Versorgungsnahe Daten - Anwendbarkeit und Evidenzqualität

Holger Schünemann, MD, MSc, PhD, FRCPC

Professor of Medicine and Clinical Epidemiology

McMaster University

schuneh@mcmaster.ca

@schunemann_mac

McMaster University | GRADE Centre

Michael G. DeGroote COCHRANE CANADA CENTRE

Cochrane Canada

WHO Collaborating Center for Infectious Diseases, Research Methods and Recommendations

Disclosures

- No direct financial conflicts
- GRADE Working Group Co-Chair
- Cochrane Canada - Director
- Guidelines International Network – board
- Views expressed my own

Topics for today

- The questions we ask to inform health decisions
- Applicability versus risk of bias for the body of evidence on interventions
- What is “real world evidence” in that context?
- Which evidence for which question in the decision-making process?
- Overall certainty of evidence in answers addressing desirable and undesirable consequences of health care options

CA DE

Decision maker perspective!

Guideline developer (also conducting systematic reviews and HTA)

Clinician supporting people

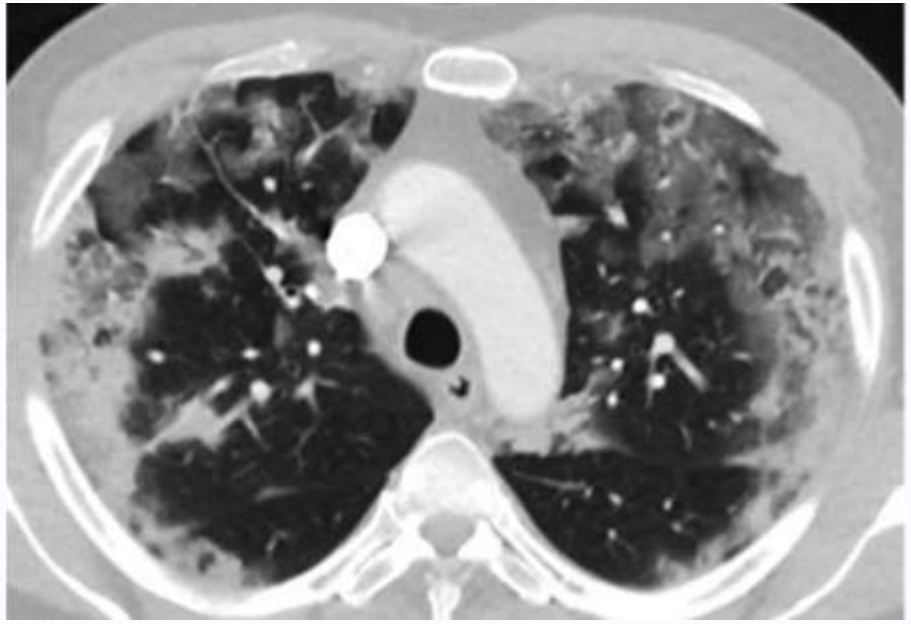
Benefits
Importance
Low co$t
Doable

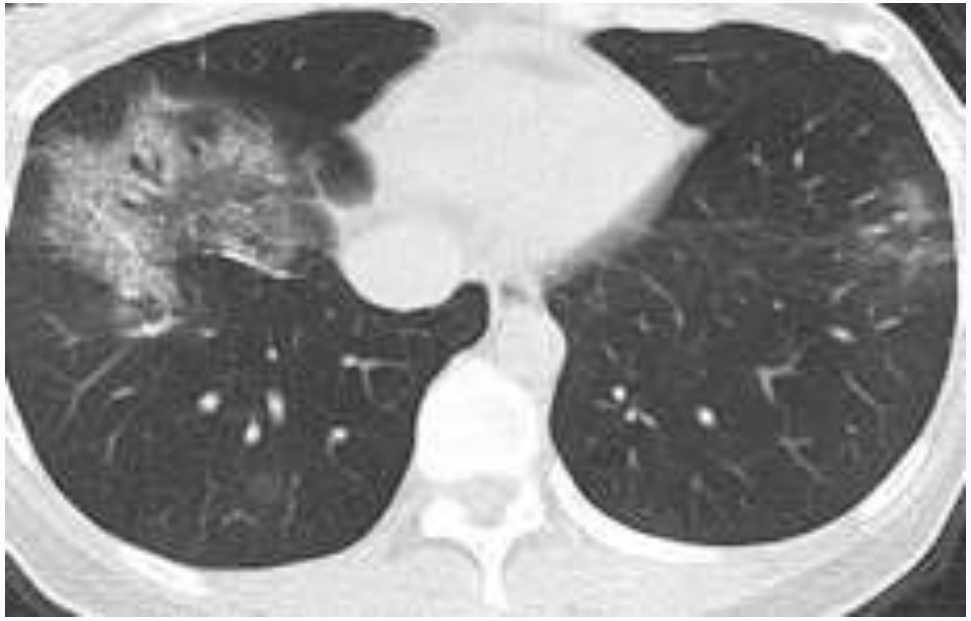
Harms
Importance
High co$$$t
~~Doable~~

The key question for people typically:
Health benefits outweigh the harms?
But many factors influence a decision

Balance of options and consequences

Covid-19 – respiratory failure: Antithrombotics for people with COVID-19





Foreground question

Population, Interventions, Comparison, Outcomes (PICO)

P: Critically ill people with COVID-19

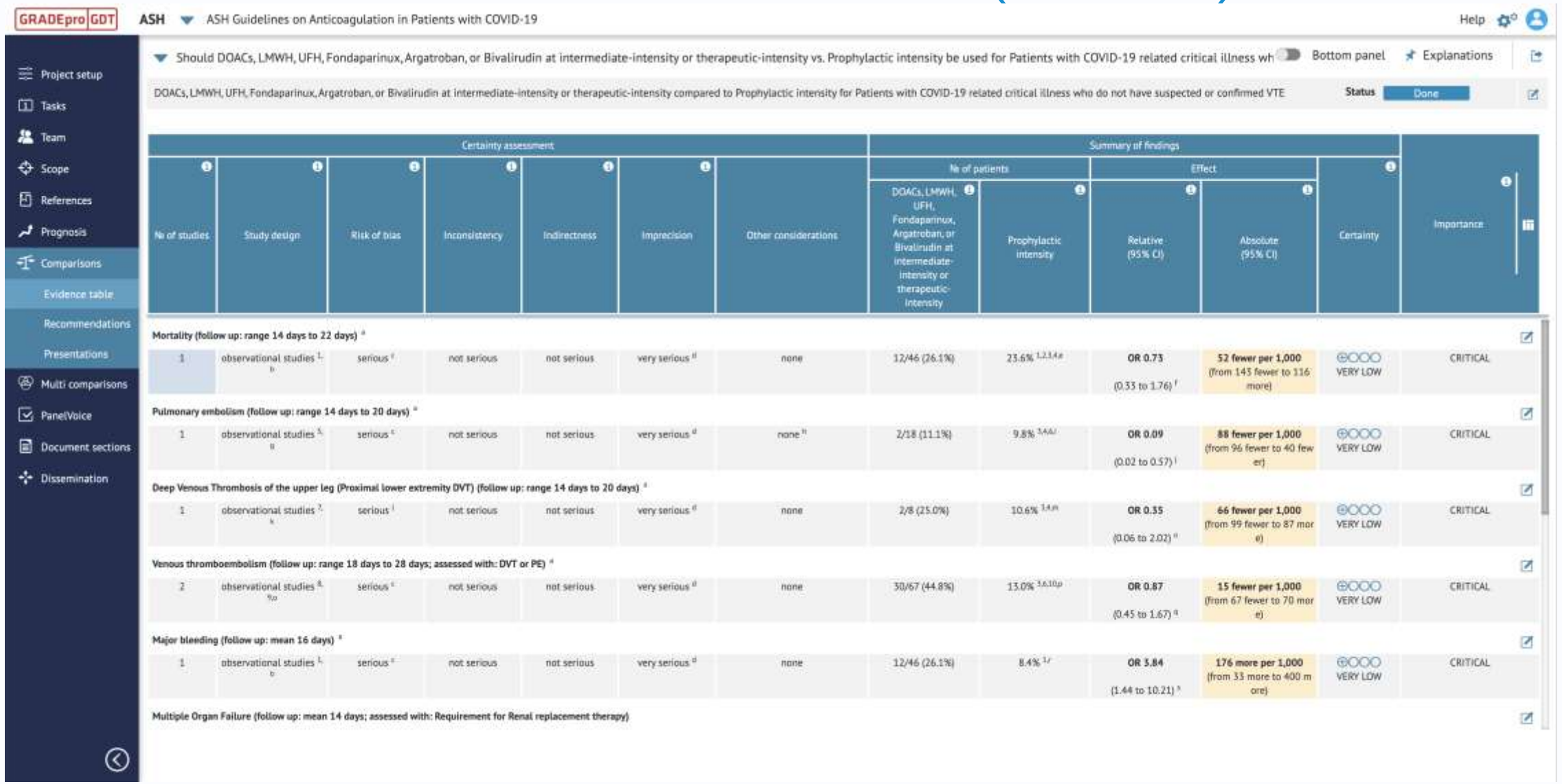
I: prophylactic anticoagulation (dose, type)

C: intermediate or therapeutic anticoagulation

O: Mortality, venous thromboembolism, bleeding…

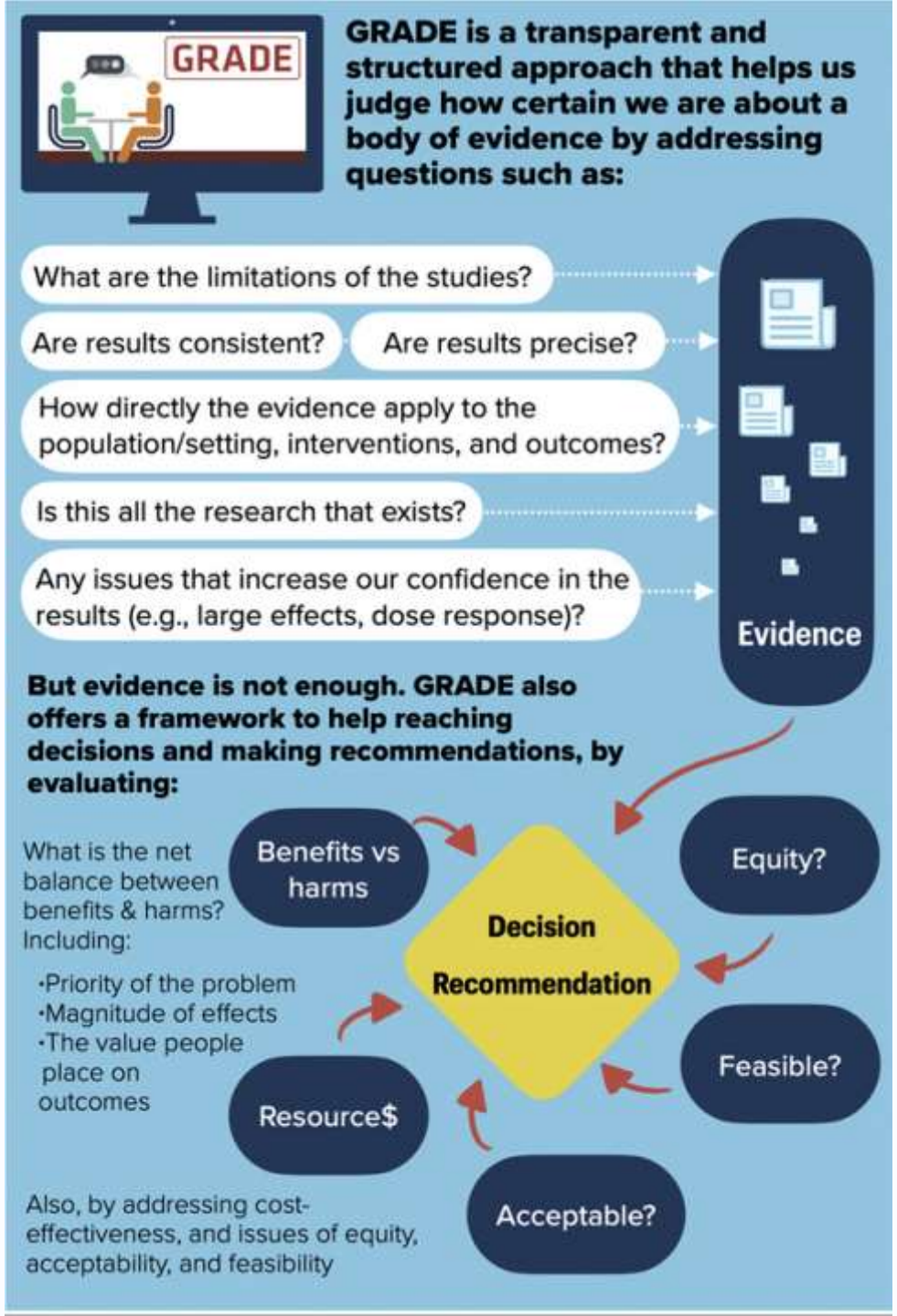
Systematic review ….

Randomized trials & Observational studies (Fall 2020)



GRADE evidence assessment

GRADE based recommendation and decision-making



GRADE working group

Unifying, transparent and sensible system for grading the certainty of evidence and making decisions

- WHO, NICE, CADTH, CDC, AHRQ, professional societies, academic institutions since 2000 – over 100 helped develop & use GRADE (>1000 people globally)
- For systematic reviews, HTA and guidelines, 1000's of recommendations
- Over 80,000 citations of publications:
  - Guidance articles, Concept articles, GRADE notes: BMJ, JCE, EHI
  - GRADE Handbook
  - GRADEpro app – guideline authoring, GRADE, dissemination, adaptation

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, AJRCCM 2006, Chest 2006, BMJ 2008, JCE 2011-2021

Best evidence

Effects: observational studies – “real world evidence”

Baseline risk: observational studies – “real world evidence”

Recommendation 1. The American Society of Hematology (ASH) guideline panel *suggests* using prophylactic-intensity over intermediate-intensity or therapeutic-intensity anticoagulation for patients with coronavirus disease 2019 (COVID-19)–related critical illness who do not have suspected or confirmed venous thromboembolism (VTE) (conditional recommendation based on very low certainty in the evidence about effects ⨁◯◯◯).

What is real world evidence in that context?

Language of *real world evidence* often suggests randomized trials come from an *unreal world*

Obscures risk of bias by emphasizing directness

Emphasis on *real world evidence* plays into the hands of those *manipulating data for gain*

Waters down what we should be focusing on – criteria for best health decisions

thebmjopinion

Holger J. Schünemann: All evidence is real world evidence

March 29, 2019

FDA revised definition

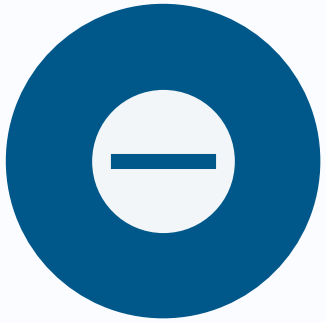
What is RWE?

FDA June 2019

Real-world ***evidence*** is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective).

Obscures risk of bias by emphasizing directness

Clinical questions in RCTs may be narrow but risk of bias lower



Non-randomized studies more applicable but higher risk of bias

Non-randomized *real world evidence* studies on intervention effects

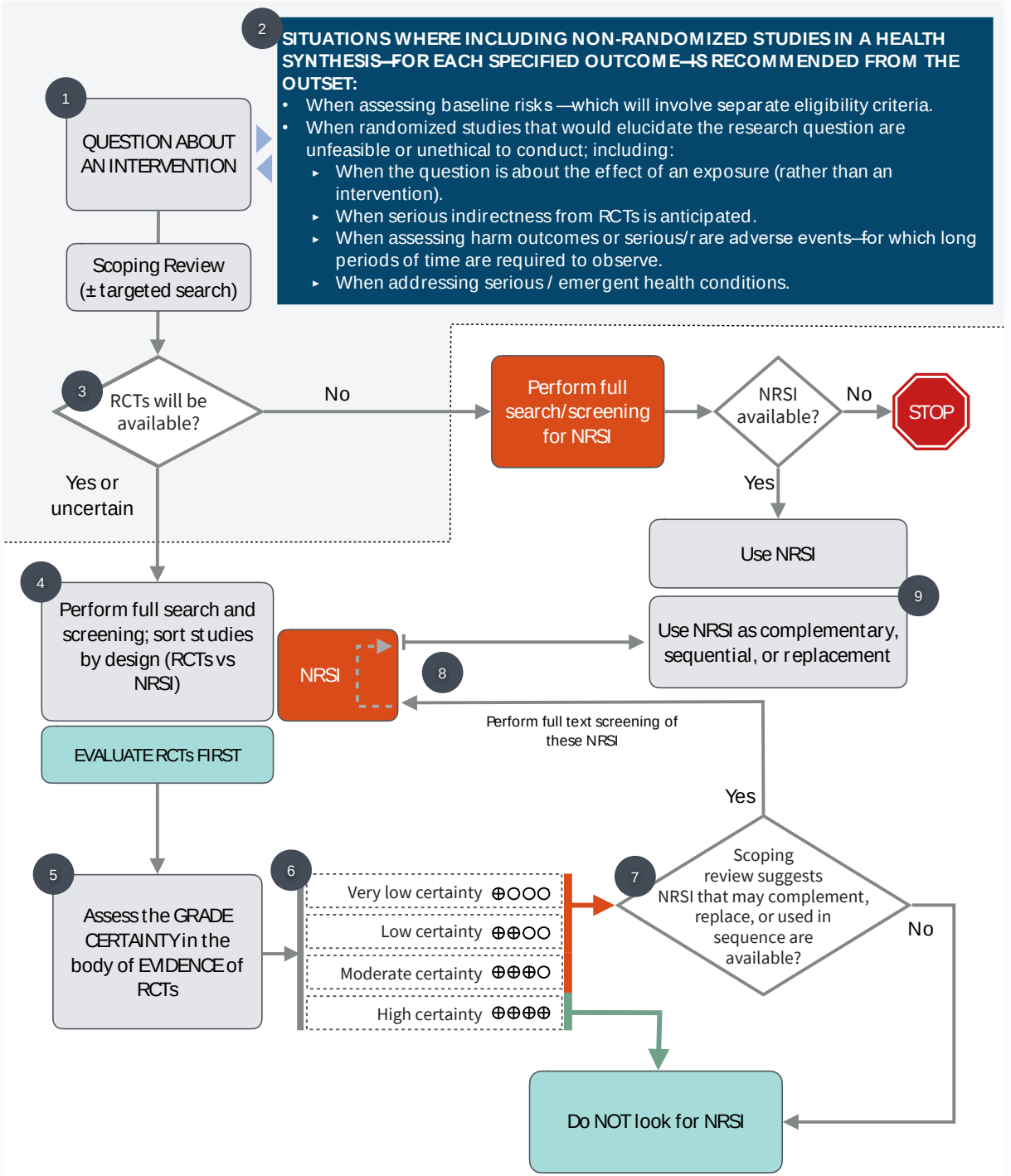
Directness

Limitations in study design and execution

Systematic error remains even if large or direct, error just gets more precise

“Even the **most direct real world evidence will not be sufficient to provide certainty** that an intervention has the intended effects if other bias is introduced when **decision makers assign interventions and fail to achieve a fair comparison** between an intervention and an alternative in producing *real world evidence*.”

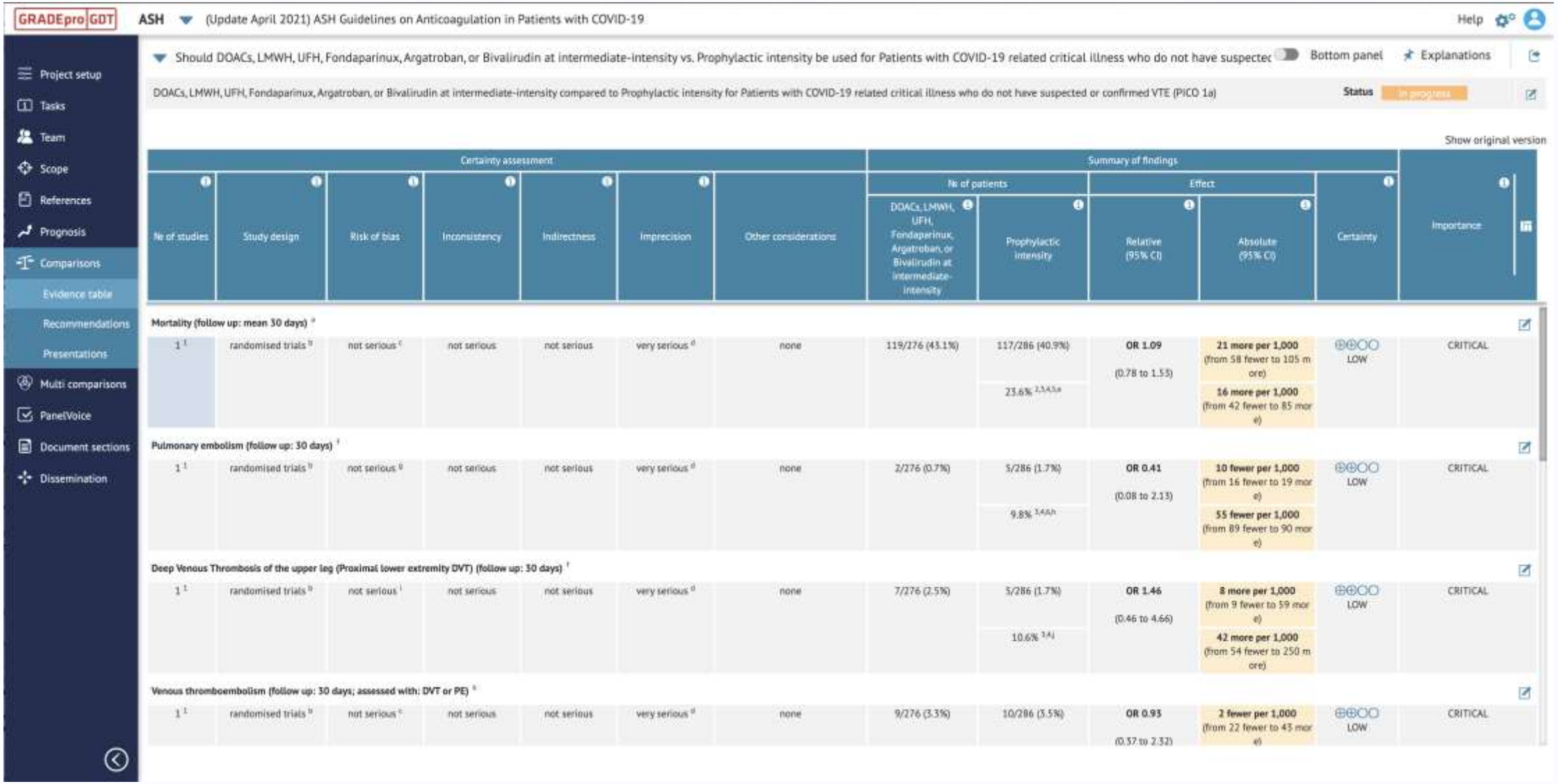
When do we look for non-randomized study evidence?



Cuello et al. Submitted

Systematic review ….

Randomized trials & Observational studies (Spring 2021)



GRADEpro GDT | ASH | (Update April 2021) ASH Guidelines on Anticoagulation in Patients with COVID-19 | Help

Project setup
Tasks
Team
Scope
References
Prognosis
Comparisons
Evidence table
Recommendations
Presentations
Multi comparisons
PanelVoice
Document sections
Dissemination

Should DOACs, LMWH, UFH, Fondaparinux, Argatroban, or Bivalirudin at intermediate-intensity vs. Prophylactic intensity be used for Patients with COVID-19 related critical illness who do not have suspected | Bottom panel | Explanations

DOACs, LMWH, UFH, Fondaparinux, Argatroban, or Bivalirudin at intermediate-intensity compared to Prophylactic intensity for Patients with COVID-19 related critical illness who do not have suspected or confirmed VTE (PICO 1a) | Status: In progress

Show original version

Certainty assessment							Summary of findings					
							№ of patients		Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	DOACs, LMWH, UFH, Fondaparinux, Argatroban, or Bivalirudin at intermediate-intensity	Prophylactic intensity	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
Mortality (follow up: mean 30 days) [a]												
1 [1]	randomised trials [b]	not serious [c]	not serious	not serious	very serious [d]	none	119/276 (43.1%)	117/286 (40.9%) 23.6% [2,3,4,5,e]	OR 1.09 (0.78 to 1.53)	21 more per 1,000 (from 58 fewer to 105 more) 16 more per 1,000 (from 42 fewer to 85 more)	⨁◯◯◯ LOW	CRITICAL
Pulmonary embolism (follow up: 30 days) [f]												
1 [1]	randomised trials [b]	not serious [g]	not serious	not serious	very serious [d]	none	2/276 (0.7%)	5/286 (1.7%) 9.8% [3,4,6,h]	OR 0.41 (0.08 to 2.13)	10 fewer per 1,000 (from 16 fewer to 19 more) 55 fewer per 1,000 (from 89 fewer to 90 more)	⨁◯◯◯ LOW	CRITICAL
Deep Venous Thrombosis of the upper leg (Proximal lower extremity DVT) (follow up: 30 days) [f]												
1 [1]	randomised trials [b]	not serious [i]	not serious	not serious	very serious [d]	none	7/276 (2.5%)	5/286 (1.7%) 10.6% [3,4,j]	OR 1.46 (0.46 to 4.66)	8 more per 1,000 (from 9 fewer to 59 more) 42 more per 1,000 (from 54 fewer to 250 more)	⨁◯◯◯ LOW	CRITICAL
Venous thromboembolism (follow up: 30 days; assessed with: DVT or PE) [k]												
1 [1]	randomised trials [b]	not serious [c]	not serious	not serious	very serious [d]	none	9/276 (3.3%)	10/286 (3.5%)	OR 0.93 (0.37 to 2.32)	2 fewer per 1,000 (from 22 fewer to 43 more)	⨁◯◯◯ LOW	CRITICAL

Best evidence (May 2021)

Effects: randomized trial (over observational studies)

Baseline risk: observational studies – “real world evidence”

The American Society of Hematology (ASH) guideline panel *suggests* using prophylactic-intensity over intermediate-intensity anticoagulation for patients with coronavirus disease 2019 (COVID-19)–related critical illness who do not have suspected or confirmed venous thromboembolism (VTE) (low certainty of evidence).

Remark:

- The ASH guideline panel plans to continue to update this recommendation when the full results of other trials become available. **Clinicians should weigh the potential benefits and harms based on the most up-to-date available evidence in caring for their patients.**
- **Separated intermediate and therapeutic anticoagulation.**
- Patients with COVID-19–related critical illness are defined as those suffering from an immediately life-threatening condition who would typically be admitted to an intensive care unit (ICU). Examples include patients requiring hemodynamic support, ventilatory support, and renal-replacement therapy.
- An individualized assessment of the patient's risk of thrombosis and bleeding is important when deciding on anticoagulation intensity. Risk-assessment models to estimate thrombotic and bleeding risk in hospitalized patients are available, but they have not been validated for patients with COVID-19. The panel acknowledges that higher-intensity anticoagulation may be preferred for patients judged to be at high thrombotic risk and low bleeding risk.
- At present, there is no direct high-certainty evidence comparing different types of anticoagulants. The selection of a specific agent (eg, low-molecular-weight heparin, unfractionated heparin, etc) may be based on availability, resources required, familiarity, and the aim of minimizing personal protective equipment (PPE) use or staff exposure to COVID-19–infected patients as well as patient-specific factors (eg, renal function, history of heparin-induced thrombocytopenia, concerns about gastrointestinal tract absorption).

Beyond intervention effects

The questions we ask to inform decisions

Guideline “should” question - Population, interventions, comparison

- Should therapeutic, intermediate or prophylactic doses be used in acutely ill COVID-19 patients?
- Should COVID-19 patients after discharge from the hospital receive thromboprophylaxis?

Evidence (systematic) reviews – Population, interventions, comparison, outcomes (PICO) questions

- In people with COVID-19, what is the impact of therapeutic compared with prophylactic anticoagulation on mortality, venous thromboembolism, bleeding, ….
- In people with COVID-19, what is the accuracy of pulmonary angiogram compared with compression ultrasound – review of test accuracy
- What value do people with COVID-19 place on outcomes mortality, venous thromboembolism, bleeding? – review of values and preferences
- In people with COVID-19, how cost effective is therapeutic compared with prophylactic anticoagulation? – review of cost-effectiveness studies
- In people with COVID-19 who are discharged, when compared to no anticoagulation is prophylactic anticoagulation feasible and/or acceptable to implement? – review of feasibility studies
- …

Decision/Recommendation - provide the answers to the “should” questions

- In people with COVID-19, the guideline panel recommends/suggests using prophylactic over therapeutic anticoagulation

Kriterien, Evidenzlage und zusätzliche Überlegungen, die Stärke und Ausrichtung der GRADE-Empfehlung beeinflussen

1. Problem: Ist das Problem eine Priorität?
2. Erwünschte Effekte: Wie groß sind die zu erwartenden erwünschten Effekte?
3. Unerwünschte Effekte: Wie groß sind die zu erwartenden unerwünschten Effekte?
4. Vertrauenswürdigkeit der Evidenz: Wie ist das gesamte Vertrauen in die Evidenz der Effekte?
5. Wertvorstellungen: Gibt es wesentliche Unsicherheit darüber, welche Wichtigkeit Betroffene den Hauptendpunkten beimessen oder variieren die Einschätzungen?
6. Abwägen der Effekte: Spricht ein Abwägen von erwünschten und unerwünschten Wirkungen für die Intervention oder die Vergleichsintervention?
7. Benötigte Ressourcen: Wie hoch ist der Ressourcenbedarf (Kosten)?
8. Vertrauenswürdigkeit der Evidenz der benötigten Ressourcen: Wie groß ist das Vertrauen in die Evidenz zum Ressourcenbedarf (Kosten)?
9. Kostenwirksamkeit: Spricht die Kostenwirksamkeit für die Intervention oder die Vergleichsintervention?
10. Gleichstellung: Was wären die Auswirkungen auf die Gerechtigkeit im Gesundheitswesen?
11. Akzeptanz: Ist die Intervention für die wichtigsten Interessengruppen akzeptabel?
12. Machbarkeit: Ist es möglich die Intervention umzusetzen?

Quelle: GRADE EtD framework

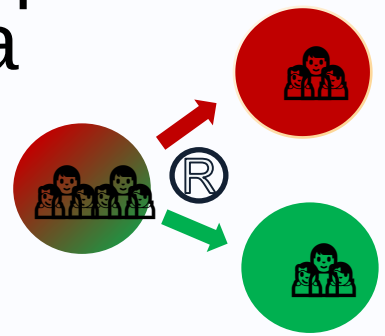
Abbildung 2: Kriterien, Forschungsevidenz, Entscheidungen und weitere Überlegungen, die die Ausprägung und Ausrichtung bei den EtD-Ansätzen der GRADE-Methodik beeinflussen.

EtDs by GRADE Working Group, e.g. Schünemann Aerzteblatt 2019

1. **Problem** ❗
   Is the problem a priority?
2. **Desirable Effects**
   How substantial are the desirable anticipated effects?
3. **Undesirable Effects**
   How substantial are the undesirable anticipated effects?
4. **Certainty of evidence**
   What is the overall certainty of the evidence of effects?
5. **Values** ❗
   Is there important uncertainty about or variability in how much people value the main outcomes?
6. **Balance of effects** ❗
   Does the balance between desirable and undesirable effects favor the intervention or the comparison?
7. **Resources required**
   How large are the resource requirements (costs)?
8. **Certainty of evidence of required resources**
   What is the certainty of the evidence of resource requirements (costs)?
9. **Cost effectiveness**
   Does the cost-effectiveness of the intervention favor the intervention or the comparison?
10. **Equity**
    What would be the impact on health equity?
11. **Acceptability**
    Is the intervention acceptable to key stakeholders?
12. **Feasibility**
    Is the intervention feasible to implement?

Evidence to Decision Criteria

Randomized designs applicable to many of these criteria but EHR, registries etc. provide informative data



Non-randomized designs definitely for priority of the problem, baseline risk, values

Balance of benefits and harms

- Reduction in venous thromboembolism, respiratory failure?
- Increased in major bleeding
- Not without knowing the values placed on these outcomes

Systematic review of **values & preference studies: non-randomized studies** (e.g. health services research)

Table 1. Summary of findings for overall population: utility, RIO, or health state value information

Health state/outcome (categories of values and preferences)	Estimates of utilities No. of participants/studies	Certainty in evidence	Interpretation of findings
DVT* (Hogg et al,[27,28] Lloyd et al,[31] Locadia et al,[32] Marvig et al,[34] Utne et al[37])	Range across studies: 0.61-0.99 Standard gamble: 0.81-0.99 Time trade-off: 0.84 VAS: 0.65-0.72 EQ-5D utility: 0.61-0.79 SF-6D: 0.64 1702 participants from 6 studies‡ Standard gamble: 260 participants from 2 studies Time trade-off: 124 participants from 1 study VAS: 485 participants from 3 studies EQ-5D utility: 1318 participants from 4 studies SF-6D: 44 participants from 1 study	⊕⊕⊕◯ Moderate certainty due to inconsistency†	People may probably find DVT having a moderate or a trivial impact on their lives. There is likely an important variability for this assessment.
Gastrointestinal tract bleeding event (Hogg et al,[27] Lloyd et al,[31] Locadia et al[32])	Range across studies: 0.59-0.65 Standard gamble: 0.65 (IQR, 0.15-0.86) Time trade-off: 0.65 (IQR, 0.49-0.86) Rating scale modeled: EQ-5D 0.59 (95% CI, 0.46-0.69) 1217 participants Standard gamble: 216 participants from 1 study Time trade-off: 124 participants from 1 study Rating scale modeled: EQ-5D: 877 patients from 1 study	⊕⊕⊕◯ Moderate certainty due¶ to indirectness**	People may probably find gastrointestinal tract bleeding having a moderate impact on their lives. There is likely an important variability for this assessment.

Zhang et al. 2019, Etxeandia-Ikobaltzeta et al. 2020

Interaktive Evidence-to-Decision-Ansätze

- Settings
- Tasks
- Team
- Scope
- References
- Prognosis
- Comparisons
- Multi comparisons
- PanelVoice
- Document sections
- Dissemination

Frage
- Details – PICO-Subgruppen
- Hintergrund und Interessenskonflikte

Bewertung
- Kriterien
- Entscheidungen
- Forschungsevidenz (HTA und systematische Überprüfungen)
- Weitere Überlegungen

GRADE

Schlussfolgerungen
- Art der Entscheidung – Empfehlung
- Entscheidung
- Überlegungen zur Implementierung – Überwachung und Beurteilung
- Forschungsüberlegungen

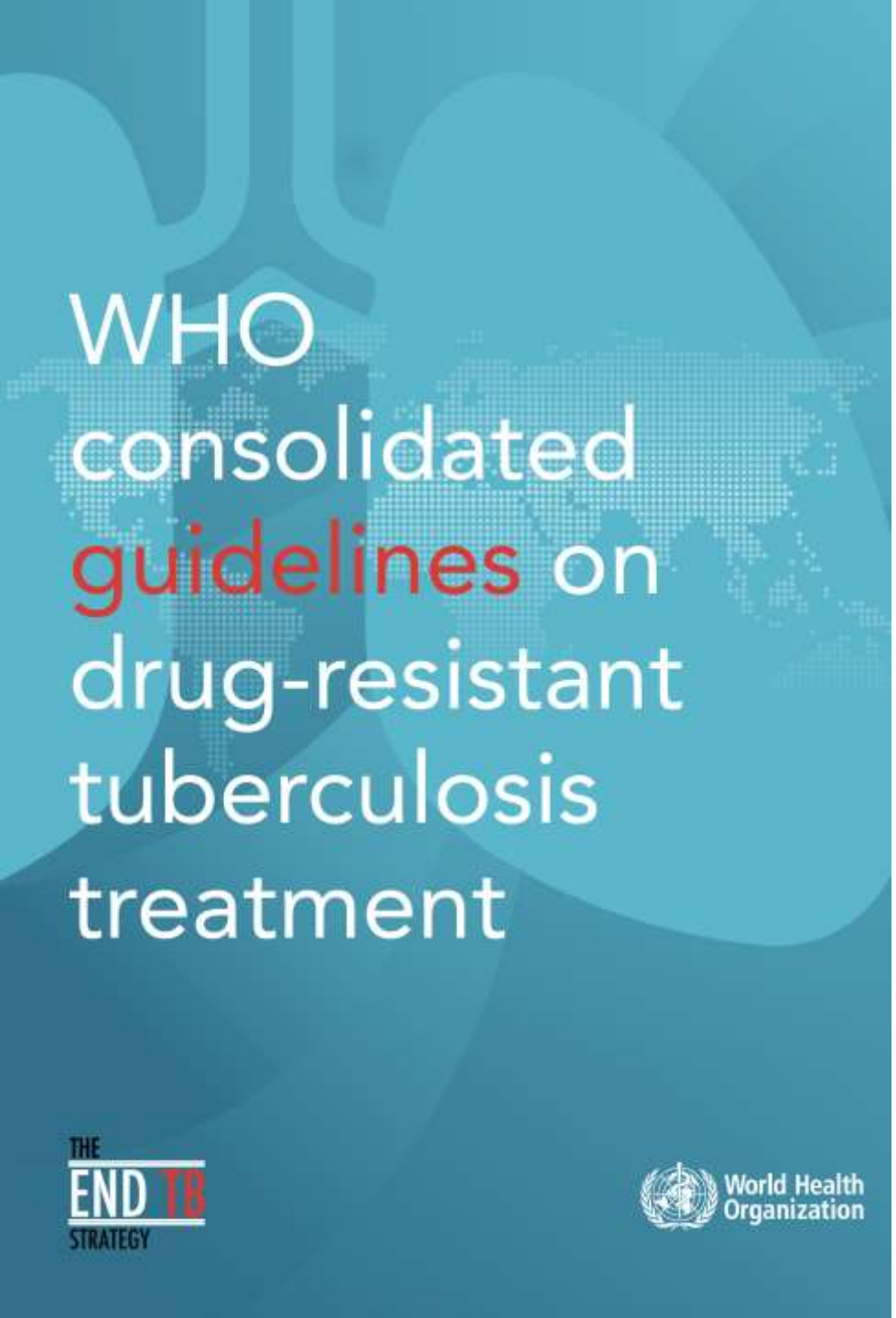
Präsentation
- Treffen der Richtliniengruppe & Mitteilung von Kostenübernahmeentscheidungen
- Datenbank mit Entscheidungsrahmen
- Interactive Decision Aids (iDeAs), Apps

Perspektiven
- Klinische – persönliche
- Klinische – Bevölkerung
- Gesundheitssysteme & Öffentliche Gesundheit
- Gesundheitssysteme & Politik

Art von Entscheidungen
- Empfehlung
- Richtlinie
- Kostenübernahme

Verwendung
- Gruppenentscheidungsfindung
- Persönlich / Online

Quelle: gradepro.org



Of course there are scenarios where non-randomized studies trump

Bedaquiline as add-on for multi-drug resistant tuberculosis – outcome mortality

First evaluation - 2013: Increase in death from **two RCTs** (n = 160) : imprecision (10 events) and indirectness (few countries, few patients)

Very low certainty ⊕○○○

Second evaluation - 2016: Reduction in death from non-randomized studies summarized in an **individual patient data meta-analysis (IPDMA) with selection bias** (n = 1556) + indirectness

Very low certainty ⊕○○○

Third evaluation - 2018: Large reduction in death from non-randomized studies **IPDMA (n = 4,600 patients - 1391 received bedaquiline – 53 studies, 40 countries) - large effect**, adjustment without further concerns about risk of bias; no indirectness

Moderate certainty ⊕⊕⊕○

Summary

- One question about an intervention
  - Answer requires different types of systematic reviews to address the criteria that influence a recommendation or decision

- Non randomized studies, e.g. from health services research will play a role to address many of these criteria
  - But there is a risk in trading off bias against applicability