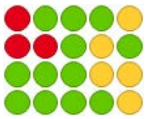
Determinants of certainty of evidence

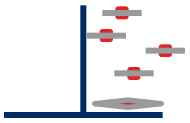
RCTs ⊕⊕⊕⊕ | high

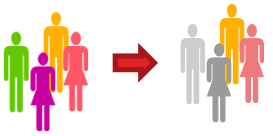
Non randomized studies ⊕⊕◯◯ | low

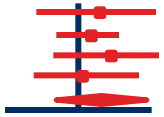
5 factors that can lower quality

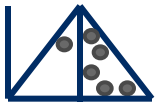
1. limitations in detailed study design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision
5. Publication bias





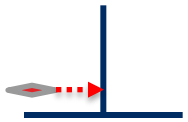


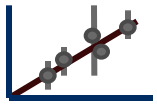




3 factors can increase quality

1. large magnitude of effect
2. opposing plausible residual bias or confounding
3. dose-response gradient





Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Large effects
High certainty

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

BMJ, 2003

3.2.2 Dramatic effect

If the course of a disease is certainly or almost certainly predictable, and no treatment options are available to influence this course, then proof of a benefit of a medical intervention can also be provided by the observation of a reversal of the (more or less) deterministic course of the disease in well-documented case series of patients. If, for example, it is known that it is highly probable that a disease leads to death within a short time after diagnosis, and it is

Large effects

3.2.2 Dramatic effect

Dramatic oversights (history and text)

If the course of a disease is certainly or almost certainly predictable, and no treatment options are available to influence this course, then proof of a benefit of a medical intervention can also be provided by the observation of a reversal of the (more or less) deterministic course of the disease in well-documented case series of patients. If, for example, it is known that it is highly probable that a disease leads to death within a short time after diagnosis, and it is

Final certainty – by outcome

GRADE

3.1.4 Outcome-related assessment

The benefit assessment and the estimation of the extent of the (un)certainty of results generally follow international EBM standards as developed, for example, by the GRADE[13] group [23].

For body of evidence from RCTs often low

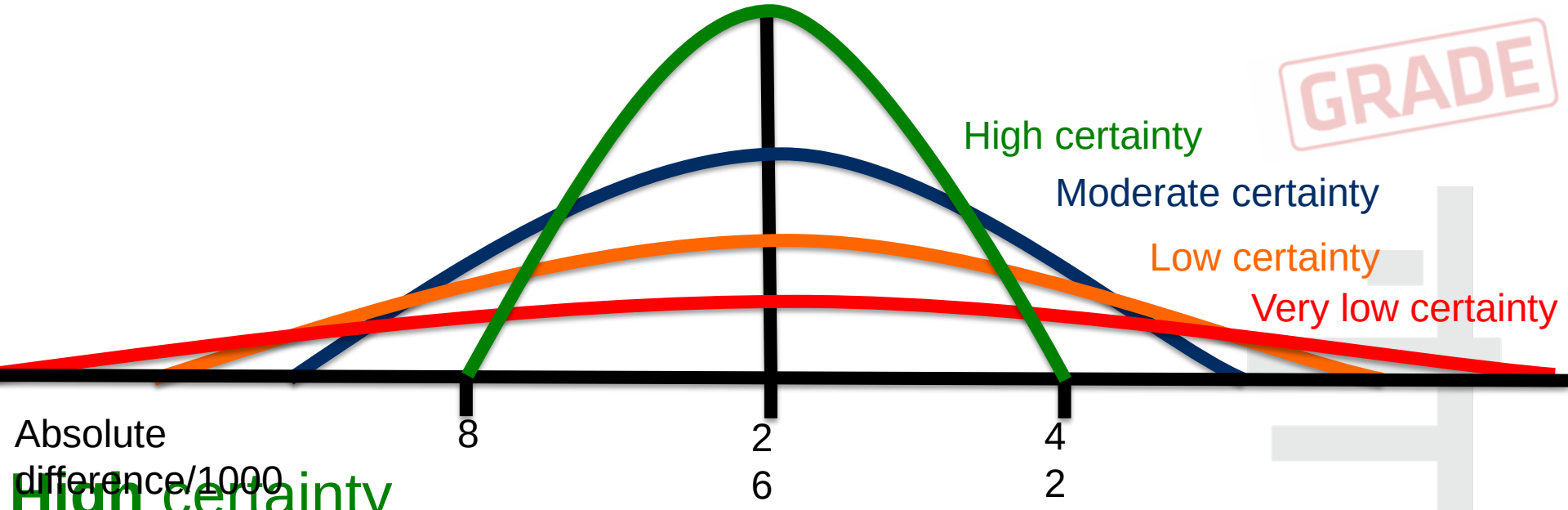
Non-randomized studies can end up as high but that is rare

Participants: *MDR TB patients*
Intervention: *bedaquiline + background MDR TB treatment*
Comparison: *background MDR TB treatment alone*

▸ About this summary

Add or remove columns: ⇕ Visual overview

Outcome	Plain language summary	Absolute Effect: Without bedaquiline	Absolute Effect: With bedaquiline	Relative effect (95% CI) Nº of participants & studies	Certainty of the evidence (GRADE)
Cured by end of study Follow-up: 120 weeks	*Bedaquiline may increase the number of patients cured.*	32 per 100	58 per 100	RR 1.81 (1.26 to 2.31) Based on data from 132 patients in 1 study	⊕⊕◯◯ Low
		Difference 26 more per 100 patients (95% CI: 8 to 42 more per 100 patients)			
Serious adverse events Follow-up: 24 week treatment phase	*It is uncertain whether bedaquiline increases the number of patients who have adverse effects.*	2 per 100	7 per 100	RR 3.6 (0.77 to 14.00) Based on data from 207 patients in 2 studies	⊕◯◯◯ Very low
		Difference 5 more per 100 patients (95% CI: 0 to 25 more per 100 patients)			
Mortality Follow-up: 120 weeks	*It is uncertain whether bedaquiline increases the number of patients who die.*	3 per 100	13 per 100	RR 9.23 (1.20 to 72.95) Based on data from 160 patients in 1 study	⊕◯◯◯ Very low
		Difference 10 more per 100 patients (95% CI: 0 to 53 more per 100 patients)			



High certainty

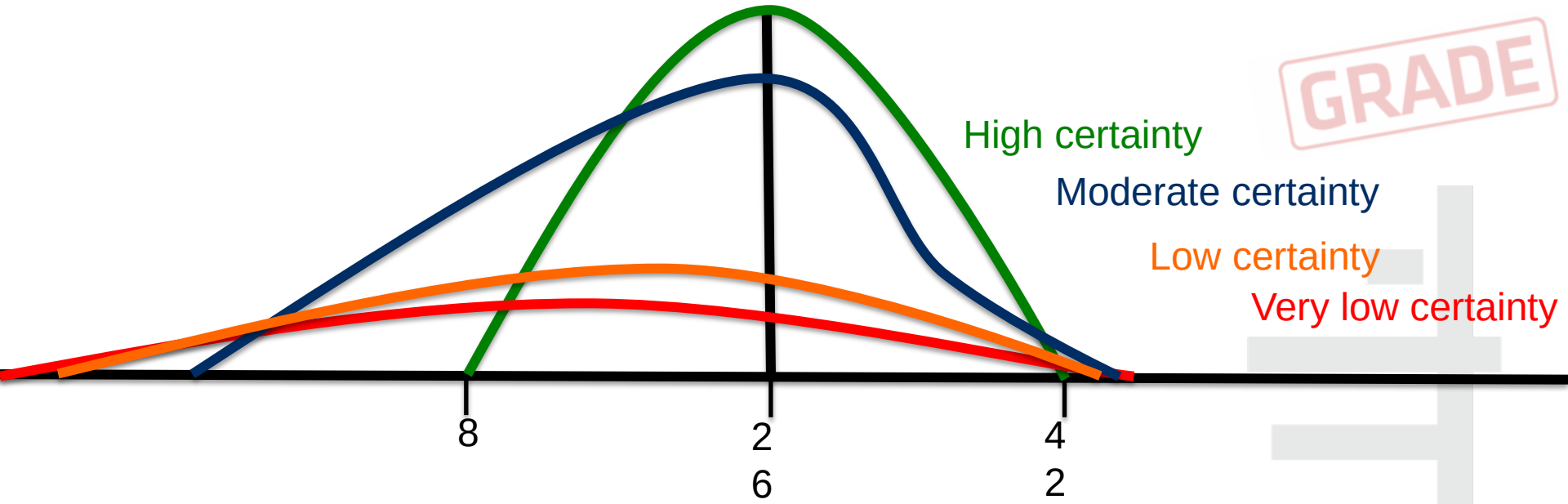
Certainty range identical to CI: distribution known

M**oderate** certainty due to indirectness or other downgrading domain including imprecision – wider certainty range **shape and width** not exactly known

Certainty due to risk of bias and indirectness – very wide certainty range despite narrow confidence intervals

Very low certainty due to risk of bias, indirectness and publication bias – extremely wide certainty range

Schünemann, JCE 2016



Is frequent reference to statistical significance appropriate?

GRADEpro GDT | ASH Guideline on Prevention of VTE in Medical Hospitalized Patients (Working Copy) | schuneh@mcmaster.ca

Should any DOAC vs. other prophylactic LMWH be used in acutely ill inpatient medical patients? | Bottom panel | Explanations | Help

Any DOAC compared to other prophylactic LMWH in acutely ill inpatient medical patients

Quality assessment							Summary of findings					
							№ of patients		Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Any DOAC	Other prophylactic LMWH	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Pulmonary Embolism – representing the moderate marker state (assessed with: Non fatal PE)												
2	randomised trials	not serious	not serious	not serious	serious	none	9/6190 (0.1%)	5/6264 (0.1%) [a] 0.4% [a]	**RR 1.75** (0.57 to 5.43)	**1 more per 1,000** (from 0 fewer to 4 more) **3 more per 1,000** (from 2 fewer to 18 more)	⨁⨁⨁◯ MODERATE	CRITICAL
Proximal Deep Vein Thrombosis – representing the moderate marker state (assessed with: Symptomatic DVT)												
2	randomised trials	not serious	not serious	not serious	serious	none	8/6193 (0.1%)	10/6266 (0.2%) [b,c] 0.2% [b,c]	**RR 0.74** (0.18 to 3.03)	**0 fewer per 1,000** (from 1 fewer to 3 more) **1 fewer per 1,000** (from 2 fewer to 4 more)	⨁⨁⨁◯ MODERATE	CRITICAL
Distal Deep Vein Thrombosis – representing the moderate distal DVT marker state (assessed with: Symptomatic DVTs)												

What's not so clear

Reliance on number of studies and statistical significance

- 4 or more studies:
  - All studies show statistically significant results in the same direction of effects: The effects in the same direction are clearly in the same direction.
  - The prediction interval does not cover the zero effect: The effects in the same direction are clearly in the same direction.
  - The prediction interval covers the zero effect: The effects in the same direction are moderately in the same direction.
- The study is a multi-centre study with at least 10 centres.
- The effect estimate observed has a very small corresponding p-value ($p < 0.001$).

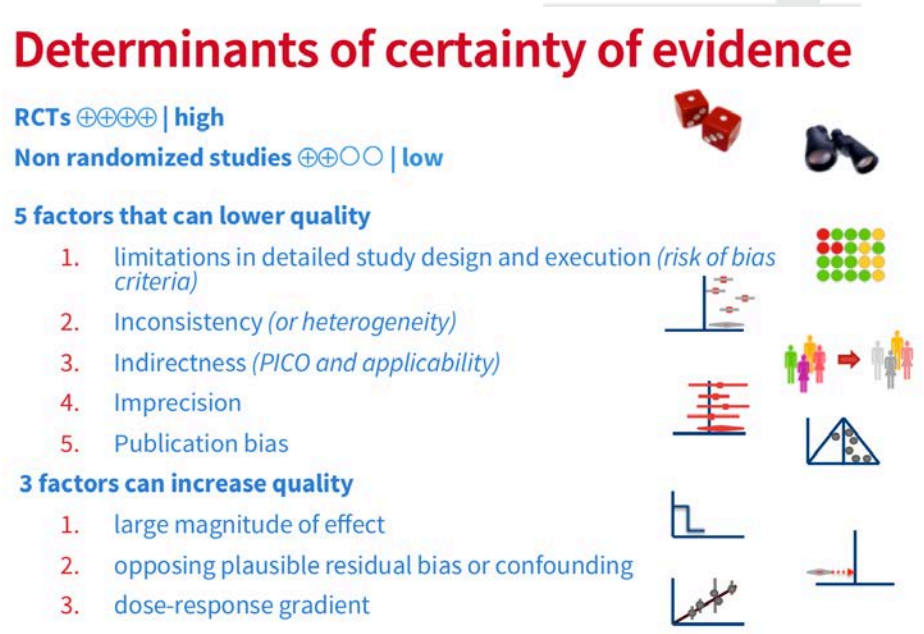
How does proof, indication, hint differ from certainty of the evidence?

What's not so clear

Question formulation – importance of outcomes

Often detailed but not practical – examples

A simple depiction of certainty evaluation



2 aspects

Certainty of evidence

- Evidence assessed transparently across all certainty domains
- Confidence in an estimate?
- Starts with single research studies
- Ends with a body of evidence by health outcome
  - High, moderate, low, very low certainty

Recommendations/Decisions

- Involves making judgments and decisions transparent
- Evidence to Decision (EtD) frameworks
- Comprehensive list of criteria that influence a decision or recommendation
- Clearly developed & formulated action message
  - Strong or conditional action for or against an option

OPEN ACCESS Freely available online

PLoS MEDICINE

Health in Action

Transparent Development of the WHO Rapid Advice Guidelines

Holger J. Schünemann*, Suzanne R. Hill, Meetali Kakad, Gunn E. Vist, Richard Bellamy, Lauren Stockman, Torbjørn Fosen Wisløff, Chris Del Mar, Frederick Hayden, Timothy M. Uyeki, Jeremy Farrar, Yazdan Yazdanpanah, Howard Zucker, John Beigel, Tawee Chotpitayasunondh, Tran Tinh Hien, Bülent Özbay, Norio Sugaya, Andrew D. Oxman

Summary

Emerging health problems require rapid advice. We describe the development and pilot testing of a systematic, transparent approach used by the World Health Organization (WHO) to develop rapid advice guidelines in response to requests from member states confronted with uncertainty about the pharmacological management of avian influenza A (H5N1) virus infection. We first searched for systematic reviews of randomized trials of treatment and prevention

Clinical practice guidelines generally, and some WHO guidelines specifically, have been criticized for not being based on the best available evidence, for being exposed to undue influence by industry and experts who participate in guideline panels, and for not adhering to guidelines for preparing guidelines [1–7]. Guidance that is not informed by the best available evidence or by statements that the available evidence is of low quality can harm patients, waste limited resources, and hinder research to address important uncertainties [8]. development of guidelines can take two years or more [13,14]. This timeframe is not practical for providing rapid advice, for example for emerging infectious diseases such as avian influenza (H5N1 infection) or severe acute respiratory syndrome (SARS). Indeed, one of the most frequently cited weaknesses in guideline development is the length of time that it takes to develop a guideline [15]. Organizations including the National Centre for Health and Clinical Excellence in the United Kingdom and the National Institutes of Health in the

Factors that can weaken the strength of a recommendation. Example: treatment of H5N1 patients with oseltamivir	Decision	Explanation
Lower quality evidence	☒ Yes ☐ No	The quality of evidence is very low.
Uncertainty about the balance of benefits versus harms and burdens	☒ Yes ☐ No	The benefits are uncertain because several important or critical outcomes were not measured.
Uncertainty or differences in values	☐ Yes ☒ No	All patients and care providers would accept treatment for H5N1 disease.
Marginal net benefits or downsides	☐ Yes ☒ No	The potential benefit is very large despite potentially small relative risk reductions.
Uncertainty about whether the net benefits are worth the costs	☐ Yes ☒ No	For treatment of sporadic patients the price is not too high.

Frequent "yes" answers will increase the likelihood of a weak recommendation.

doi:10.1371/journal.pmed.0040119.g003

Figure 3. Decisions about the Strength of a Recommendation

Schünemann et al. Lancet ID & PLOS Med 2007

Evidence to decision tables

Transparent for decision-making

Not granular enough for complex decision-making in health policy and public health

Feasibility and acceptability issues important

Different decisions need adaptable frameworks

- Coverage, health systems (perspectives), tests!

GRADE's DECIDE project (2011-2015)

- Improving EtD tables

Development

GRADE Evidence to Decision (EtD) Framework

An iterative 5-year process:

GRADE Working Group's approach to EtD

- NICE, SIGN, WHO partners

Review of relevant literature and surveys

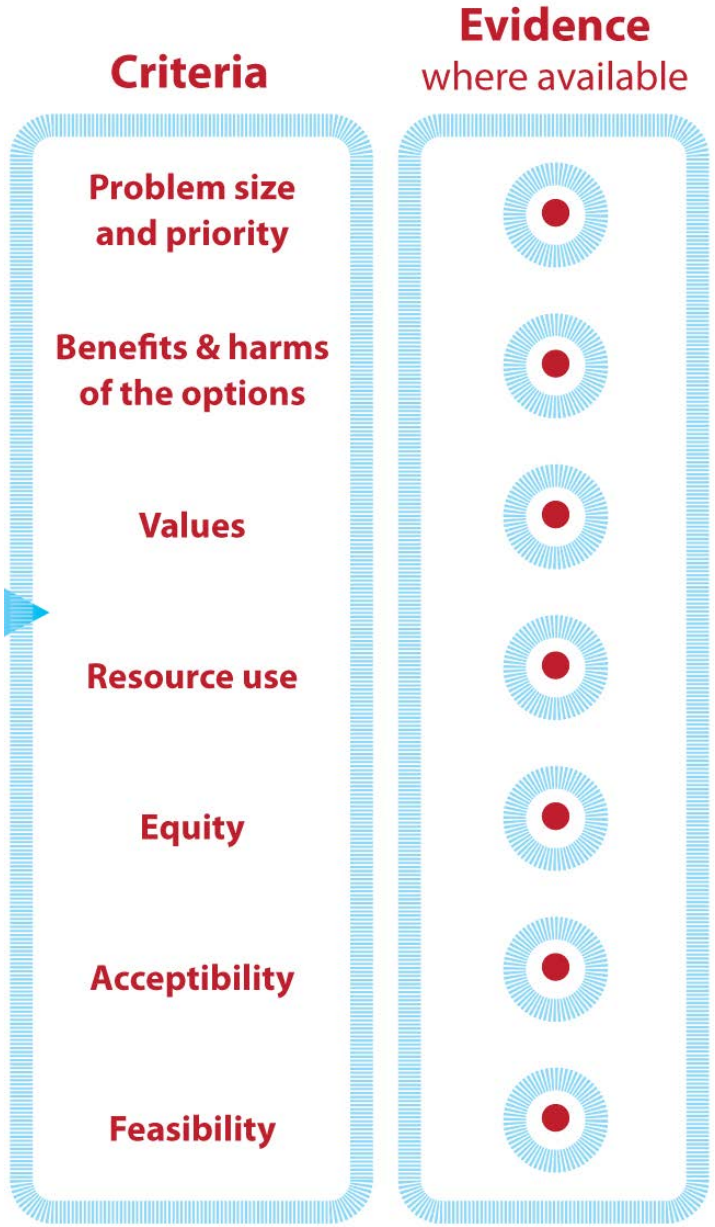
Brain storming

Feedback from stakeholders

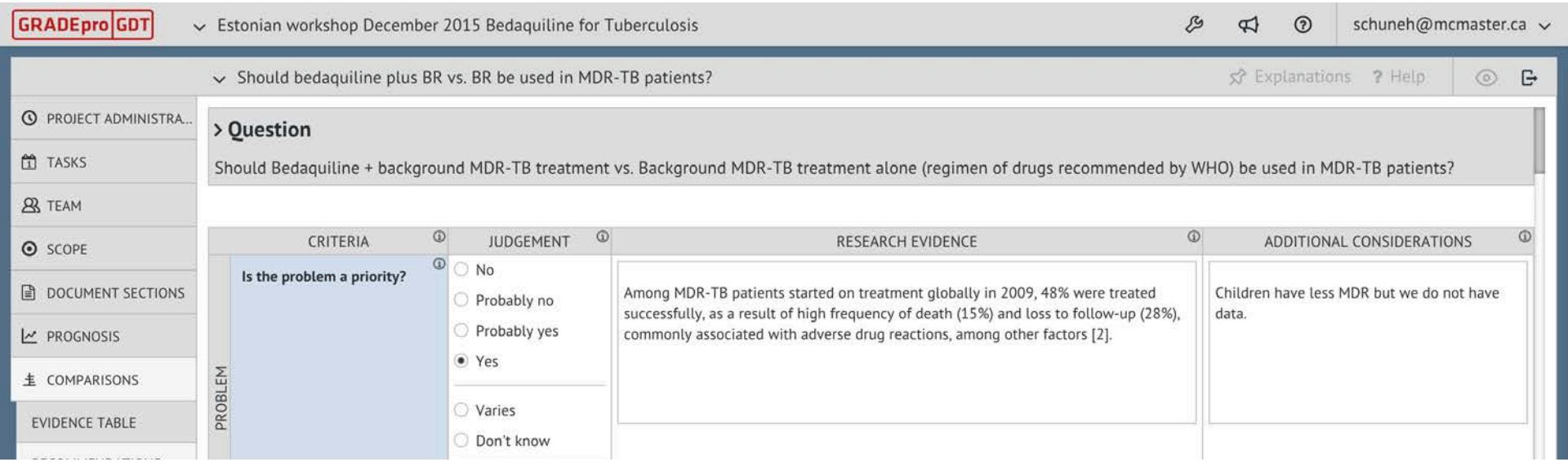
User testing

Application to examples (>100 recs) across health topics

Decision criteria



EtD frameworks



Criteria on which a decision is based

Judgements that must be made in relation to each criterion

Research evidence to inform each judgement

Additional considerations that inform or explain each judgement

The use of bedaquiline in the treatment of multidrug-resistant tuberculosis

Interim policy guidance

World Health Organization

WHO 2013

Contentious issue

FDA

Citizen groups

Pharma

Program managers

Overall low to very low certainty in the evidence

GRADE

Quality assessment							No of patients	Effect	Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations				
Subjects cured by end of study: 120 weeks (C208 Stage 2: mITT) [1,2]										
1	randomized	no serious	no serious	no serious	serious[5]	none	59/132	RR = 1.81 for	26/100 more patients cured	Critical
Mortality up to end of study at 120 weeks (C208 Sta										
1[1]	randomized	no serious	no serious	no serious	serious[5]	none	9/79[1]	1/81[1]	RR 9.23 (1.20	10 more per 100	+OOO	Critical

1 Mortality – SAE?

2 1 “phase 2” RCT evaluating cure

3 59 events in 132 patients 120 weeks

4 RR = 9.2 for death

5 10/100 more patients dead

1 The mITT modified intention to treat population i … group after excluding 13 subjects (16.5%) treated with bedaquiline and 15 subjects (18.5%) with placebo who did not have MDR or pre-XDR-TB at … uable.

2 Cure defined as 5 consecutive negative cultures from samples collected at least 30 days apart in the final 12 months of treatment , OR if only 1 culture is reported positive during that period, then a further 3 consecutive negative cultures from samples taken at least 30 days apart.

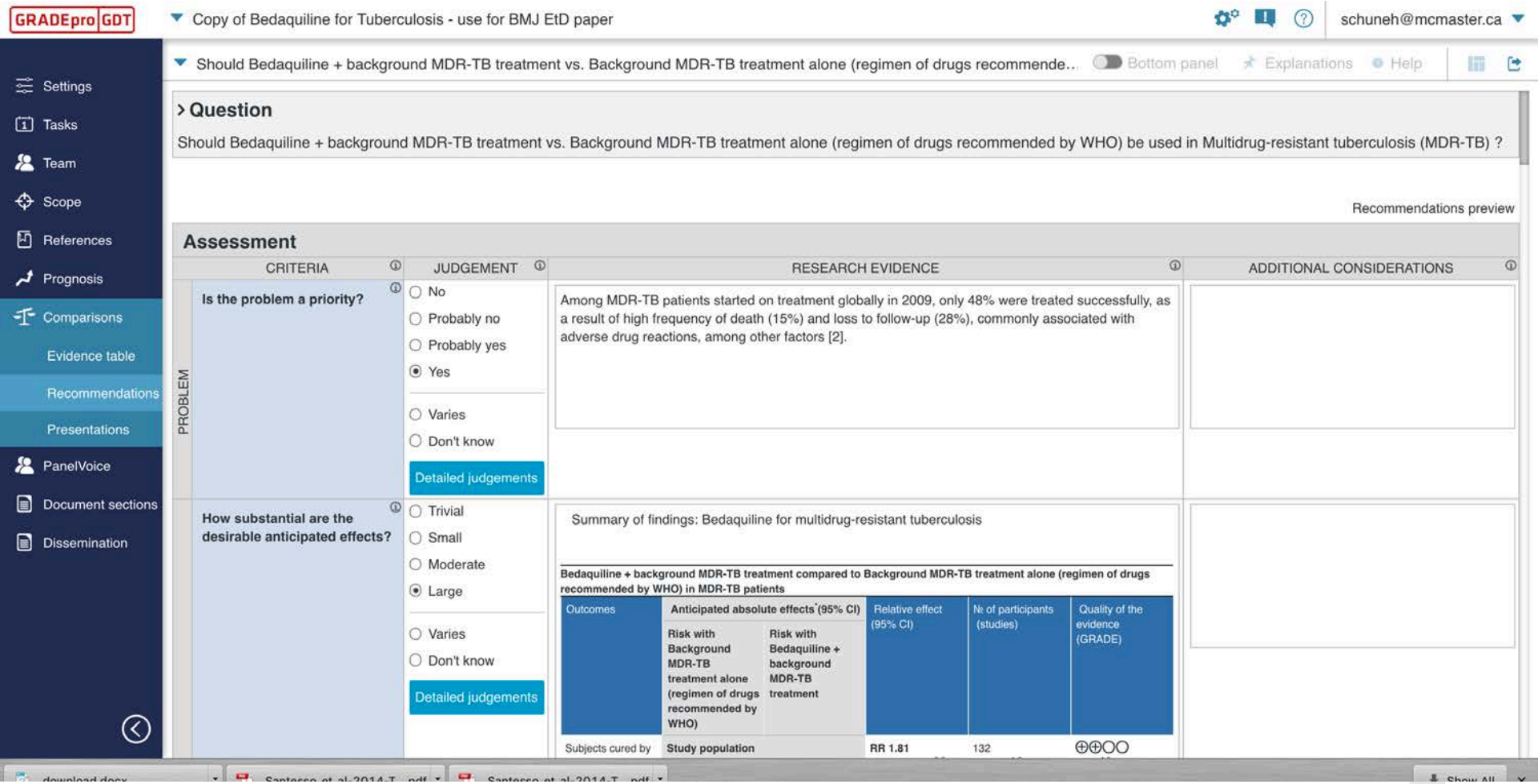
3 End of study data slide supplied by Janssen subsequent to US-FDA meeting. In this slide, mention is made of ‘treatment success’, but the company further clarified that the strict WHO definition of ‘cure’ was being used.

4 Representativeness of the mITT population (assumptions made for ITT population).

5 Small sample size and resulting large confidence interval limits precision: few (= serious) or very few (= very serious) observations.

6 This difference is statistically significant (Fisher p=0.005; Pearson p=0.003).

WHO, 2013


GRADEpro GDT
Copy of Bedaquiline for Tuberculosis - use for BMJ EtD paper
schuneh@mcmaster.ca
Settings
Tasks
Team
Scope
References
Prognosis
Comparisons
Evidence table
Recommendations
Presentations
PanelVoice
Document sections
Dissemination
Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommende..
Bottom panel
Explanations
Help
Question
Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommended by WHO) be used in Multidrug-resistant tuberculosis (MDR-TB) ?
Recommendations preview
Assessment
CRITERIA
JUDGEMENT
RESEARCH EVIDENCE
ADDITIONAL CONSIDERATIONS
PROBLEM
Is the problem a priority?
No
Probably no
Probably yes
Yes
Varies
Don't know
Detailed judgements
Among MDR-TB patients started on treatment globally in 2009, only 48% were treated successfully, as a result of high frequency of death (15%) and loss to follow-up (28%), commonly associated with adverse drug reactions, among other factors [2].
How substantial are the desirable anticipated effects?
Trivial
Small
Moderate
Large
Varies
Don't know
Detailed judgements
Summary of findings: Bedaquiline for multidrug-resistant tuberculosis
Bedaquiline + background MDR-TB treatment compared to Background MDR-TB treatment alone (regimen of drugs recommended by WHO) in MDR-TB patients
Outcomes
Anticipated absolute effects* (95% CI)
Risk with Background MDR-TB treatment alone (regimen of drugs recommended by WHO)
Risk with Bedaquiline + background MDR-TB treatment
Relative effect (95% CI)
№ of participants (studies)
Quality of the evidence (GRADE)
Subjects cured by
Study population
RR 1.81
132
⨁⨁◯◯

Settings
Tasks
Team
Scope
References
Prognosis
Comparisons
Evidence table
Recommendations
Presentations
PanelVoice
Document sections
Dissemination

Question

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommended by WHO) be used

Assessment

	CRITERIA	JUDGEMENT	RESEARCH EVIDENCE
PROBLEM	**Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Among MDR-TB patients started on treatment globally in 2009, only 48% were treated successfully, as a result of high frequency of death (15%) and loss to follow-up (28%), commonly associated with adverse drug reactions, among other factors [2].
	How substantial are the desirable anticipated effects?	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	Summary of findings: Bedaquiline for multidrug-resistant tuberculosis

Bedaquiline + background MDR-TB treatment compared to Background MDR-TB treatment alone (regimen of drugs recommended by WHO) in MDR-TB patients

Outcomes	Anticipated absolute effects*(95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)
	Risk with Background MDR-TB treatment alone (regimen of drugs recommended by WHO)	Risk with Bedaquiline + background MDR-TB treatment			

Participants: *MDR TB patients*
Intervention: *bedaquiline + background MDR TB treatment*
Comparison: *background MDR TB treatment alone*

About this summary

Add or remove columns:

Visual overview

Outcome	Plain language summary	Absolute Effect: Without bedaquiline	Absolute Effect: With bedaquiline	Relative effect (95% CI) Nº of participants & studies	Certainty of the evidence (GRADE)
Cured by end of study Follow-up: 120 weeks	*Bedaquiline may increase the number of patients cured.*	32 per 100	58 per 100	RR 1.81 (1.26 to 2.31) Based on data from 132 patients in 1 study	⨁⨁◯◯ Low
		Difference 26 more per 100 patients (95% CI: 8 to 42 more per 100 patients)			
Serious adverse events Follow-up: 24 week treatment phase	*It is uncertain whether bedaquiline increases the number of patients who have adverse effects.*	2 per 100	7 per 100	RR 3.6 (0.77 to 14.00) Based on data from 207 patients in 2 studies	⨁◯◯◯ Very low
		Difference 5 more per 100 patients (95% CI: 0 to 25 more per 100 patients)			
Mortality Follow-up: 120 weeks	*It is uncertain whether bedaquiline increases the number of patients who die.*	3 per 100	13 per 100	RR 9.23 (1.20 to 72.95) Based on data from 160 patients in 1 study	⨁◯◯◯ Very low
		Difference 10 more per 100 patients (95% CI: 0 to 53 more per 100 patients)			

Balance of health effects

26 more cures vs. 10 deaths?

26 more cures vs. 26 deaths?

Relative value of the health outcome

Values and preferences (utilities) = relative importance of outcomes

Even when we are certain the effects – what about the utilities

Article in Press

42 systematic reviews generated 23 items for assessing the risk of bias in Values and Preferences´ studies

Juan Jose Yepes Nuñez, Yuan Zhang, Feng Xie, Pablo Alonso-Coello, Anna Selva, Holger Schünemann, Gordon Guyatt

DOI: http://dx.doi.org/10.1016/j.jclinepi.2017.04.019

Article Info

Abstract

Abstract

Objective

In systematic reviews of studies of patients' values and preferences, to summarize items and domains authors have identified when considering the risk of bias associated with primary studies.

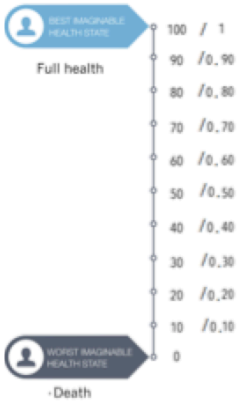
Applying GRADE domains to utility/importance of outcomes

GRADE

Summary of finding table

Question: What are the views about the relative value/importance of outcomes of interest in decision making for patients with chronic obstructive pulmonary disease?

Health state/Outcome (Categories of values and preferences)	Estimates of outcome importance (range across studies / pooled mean, 95% CI)	No. of participants /studies	Certainty in evidence	Interpretation of findings
Exacerbation (Utility* measured with visual analogue scale [1])	range across studies: 0.259-0.466/ pooled mean: 0.377 (95% CI: 0.294, 0.461) [2]	1076 participants/ 4 studies [2]	⊕⊕⊕ Moderate certainty due to inconsistency [2]	Most people find exacerbation of COPD probably has a large impact on lives. There is likely no important variability for this assessment.
Exacerbation (EQ-5D Utility [3])	range across studies 0.43-0.683/ pooled mean: 0.525 (95% CI: 0.434, 0.615) [4]	927 participants/ 3 studies [4]	⊕⊕ Low certainty due to inconsistency and indirectness [4,5]	Most people find exacerbation of COPD probably has a large impact on lives. There is likely no important variability for this assessment.
Exacerbation (disutility) [6]	Visual analogue scale: One non-serious exacerbation: -0.037 (0.005); Two non-serious exacerbations: -0.068 (0.005); One serious exacerbation: -0.090 (0.007); One non-serious and one serious exacerbation: -0.130 (0.007) Time trade off: One non-serious exacerbation: -0.010 (0.007); Two non-serious exacerbations: -0.021 (0.007); One serious exacerbation: -0.042 (0.009); One non-serious and one serious exacerbation: -0.088 (0.009)	239 participants/ 1 study	⊕⊕⊕⊕ High certainty	Most people find exacerbation of COPD has an impact on lives, which grows larger as the severity of exacerbation progresses. There is likely no important variability for this assessment.



*Utilities represent the strength of an individual's preferences for different outcomes. They are measured on an interval scale, with zero reflecting states of health equivalent to death/worst imaginable health and one (or 100 in some cases) reflecting perfect health/ best imaginable health.

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommended by WHO) be a...

DESIRABLE EFFECTS

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

Detailed judgements

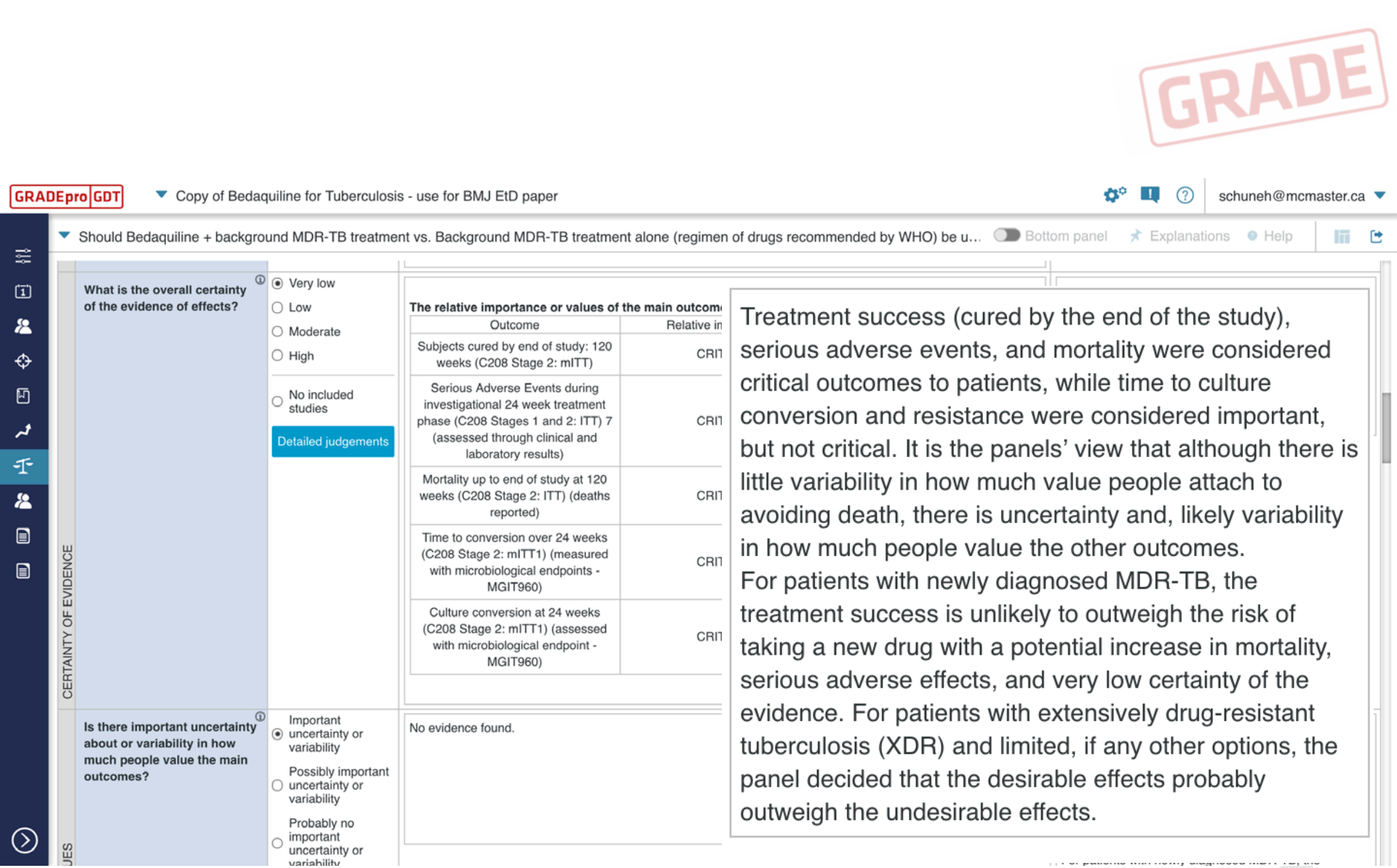
How substantial are the undesirable anticipated effects?

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

Detailed judgements

Bedaquiline + background MDR-TB treatment compared to Background MDR-TB treatment alone (regimen of drugs recommende WHO) in MDR-TB patients

Outcomes	Anticipated absolute effects*(95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)
	Risk with Background MDR-TB treatment alone (regimen of drugs recommended by WHO)	Risk with Bedaquiline + background MDR-TB treatment			
Subjects cured by end of study: 120 weeks (C208 Stage 2: mITT)[1,2]	**Study population**		**RR 1.81** (1.26 to 2.31)[3,6]	132 (1 RCT)[1,3]	⨁⨁◯◯ LOW[4,5]
	32 per 100[1]	**58 per 100** (40 to 74)[1]			
Serious Adverse Events during investigational 24 week treatment phase (C208 Stages 1 and 2: ITT) 7 (assessed through clinical and laboratory results)	**Study population**		**RR 3.60** (0.77 to 14.00)	207 (2 RCTs)[7,9]	⨁◯◯◯ VERY LOW[5,8]
	2 per 100	**7 per 100** (1 to 27)[9]			
Mortality up to end of study at 120 weeks (C208 Stage 2: ITT) (deaths reported)	**Study population**		**RR 9.23** (1.20 to 72.95)[12,13]	160 (1 RCT)[10]	⨁◯◯◯ VERY LOW[3,11]
	1 per 100[10]	**11 per 100** (1 to 90)[10]			
Time to conversion over 24 weeks (C208 Stage 2: mITT1) (measured with microbiological endpoints - MGIT960)	**Study population**		not estimable	(1 RCT)[14]	⨁⨁◯◯ LOW[4,5,15]
	0 per 100	**NaN per 100** (NaN to NaN)			
Culture conversion at 24 weeks (C208 Stage 2: mITT1) (assessed with microbiological endpoint - MGIT960)	**Study population**		**RR 1.37** (1.10 to 1.77)[17]	132 (1 RCT)[1,16]	⨁⨁◯◯ LOW[4,5,15]
	58 per 100[1]	**79 per 100** (63 to 100)[1]			
Acquired resistance to	**Study population**		**RR 0.39**	37	⨁◯◯◯


GRADE
GRADEpro GDT
Copy of Bedaquiline for Tuberculosis - use for BMJ EtD paper
schuneh@mcmaster.ca
Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommended by WHO) be u...
Bottom panel
Explanations
Help
CERTAINTY OF EVIDENCE
What is the overall certainty of the evidence of effects?
Very low
Low
Moderate
High
No included studies
Detailed judgements
The relative importance or values of the main outcom
Outcome
Relative in
Subjects cured by end of study: 120 weeks (C208 Stage 2: mITT)
CRIT
Serious Adverse Events during investigational 24 week treatment phase (C208 Stages 1 and 2: ITT) 7 (assessed through clinical and laboratory results)
CRIT
Mortality up to end of study at 120 weeks (C208 Stage 2: ITT) (deaths reported)
CRIT
Time to conversion over 24 weeks (C208 Stage 2: mITT1) (measured with microbiological endpoints - MGIT960)
CRIT
Culture conversion at 24 weeks (C208 Stage 2: mITT1) (assessed with microbiological endpoint - MGIT960)
CRIT
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 evidence found.
Treatment success (cured by the end of the study), serious adverse events, and mortality were considered critical outcomes to patients, while time to culture conversion and resistance were considered important, but not critical. It is the panels' view that although there is little variability in how much value people attach to avoiding death, there is uncertainty and, likely variability in how much people value the other outcomes.
For patients with newly diagnosed MDR-TB, the treatment success is unlikely to outweigh the risk of taking a new drug with a potential increase in mortality, serious adverse effects, and very low certainty of the evidence. For patients with extensively drug-resistant tuberculosis (XDR) and limited, if any other options, the panel decided that the desirable effects probably outweigh the undesirable effects.

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs reco

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

Detailed judgements

See evidence profile above

URCES REQUIRED

How large are the resource requirements (costs)?

- ○ Large costs
- ○ Moderate costs
- ○ Negligible costs and savings
- ○ Moderate savings
- ○ Large savings

- ◉ Varies
- ○ Don't know

Cost data for the base case in each country were sourced from published s
data provided by study authors. For the primary estimates for the unit cost p
regimen cost of US $900 (for Global Fund Eligible countries) and US $3000
full course of bedaquiline based on estimates from Janssen. In addition the
added.
To estimate the possible cost savings from a shortened course with bedaqu
six months were estimated. Eight month intensive phase drug costs were a
hospitalization and required length of second-line parenteral agents (injecta
hospitalization was not used extensively in the intensive phase of treatment
in the cost of clinic visits. All other costs (programme management, testing
to remain the same as the non-shortened bedaquiline regimen.

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommende

COST EFFECTIVENESS

Does the cost-effectiveness of the intervention 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
- ○ No included studies

Detailed judgements

Modelling of the incremental cost-effectiveness of adding bedaquiline to WHO recom was conducted by an independent consultant contracted by WHO for review by the e assumed that bedaquiline would be added to treatment for all patients starting MDR- were explored to appraise the cost-effectiveness of bedaquiline in these settings. Un bedaquiline-containing regimens were assessed as relatively cost-effective in most s ambiguous in low-income settings, and highly dependent on the assumptions made a results to routine settings.

EQUITY

What would be the impact on health equity?

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

Detailed judgements

No research evidence found

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs re

ACCEPTABILITY	**Is the intervention acceptable to key stakeholders?**	○ No ○ Probably no ○ Probably yes ○ Yes ◉ Varies ○ Don't know Detailed judgements	No evidence found.
FEASIBILITY	**Is the intervention feasible to implement?**	○ No ○ Probably no ○ Probably yes ○ Yes ◉ Varies ○ Don't know Detailed judgements	No evidence found.

BALANCE OF EFFECTS	Favors the comparison	Probably favors the...	Does not favor either the...	**Probably favors the...**	Favors the intervention	Varies	Don't know	
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and...	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...	Does not favor either the...	**Probably favors the...**	Favors the intervention	Varies	No...	
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		

Conclusions

Should Bedaquiline + background MDR-TB treatment vs. Background MDR-TB treatment alone (regimen of drugs recommended by WHO) be used in Multidrug-resistant tuberculosis (MDR-TB) ?

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
	○	○	○	◉	○

6. WHO Interim policy recommendations

In view of the aforementioned evidence assessment and advice provided by the EG, WHO recommends that ***bedaquiline may be added to a WHO-recommended regimen in adult patients with pulmonary MDR-TB (conditional recommendation, very low confidence in estimates of effects).***

Given the limited data available on bedaquiline and its use under the various situations that may be encountered in different clinical settings, adequate provisions for safe and effective use of the drug must be in place. Consequently, countries are advised to follow a phased approach to bedaquiline implementation, ideally through observational

5. **Pharmacovigilance and proper management of adverse drug reactions and prevention of drug–drug interactions.**
    a. Special measures need to be put in place to ensure the early detection and timely reporting of adverse events using active pharmacovigilance methods, such as 'cohort event monitoring'. Any adverse drug reaction attributed to bedaquiline should also be reported to the national pharmacovigilance centre as part of the spontaneous reporting mechanism in the country. As for any other drug in the MDR-TB regimen the patient should be encouraged to report to the attending health worker any adverse event that occurs during the time the drug is being

Report of the Guideline Development Group Meeting on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis

A review of available evidence (2016)

28 - 29 June 2016

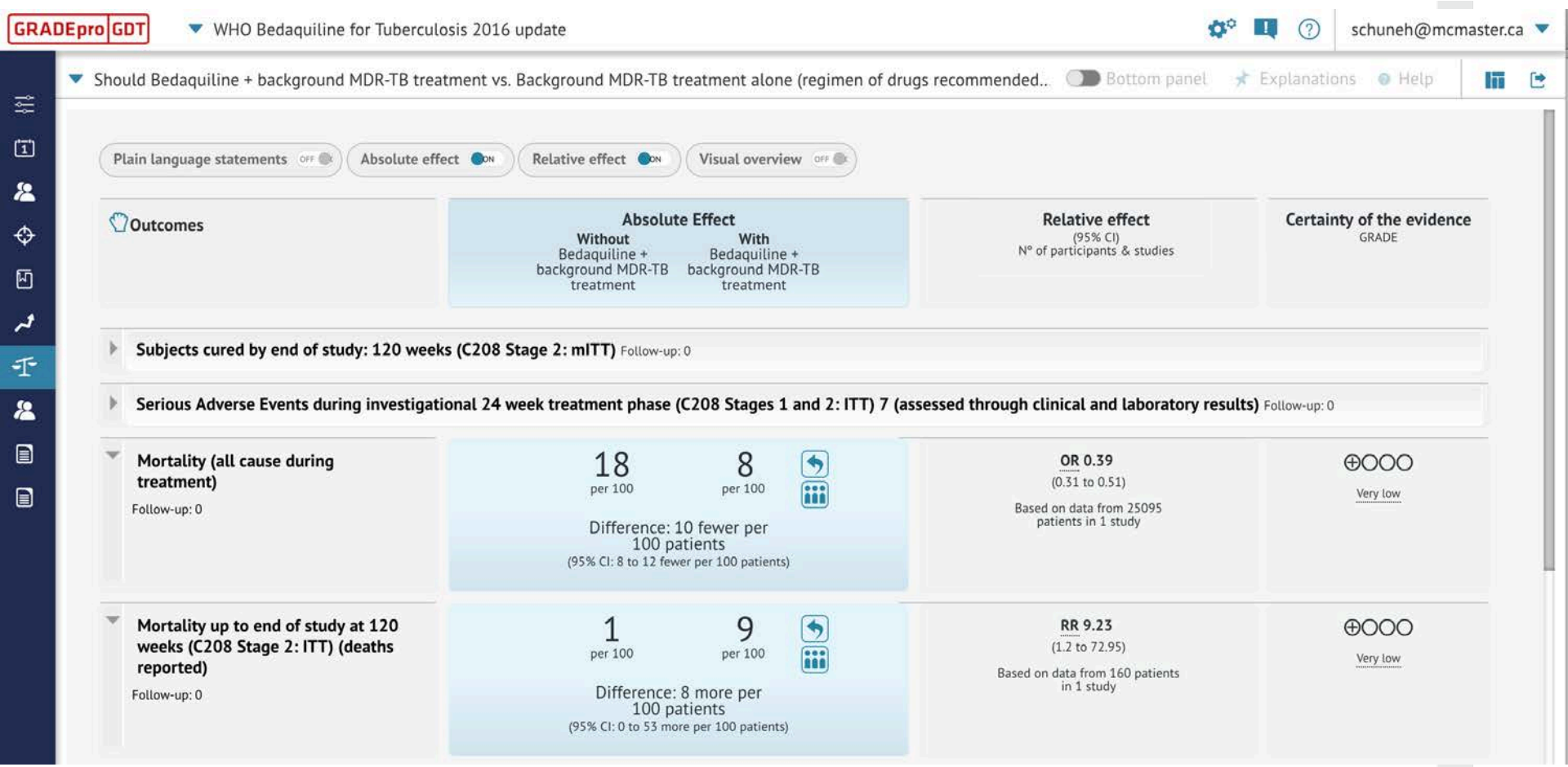
Geneva, Switzerland

World Health Organization

THE END TB STRATEGY

GRADE

Individual patient data meta-analysis of bedaquiline in MDR-TB



Modelling: benefits > harm?

GRADE

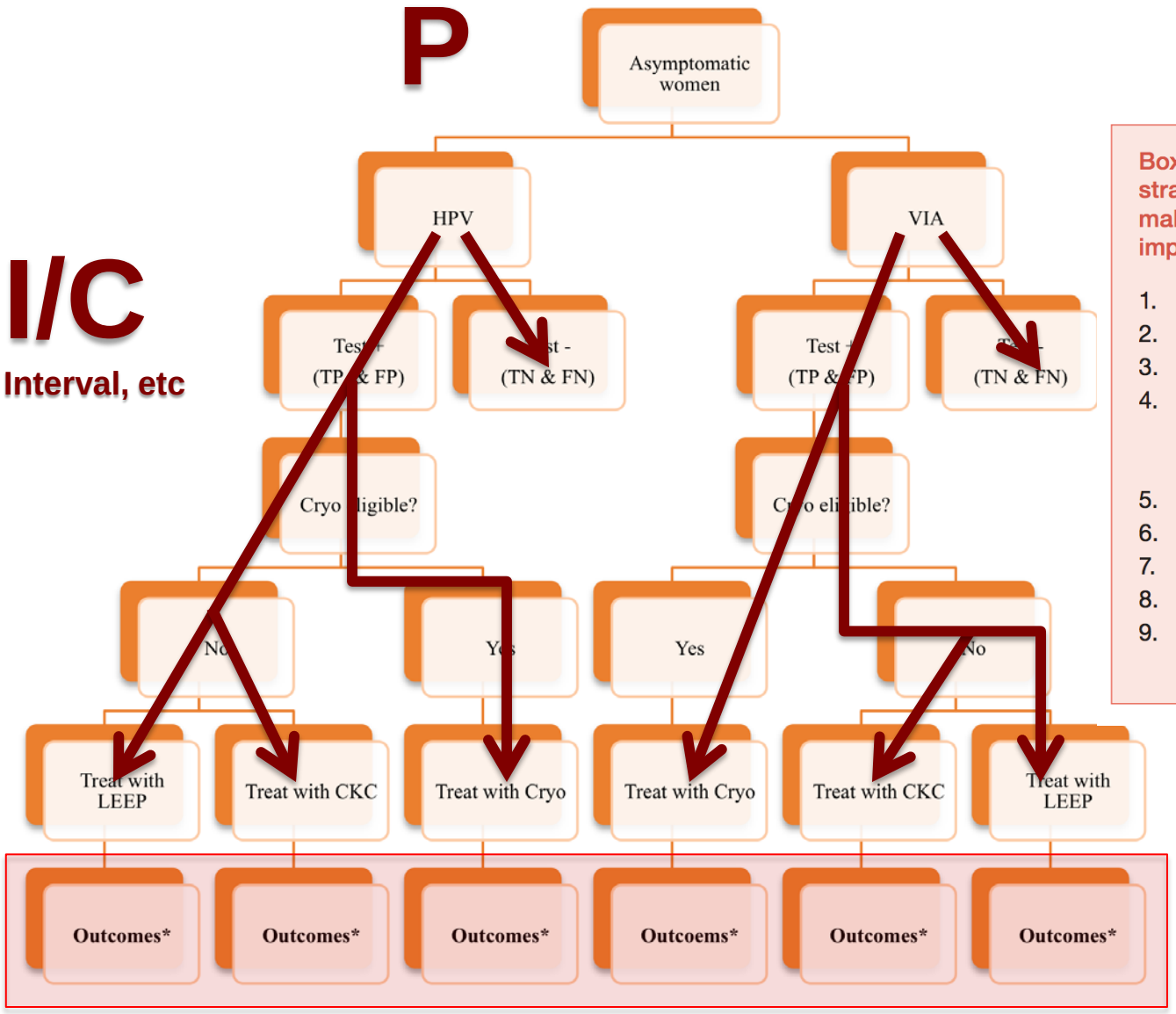
Participants: *MDR TB patients*
Intervention: *bedaquiline + background MDR TB treatment*
Comparison: *background MDR TB treatment alone*

About this summary

Add or remove columns:

Visual overview

Outcome	Plain language summary	Absolute Effect: Without bedaquiline	Absolute Effect: With bedaquiline	Relative effect (95% CI) N° of participants & studies	Certainty of the evidence (GRADE)
Cured by end of study Follow-up: 120 weeks	*Bedaquiline may increase the number of patients cured.*	32 per 100	58 per 100	RR 1.81 (1.26 to 2.31) Based on data from 132 patients in 1 study	⨁⨁◯◯ Low
		Difference 26 more per 100 patients (95% CI: 8 to 42 more per 100 patients)			
Serious adverse events Follow-up: 24 week treatment phase	*It is uncertain whether bedaquiline increases the number of patients who have adverse effects.*	2 per 100	7 per 100	RR 3.6 (0.77 to 14.00) Based on data from 207 patients in 2 studies	⨁◯◯◯ Very low
		Difference 5 more per 100 patients (95% CI: 0 to 25 more per 100 patients)			
Mortality Follow-up: 120 weeks	*It is uncertain whether bedaquiline increases the number of patients who die.*	3 per 100	13 per 100	RR 9.23 (1.20 to 72.95) Based on data from 160 patients in 1 study	⨁◯◯◯ Very low
		Difference 10 more per 100 patients (95% CI: 0 to 53 more per 100 patients)			



Box 2: Outcomes for screen-and-treat strategies identified as important for making recommendations (in order of importance)

1. Mortality from cervical cancer
2. Cervical cancer incidence
3. Detected CIN2, CIN3
4. Major infections (requiring hospital admission and antibiotics, e.g. pelvic inflammatory disease)
5. Maternal bleeding
6. Premature delivery
7. Fertility
8. Identification of STIs (benefit)
9. Minor infections (requiring outpatient treatment only)

*Outcomes are: Mortality from cervical cancer, Rate of cervical cancer detection, Rate of CIN 2 & 3 detection, major bleeding, premature delivery, infertility, STI detection, major infections, and minor infections

Certainty in the entire model?

Own or single model

vs

Evidence across models

GRADE Domains

Domains of modeling requiring evaluation
Structure
Input
Calculation/computations
Process

Domains of modeling requiring evaluation	What is being evaluated (produced)
Structure	PICO analytical framework - Graphical representation Description of model characteristics (e.g. annual vs biannual screening) – part of EtD Assumptions (based on evidence)
Input	Assumptions (based on evidence) Certainty of the evidence summarized in evidence profiles for: • Prognostic information • Test accuracy • Effects of interventions (as part of the pathways described) • Link(ed), indirect evidence • Resources • Values and preferences
Calculation/computations	Summary of findings/evidence profiles Evidence to Decision Frameworks
Process	Involvement of (appropriate) members at relevant stages Sign off on PICO analytical framework Agreement with input variables COI management Documentation Evidence to Decision Frameworks Certainty of the evidence for the decision (GRADE)

Summary

10 years out IQWiG follows or exceeds international standards

Evidence assessment remains complex

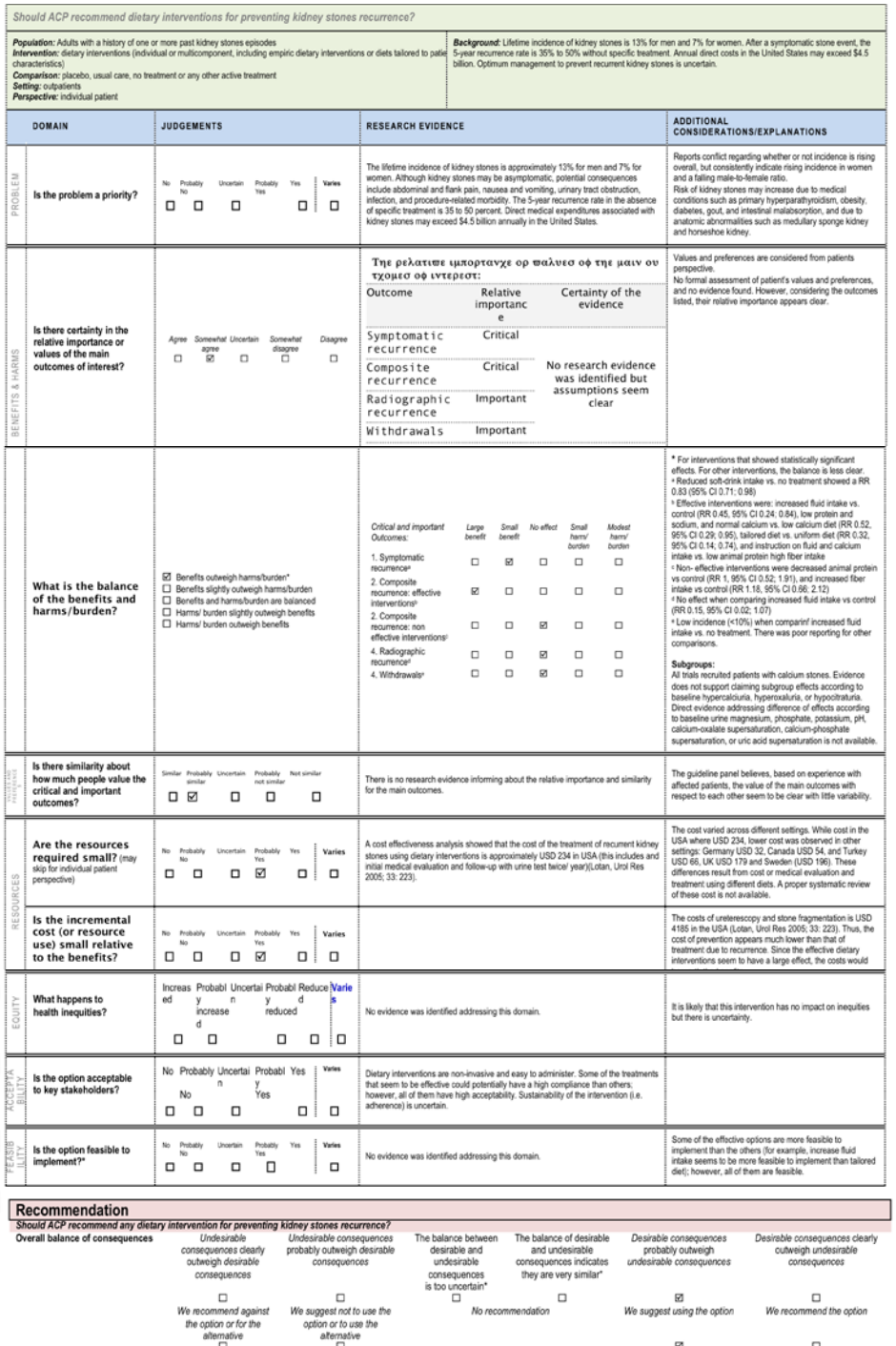
Certainty in utility evidence

Certainty in models that determine decisions – where it all comes together

Not discussed: Tests, NMA prognostic evidence, qualitative evidence …
GRADE not stopping

GRADE

- Question/Problem
- Benefits and harms
- Quality of evidence
- Values
- Resources
- Equity
- Acceptability
- Feasibility
- Recommendation



Should ACP recommend dietary interventions for preventing kidney stones recurrence?

Population: Adults with a history of one or more past kidney stones episodes
Intervention: dietary interventions (individual or multicomponent, including empiric dietary interventions or diets tailored to patie characteristics)
Comparison: placebo, usual care, no treatment or any other active treatment
Setting: outpatients
Perspective: individual patient

Background: Lifetime incidence of kidney stones is 13% for men and 7% for women. After a symptomatic stone event, the 5-year recurrence rate is 35% to 50% without specific treatment. Annual direct costs in the United States may exceed $4.5 billion. Optimum management to prevent recurrent kidney stones is uncertain.

	DOMAIN	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS/EXPLANATIONS
PROBLEM	**Is the problem a priority?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	The lifetime incidence of kidney stones is approximately 13% for men and 7% for women. Although kidney stones may be asymptomatic, potential consequences include abdominal and flank pain, nausea and vomiting, urinary tract obstruction, infection, and procedure-related morbidity. The 5-year recurrence rate in the absence of specific treatment is 35 to 50 percent. Direct medical expenditures associated with kidney stones may exceed $4.5 billion annually in the United States.	Reports conflict regarding whether or not incidence is rising overall, but consistently indicate rising incidence in women and a falling male-to-female ratio. Risk of kidney stones may increase due to medical conditions such as primary hyperparathyroidism, obesity, diabetes, gout, and intestinal malabsorption, and due to anatomic abnormalities such as medullary sponge kidney and horseshoe kidney.
BENEFITS & HARMS	**Is there certainty in the relative importance or values of the main outcomes of interest?**	Agree ☐ · Somewhat agree ☑ · Uncertain ☐ · Somewhat disagree ☐ · Disagree ☐	Τηε ρελατιωε ιμπορτανχε ορ ωαλυεσ οφ τηε μαιν ουτχομεσ οφ ιντερεστ: Outcome — Relative importance — Certainty of the evidence Symptomatic recurrence — Critical Composite recurrence — Critical Radiographic recurrence — Important Withdrawals — Important No research evidence was identified but assumptions seem clear	Values and preferences are considered from patients perspective. No formal assessment of patient's values and preferences, and no evidence found. However, considering the outcomes listed, their relative importance appears clear.
BENEFITS & HARMS	**What is the balance of the benefits and harms /burden?**	☑ Benefits outweigh harms/burden* ☐ Benefits slightly outweigh harms/burden ☐ Benefits and harms/burden are balanced ☐ Harms/ burden slightly outweigh benefits ☐ Harms/ burden outweigh benefits	*Critical and important Outcomes:* — Large benefit — Small benefit — No effect — Small harm/ burden — Modest harm/ burden 1. Symptomatic recurrence[a]: ☐ ☑ ☐ ☐ ☐ 2. Composite recurrence: effective interventions[b]: ☑ ☐ ☐ ☐ ☐ 2. Composite recurrence: non effective interventions[c]: ☐ ☐ ☑ ☐ ☐ 4. Radiographic recurrence[d]: ☐ ☐ ☑ ☐ ☐ 4. Withdrawals[e]: ☐ ☐ ☑ ☐ ☐	* For interventions that showed statistically significant effects. For other interventions, the balance is less clear. [a] Reduced soft-drink intake vs. no treatment showed a RR 0.83 (95% CI 0.71; 0.98) [b] Effective interventions were: increased fluid intake vs. control (RR 0.45, 95% CI 0.24; 0.84), low protein and sodium, and normal calcium vs. low calcium diet (RR 0.52, 95% CI 0.29; 0.95), tailored diet vs. uniform diet (RR 0.32, 95% CI 0.14; 0.74), and instruction on fluid and calcium intake vs. low animal protein high fiber intake [c] Non- effective interventions were decreased animal protein vs control (RR 1, 95% CI 0.52; 1.91), and increased fiber intake vs control (RR 1.18, 95% CI 0.66; 2.12) [d] No effect when comparing increased fluid intake vs control (RR 0.15, 95% CI 0.02; 1.07) [e] Low incidence (<10%) when comparinf increased fluid intake vs. no treatment. There was poor reporting for other comparisons. **Subgroups:** All trials recruited patients with calcium stones. Evidence does not support claiming subgroup effects according to baseline hypercalciuria, hyperoxaluria, or hypocitraturia. Direct evidence addressing difference of effects according to baseline urine magnesium, phosphate, potassium, pH, calcium-oxalate supersaturation, calcium-phosphate supersaturation, or uric acid supersaturation is not available.
VALUES	**Is there similarity about how much people value the critical and important outcomes?**	Similar ☐ · Probably similar ☑ · Uncertain ☐ · Probably not similar ☐ · Not similar ☐	There is no research evidence informing about the relative importance and similarity for the main outcomes.	The guideline panel believes, based on experience with affected patients, the value of the main outcomes with respect to each other seem to be clear with little variability.
RESOURCES	**Are the resources required small?** (may skip for individual patient perspective)	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☑ · Yes ☐ · Varies ☐	A cost effectiveness analysis showed that the cost of the treatment of recurrent kidney stones using dietary interventions is approximately USD 234 in USA (this includes and initial medical evaluation and follow-up with urine test twice/ year)(Lotan, Urol Res 2005; 33: 223).	The cost varied across different settings. While cost in the USA where USD 234, lower cost was observed in other settings: Germany USD 32, Canada USD 54, and Turkey USD 66, UK USD 179 and Sweden (USD 196). These differences result from cost or medical evaluation and treatment using different diets. A proper systematic review of these cost is not available.
RESOURCES	**Is the incremental cost (or resource use) small relative to the benefits?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☑ · Yes ☐ · Varies ☐		The costs of ureterescopy and stone fragmentation is USD 4185 in the USA (Lotan, Urol Res 2005; 33: 223). Thus, the cost of prevention appears much lower than that of treatment due to recurrence. Since the effective dietary interventions seem to have a large effect, the costs would
EQUITY	**What happens to health inequities?**	Increased ☐ · Probably increased ☐ · Uncertain ☐ · Probably reduced ☐ · Reduced ☐ · Varies ☐	No evidence was identified addressing this domain.	It is likely that this intervention has no impact on inequities but there is uncertainty.
ACCEPTABILITY	**Is the option acceptable to key stakeholders?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	Dietary interventions are non-invasive and easy to administer. Some of the treatments that seem to be effective could potentially have a high compliance than others; however, all of them have high acceptability. Sustainability of the intervention (i.e. adherence) is uncertain.	
FEASIBILITY	**Is the option feasible to implement?***	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	No evidence was identified addressing this domain.	Some of the effective options are more feasible to implement than the others (for example, increase fluid intake seems to be more feasible to implement than tailored diet); however, all of them are feasible.

Recommendation

Should ACP recommend any dietary intervention for preventing kidney stones recurrence?

Overall balance of consequences

Undesirable consequences clearly outweigh desirable consequences	*Undesirable consequences probably outweigh desirable consequences*	The balance between desirable and undesirable consequences is too uncertain*	The balance of desirable and undesirable consequences indicates they are very similar*	*Desirable consequences probably outweigh undesirable consequences*	*Desirable consequences clearly outweigh undesirable consequences*
☐	☐	☐	☐	☑	☐
We recommend against the option or for the alternative	*We suggest not to use the option or to use the alternative*	*No recommendation*		*We suggest using the option*	*We recommend the option*

Evidence to decision

GRADE DECIDE

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

For groups making recommendations

Question

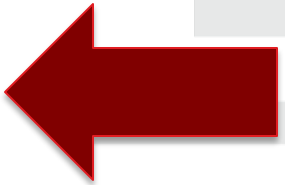
- Details
- Subgroups
- Background

Assessment

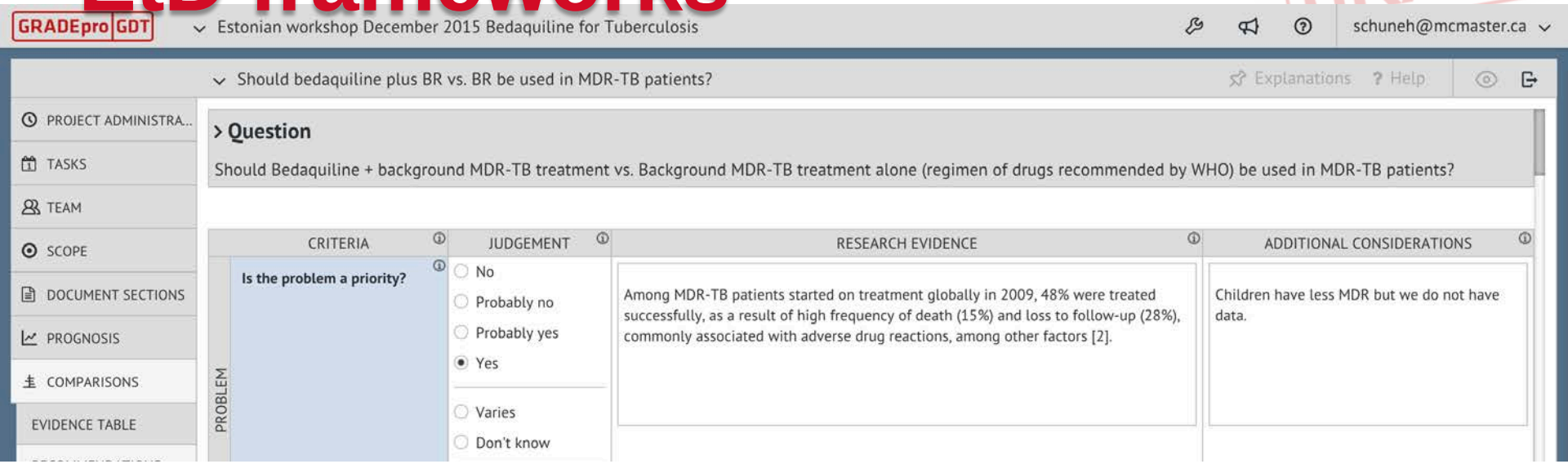
- Criteria
- Judgements
- Research evidence
- Additional considerations

Conclusions

- Type of recommendation
- Recommendation
- Justification
- Implementation considerations
- Monitoring and evaluation
- Research considerations



EtD frameworks



Criteria on which a recommendation is based

Judgements that must be made in relation to each criterion

Research evidence to inform each judgement

GRADE Evidence to Decision (EtD) framework

Can help guideline panels (and decision makers) move from evidence to a recommendation or decision by

Informing judgements about the pros and cons of each option (intervention)

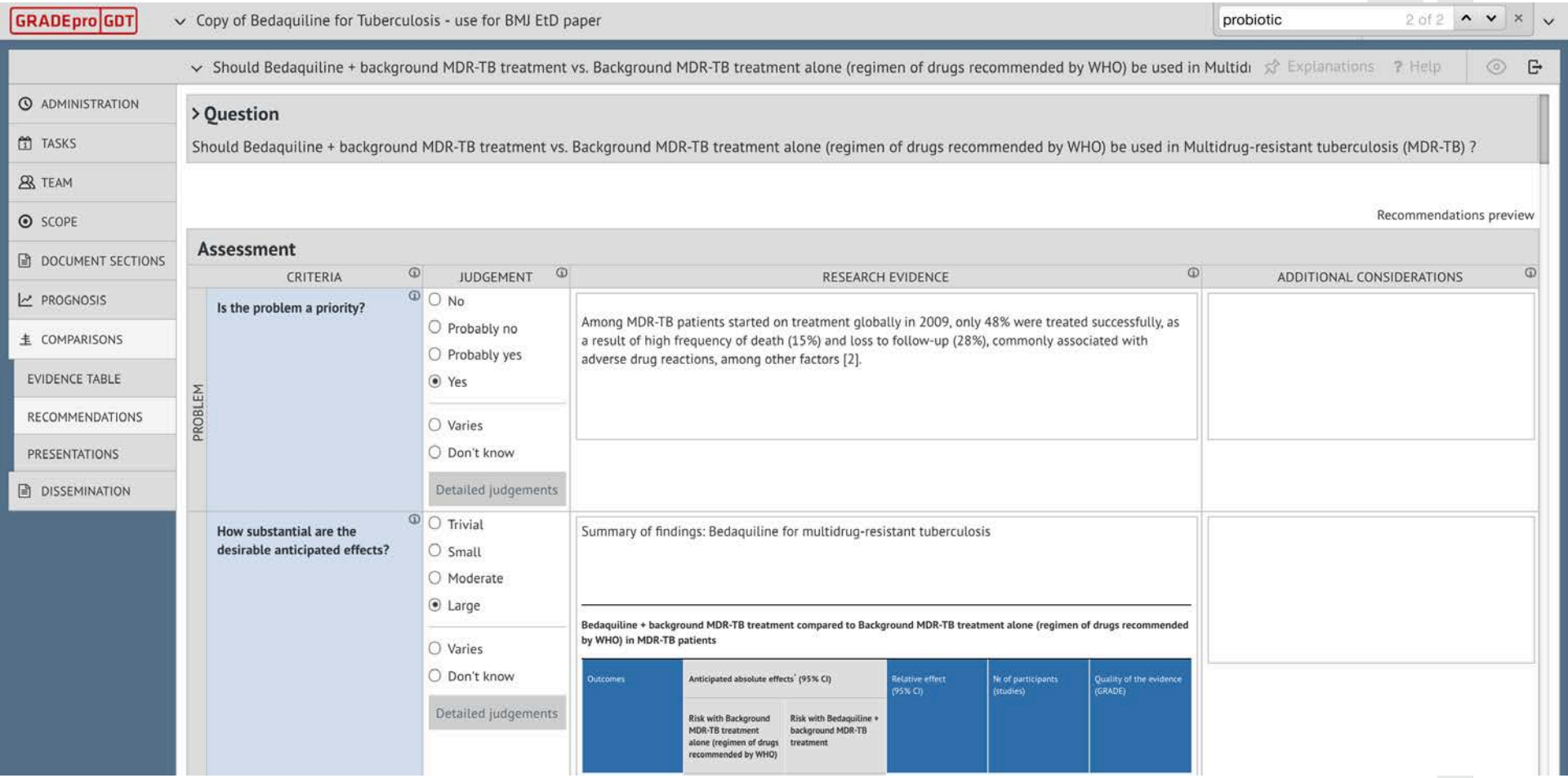
Considering each important factor that determine a decision (criteria)

Providing a concise summary of the best available research evidence to inform judgements

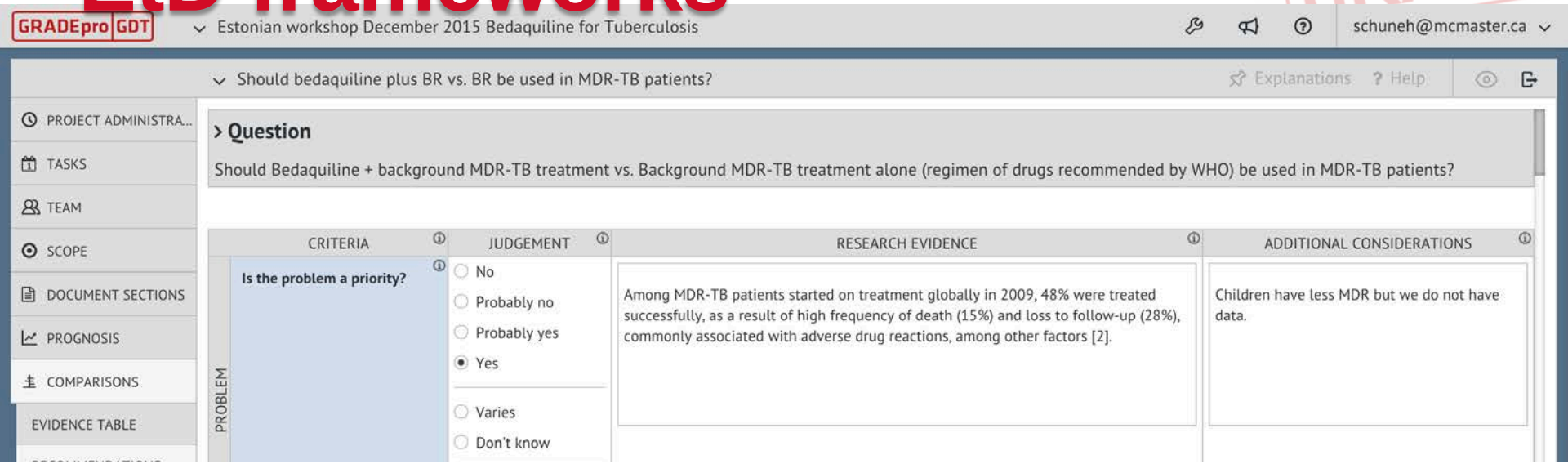
Helping to structure discussion and identify reasons for disagreements

Making the basis for decisions transparent and adaptable for target audiences

Interactive Evidence to Decision



EtD frameworks



Criteria on which a recommendation is based

Judgements that must be made in relation to each criterion

Research evidence to inform each judgement

Live use of iEtDs

EtDs are shared with panel members before the meeting and online:

Clarify the process

During the preparation for input on the evidence (all members including conflicted members could be involved)

For initial agreement on the included evidence and additional considerations

If possible, feasible and appropriate for agreement on judgments for specific decision criteria (but may all happen at an in-person meeting)

Final draft EtDs before a final meeting

What are guideline panel members doing?

Discuss evidence

GRADE

Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM **Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS **How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence.	

Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁◯◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁◯◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

Add relevant considerations

GRADE

Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

	CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	**Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS	**How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence.	

Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁◯◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁◯◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

Make judgments
(when research evidence complete) – w/o COI

> Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

	CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	**Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS	**How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence. **Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2**	

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁◯◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁◯◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

EtDs

messages

structured decision-making processes

transparent evidence syntheses that inform about the certainty in that evidence

- evidence profiles, evidence to decision frameworks with judgments

confidence in estimates of intervention effects only “a” part

accept uncertainty and be able to communicate it for better research and implementation