Understanding the GRADE approach to evidence synthesis with special reference to guideline development

Nandi Siegfried[1,2] and Tamara Kredo[1]

[1] Health Systems Research Unit
[2] Mental Health, Alcohol, Substance Use and Tobacco Research Unit

South African Medical Research Council

Cochrane Africa

South Africa GRADE Network

saMRC advancinglife

The South African Medical Research Council

recognizes the catastrophic and persisting consequences of colonialism and apartheid, including land dispossession and the intentional imposition of educational and health inequities.

Acknowledging the SAMRC's historical role and silence during apartheid, we commit our capacities and resources to the continued promotion of justice and dignity in health research in South Africa.

transformation

WORKING TOGETHER FOR EXCELLENCE

Assumptions

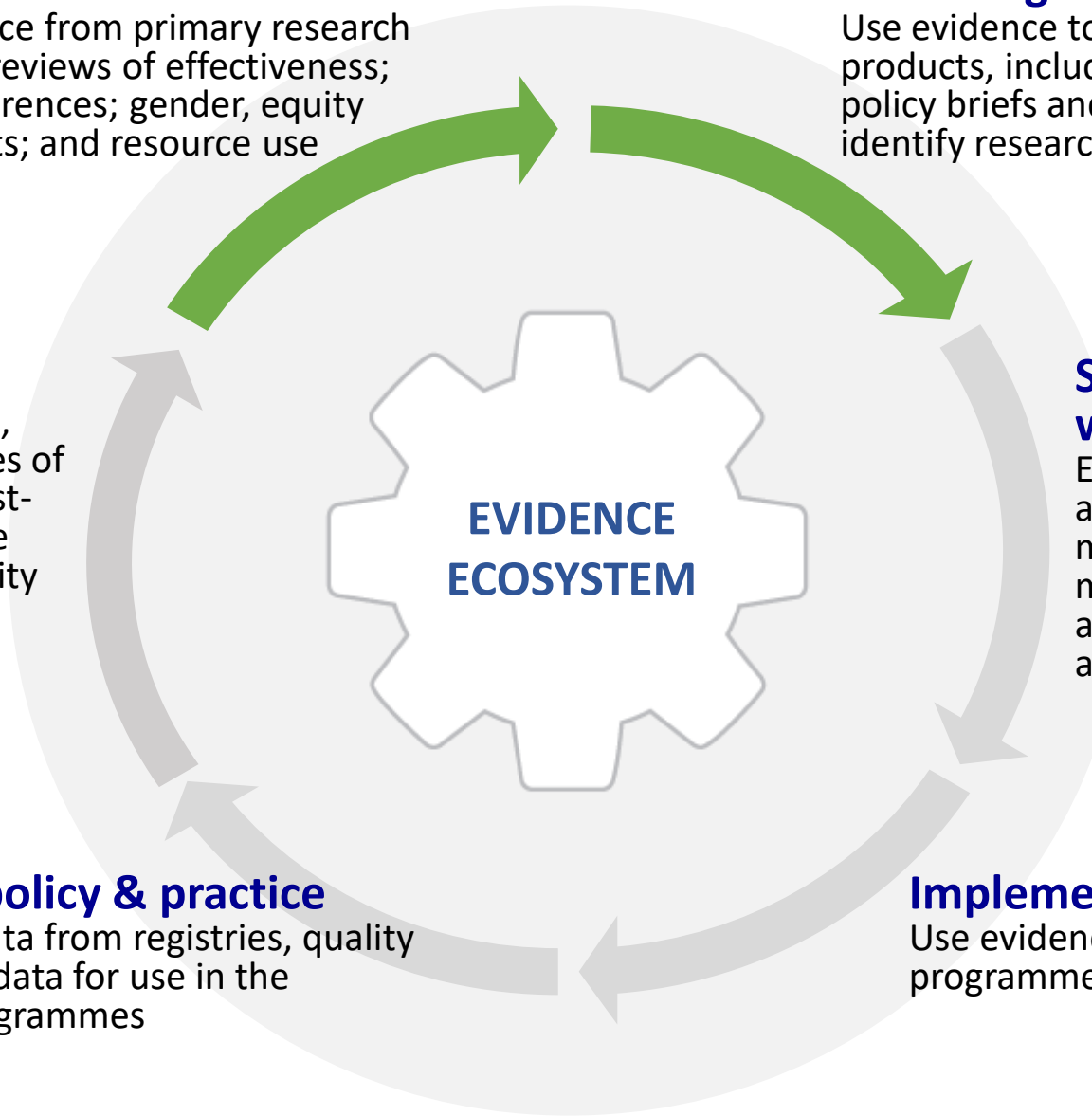
- ✓ Good understanding of difference between quantitative and qualitative evidence
- ✓ Good understanding of the hierarchy of study design related to effectiveness
- ✓ Basic knowledge of systematic reviews
- ✓ Basic knowledge of meta-analysis

Synthesize evidence

Combine evidence from primary research into systematic reviews of effectiveness; values and preferences; gender, equity and human rights; and resource use

Knowledge translation

Use evidence to inform decision support products, including guidelines, guidance, policy briefs and evidence summaries, and to identify research gaps



Produce evidence

Undertake primary research, including quantitative studies of effectiveness, safety and cost-effectiveness and qualitative studies of uptake, applicability and feasibility

Share evidence with stakeholders

Ensure evidence of beneficial and harmful interventions is made available to decision makers, healthcare providers, and the public, in an accessible and user-friendly way

Evaluate and improve policy & practice

Consider population-based data from registries, quality indicators and programmatic data for use in the evaluation of policies and programmes

Implement evidence

Use evidence to inform policies and programme

OUTLINE

What we will cover today

1. Why is there a need for a structured process for guidelines development?
2. What is GRADE for evidence synthesis?
   - ✓ GRADE for quantitative evidence including interactive learning
   - ✓ GRADE CERQual for qualitative evidence
3. How do we use GRADE for guidelines development?
   - ✓ Evidence-to-decision-making (ETD) tables
   - ✓ Formulating recommendations
4. Resources and links

WHY

Use of evidence in WHO recommendations

Andrew D Oxman, John N Lavis, Atle Fretheim

Summary

Background WHO regulations, dating back to 1951, emphasise the role of expert opinion i
recommendations. However, the organisation's guidelines, approved in 2003, emphasise the us
for evidence of effects, processes that allow for the explicit incorporation of other types of i
values), and evidence-informed dissemination and implementation strategies. We examine
particularly evidence of effects, in recommendations developed by WHO departments.

Methods We interviewed department directors (or their delegates) at WHO headquarters in Ge
reviewed a sample of the recommendation-containing reports that were discussed in the interv
background documentation). Two individuals independently analysed the interviews and revie
reports and background documentation.

Findings Systematic reviews and concise summaries of findings are rarely used for developi
Instead, processes usually rely heavily on experts in a particular specialty, rather than repres
will have to live with the recommendations or on experts in particular methodological areas.

Interpretation Progress in the development, adaptation, dissemination, and implementation o
member states will need leadership, the resources necessary for WHO to undertake these pro
and defensible way, and close attention to the current and emerging research literature related

Systematic reviews and concise summaries of findings are rarely used for developing recommendations. Instead, processes usually rely heavily on experts in a particular specialty, rather than representatives of those who will have to live with the recommendations or on experts in particular methodological areas.

WHAT

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide

Guideline developers around the world are inconsistent in how they rate quality of evidence and grade strength of recommendations. As a result, guideline users face challenges in understanding the messages that grading systems try to communicate. Since 2006 the *BMJ* has requested in its "Instructions to Authors" on bmj.com that authors should preferably use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system for grading evidence when submitting a clinical guidelines article. What was behind this decision?

In this first in a series of five articles we will explain why many organisations use formal syste... evidence and recommend... important for clini... approach t... we will exa... quality of ev... The final two ... for diagnostic ... ling the impact ... GRADE has ... (box 1). Other sy... but none, other t...

What is "quality of ...

In making healthca... ...ons, patients and clinicians must ... the benefits and downsides of alternative strategies. Decision makers will be influenced not only by the best estimates of the expected advantages and disadvantages but also by their confidence in these estimates. The cartoon depicting the weather forecaster's uncertainty captures the difference between an assessment of the likelihood of an outcome and the confidence in that assessment (figure). The usefulness of an estimate of the magnitude of intervention effects depends on our ... in that estimate.

Expert ... ...ns offering recommendations ... ...y have often erred ... ...unt of the quality ... ...ns recommended ... ...al women to use ... ...imary care physicians' practices. ... ...lly decreased ... ...ecommendations ... ...uality of evidence ... ...ave shown ... ...nal studies ... ...for a reduction in ... ...very low quality.[4] Recognition ... ...ations of the evidence would have tempered ... recommendations. Ultimately, randomised controlled trials have shown that hormone replacement therapy fails to reduce cardiovascular risk and may even increase it.[5 6]

The US Food and Drug Administration licensed the antiarrhythmic agents encainide and flecainide for use in patients on the basis of the drugs' ability to reduce asymptomatic ventricular arrhythmias associated with sudden death. This decision failed to acknowledge that because arrhythmia reduction reflected only indirectly on the outcome of sudden death the quality of the evidence for the drugs' benefit was of low quality. Subsequently, a randomised controlled trial showed that the two drugs increase the risk of sudden death.[7] Appropriate attention to the low quality of the evidence would have saved thousands of lives.

Failure to recognise high quality evidence can cause similar problems. For instance, expert recommendations lagged a decade behind the evidence from well conducted randomised controlled trials that thrombolytic therapy achieved a reduction in mortality in myocardial infarction.[8]

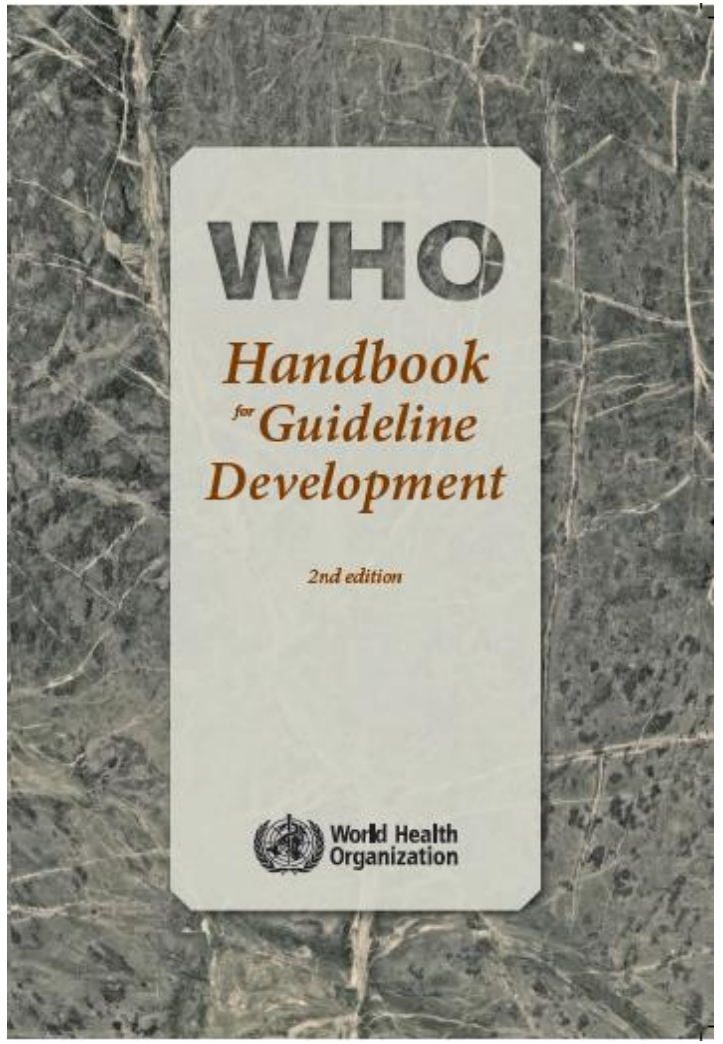
Insufficient attention to quality of evidence risks inappropriate guidelines and recommendations that may lead clinicians to act to the detriment of their

Box 1 | Advantages of GRADE over other systems

- Developed by a widely representative group of international guideline developers
- Clear separation between quality of evidence and strength of recommendations
- Explicit evaluation of the importance of outcomes of alternative management strategies
- Explicit, comprehensive criteria for downgrading and upgrading quality of evidence ratings
- Transparent process of moving from evidence to recommendations
- Explicit acknowledgment of values and preferences
- Clear, pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers
- Useful for systematic reviews and health technology assessments, as well as guidelines

Gordon H Guyatt professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada L8N 3Z5
Andrew D Oxman researcher, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs Plass, 0130 Oslo, Norway
Gunn E Vist researcher, Norwegian Knowledge Centre for the Health Services, P... St Olavs P...
...ial), Hospital ... edu, Barcelona 08041, Spain
Holger J Schünemann professor, Department of Epidemiology, Italian National Cancer Institute Regina Elena, Rome, Italy
for the GRADE Working Group
Correspondence to: G H Guyatt, CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, Room 2C12, 1200 Main Street West Hamilton, ON, Canada L8N 3Z5 guyatt@mcmaster.ca

This is the first in a series of five articles that explain the GRADE system for rating the quality of evidence and strength of recommendations.



SO WHAT

What makes GRADE special?

- Sequential assessment of
  - Certainty of evidence
  - Judgment about the balance between desirable and undesirable effects
  - Decision about the strength of a recommendation
- Separating the judgments regarding the certainty of evidence from judgments about the strength of recommendations is a critical and defining feature of the GRADE grading system

✓ Quantitative Evidence & GRADE Evidence Profiles

Step 1: Formulate the PICO question

PICO Question: Among people with TB with or without undernutrition, who are receiving TB treatment, do micronutrient supplements improve physical and mental health and wellbeing compared with TB treatment alone?

Population	Intervention	Comparison	Outcome
People with clinically or microbiologically diagnosed TB who are receiving TB treatment	• Single micronutrient supplement (e.g. Vitamin A, Folic acid…) • Multi-micronutrient supplements (a combination of two or more of the above)	• No micronutrient intervention • Different micronutrient interventions	**1. TB Treatment outcomes** • Time to sputum conversion **2. Nutritional outcomes** • Weight gain **3. Health and welfare outcomes** • Mental Health (e.g. depression)

Step 2: Selecting Outcomes

- Outcomes should be **important to those affected** by the guidelines

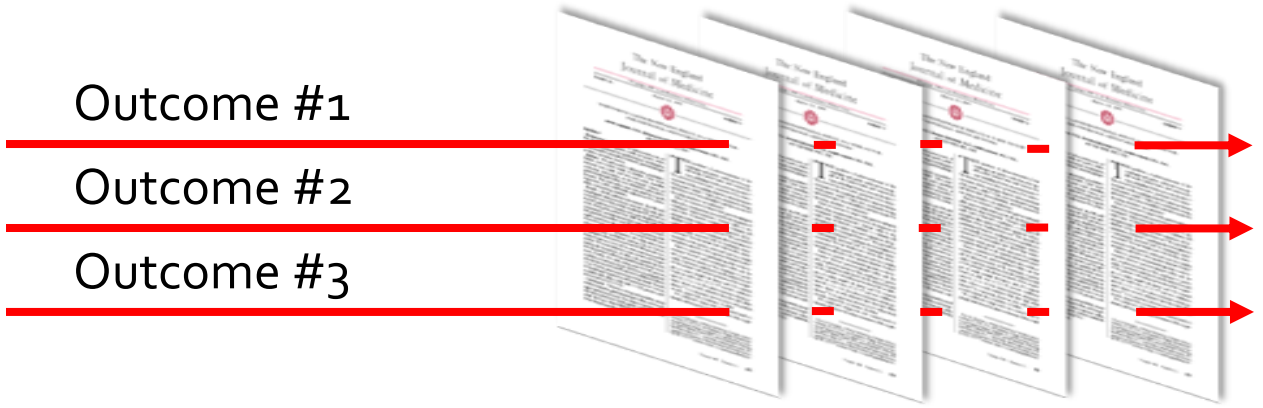
- GRADE rates the certainty of evidence for each outcome separately
  - The source of evidence may be different across outcomes
  - The same source of evidence can provide varying certainty of evidence for the different outcomes

Step 3. Rating the outcomes

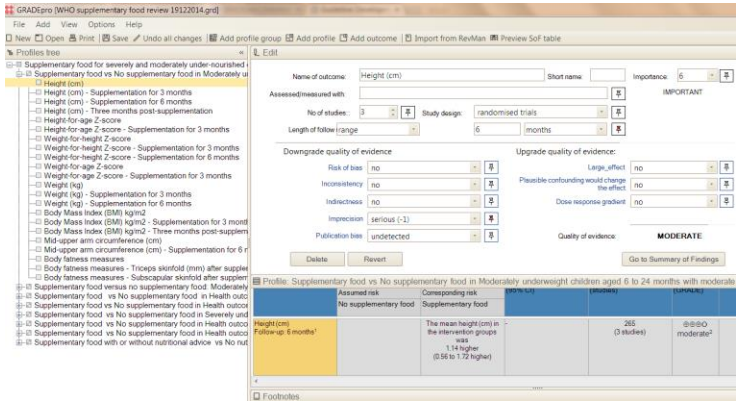
RATING	IMPORTANCE
9	Critical
8	
7	
6	Important
5	
4	
3	Not important
2	
1	

Only outcomes considered **critical** (rated 7—9) are the primary factors influencing a recommendation and should be used to determine the overall certainty of evidence supporting a recommendation

Step 4. Systematic review



Step 5. Create Evidence Profile



Step 5. Create a GRADE Evidence Profile

- Presents a graphical summary of the systematic review per PICO

- Summarises
  - Relative estimates of effect
  - Absolute estimates of effect
  - Certainty of the estimates for selected outcomes

Author(s): YFY, HJS, EAA
Date: 2008-09-14
Question: Should Parenteral anticoagulation be used for patient with cancer?
Settings: Outpatient
Bibliography: Akl EA, van Doormaal FF, Barba M, Kamath G, Kim SY, Kuipers S, Middeldorp S, Yosuico V, Dickinson HO, Schünemann HJ. Parenteral anticoagulation for prolonging survival in patients with cancer who have no other indication for anticoagulation. Cochrane Database of Systematic Reviews 2007, Issue 3

Quality assessment							Summary of findings					
							No of patients		Effect			
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Parenteral anticoagulation	control	Relative (95% CI)	Absolute	Quality	Importance
Mortality at 12 months (follow-up 1-7 years)												
5	randomized trial	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	339/586 (57.8%)	390/588 (66.3%)	RR 0.87 (0.8 to 0.95)	86 fewer per 1000 (from 33 fewer to 133 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
								50%		64 fewer per 1,000		
								90%		116 fewer per 1,000		
Major bleeding (follow-up 1-7 years)												
3	randomized trial	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	8/406 (2%)	6/408 (1.5%)	RR 1.5 (0.26 to 8.8)	7 more per 1000 (from 11 fewer to 117 more)	⊕⊕⊕O MODERATE	CRITICAL
Minor bleeding (follow-up 1-7 years)												
3	randomized trial	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	14/380 (3.7%)	5/380 (1.3%)	RR 2.07 (0.78 to 5.51)	14 more per 1000 (from 3 fewer to 59 more)	⊕⊕⊕O MODERATE	IMPORTANT
DVT (follow-up 1-7 years)												
2	randomized trial	no serious limitations	no serious inconsistency	no serious indirectness	very serious[2]	reporting bias[3]	1/232 (0.4%)	2/226 (0.9%)	RR 0.61 (0.06 to 4.91)	4 fewer per 1000 (from 8 fewer to 35 more)	⊕OOO VERY LOW	IMPORTANT

[1] The 95% confidence interval includes both no increased risk of bleeding as well as substantial increased risk of bleeding
[2] Only 2 events in the placebo group
[3] Only 2 trials reported DVT - reporting bias may be present

How to read an Evidence Profile

Question: Should a 3-month regimen of weekly rifapentine plus isoniazid be offered as an alternative regimen to daily isoniazid monotherapy for treatment of LTBI in high TB burden countries?

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3-month weekly RPT+INH	6 or 9 months INH	Relative (95% CI)	Absolute		
Active tuberculosis												
2	randomised trials	not serious	not serious	not serious [a]	serious [b]	none	26/534 (4.9%)	28/520 (5.4%)	**RR 0.733** (0.234 to 2.295)	**14 fewer per 1,000** (from 41 fewer to 70 more)	⊕⊕⊕◯ MODERATE	CRITICAL

How to read an Evidence Profile

PICO: Should 3-month regimen of weekly rifapentine plus isoniazid be offered as an alternative regimen to daily isoniazid monotherapy for treatment of LTBI in high TB burden countries?

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3-month weekly RPT+INH	6 or 9 months INH	Relative (95% CI)	Absolute		
Active tuberculosis												
2	randomised trials	not serious	not serious	not serious [a]	serious [b]	none	26/534 (4.9%)	28/520 (5.4%)	**RR 0.733** (0.234 to 2.295)	**14 fewer per 1,000** (from 41 fewer to 70 more)	⨁⨁⨁◯ MODERATE	CRITICAL

b. 95CI of both relative and absolute effect include appreciable benefit and harm with 3HP

How do we determine certainty?

- ✓ RCTs start as HIGH level certainty
- ✓ Observational studies start as LOW level certainty

- ✓ Levels can be DOWN-graded on the following five factors

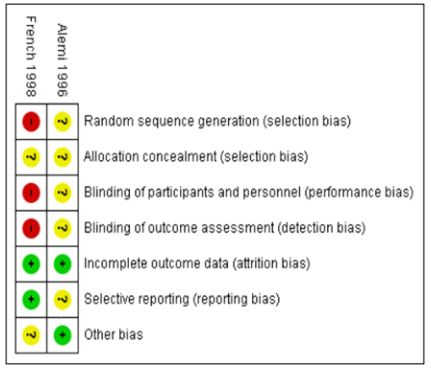
Risk of bias	Inconsistency of results	Indirectness of evidence	Imprecision of results	Other considerations

Interpretation of GRADE certainty ratings

Rating	Interpretation
High	We are *very confident* that the true effect lies close to that of the estimate of the effect
Moderate	We are *moderately confident* in the estimate of effect: The true effect is likely to be close to the estimate of effect, but possibility to be substantially different
Low	Our confidence in the effect is *limited*: The true effect may be substantially different from the estimate of the effect
Very Low	We have *very little confidence* in the effect estimate: Any estimate of effect is very uncertain

- We consider DOWN-grading
  - Inadequate allocation concealment
  - Inadequate masking
  - No true intention-to-treat principle
  - High attrition

RCTs



- We consider DOWN-grading
  - Selection Bias
  - Measurement Bias
  - Confounding
  - Incomplete or inadequate follow-up

Observational Studies

Recent advances

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

ELSEVIER

Check for updates

Journal of Clinical Epidemiology

Journal of Clinical Epidemiology 111 (2019) 105–114

ORIGINAL ARTICLE

GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence

Holger J. Schünemann[a,b,*], Carlos Cuello[a], Elie A. Akl[a,c], Reem A. Mustafa[a,d], Jörg J. Meerpohl[e], Kris Thayer[f], Rebecca L. Morgan[a], Gerald Gartlehner[g], Regina Kunz[h], S Vittal Katikireddi[i], Jonathan Sterne[j], Julian PT Higgins[j], Gordon Guyatt[a,b], GRADE Working Group

[a]Department of Health Research Methods, Evidence, and Impact and McGRADE Center, McMaster University, 1280 Main Street West, Hamilton, Ontario,

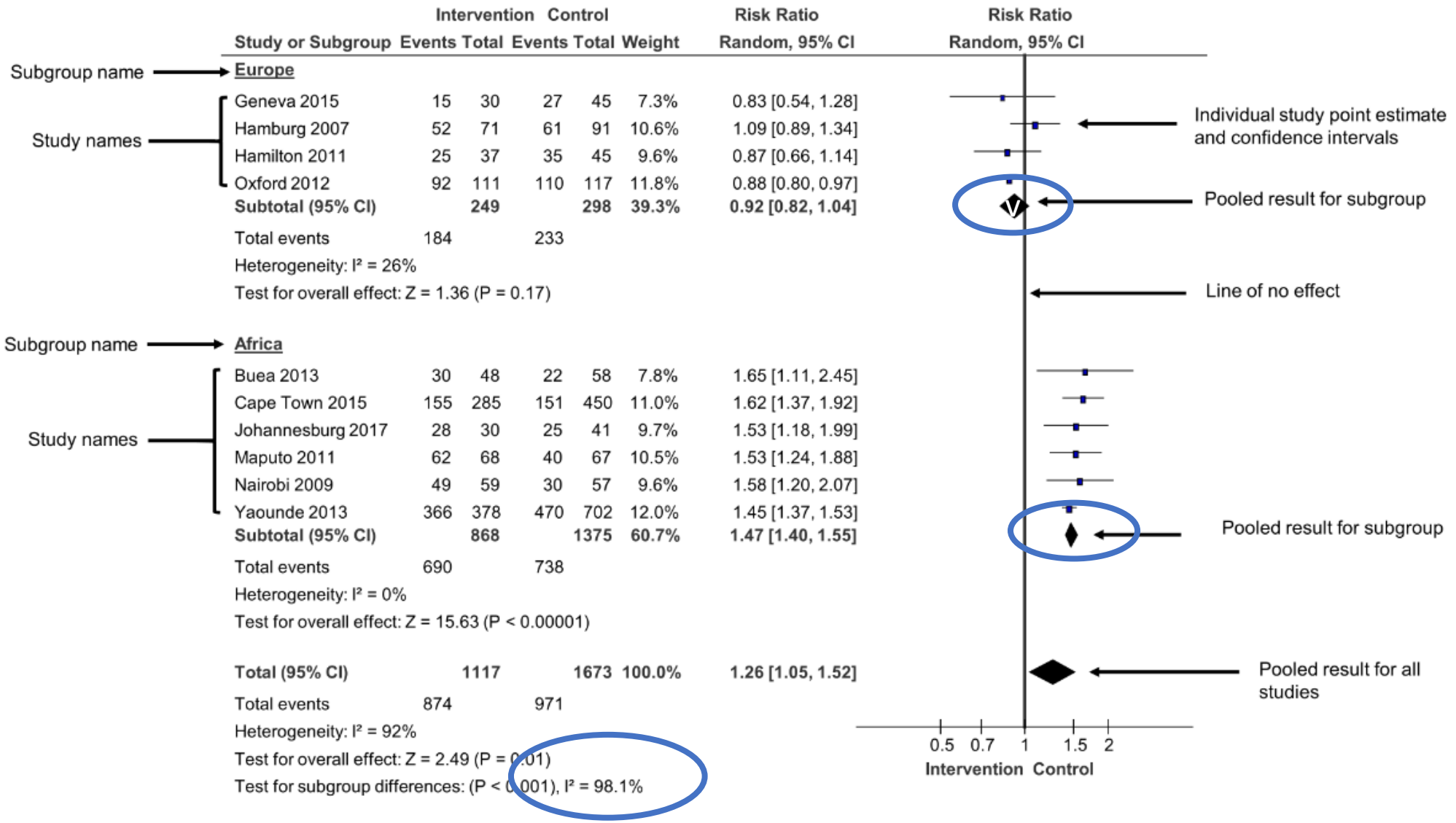
Bias in selection of the reported result — Selective reporting of results in a way that depends on the findings and prevents the estimate from being included in a meta-analysis (or other synthesis)

How do we determine levels of certainty?

- Non-randomised studies with comparators may start as HIGH certainty when ROBINS-I has been used
- Levels can be also be UP-graded on the following three factors

Large effect Size	Dose-response gradient	Plausible confounding would change the effect

- We consider DOWN-grading for *unexplained* heterogeneity
  - Large variation in effect sizes
  - Statistical tests for heterogeneity
  - No plausible explanations
- Possible explanations for inconsistency
  - Population
  - Intervention
  - Outcomes
  - Methods

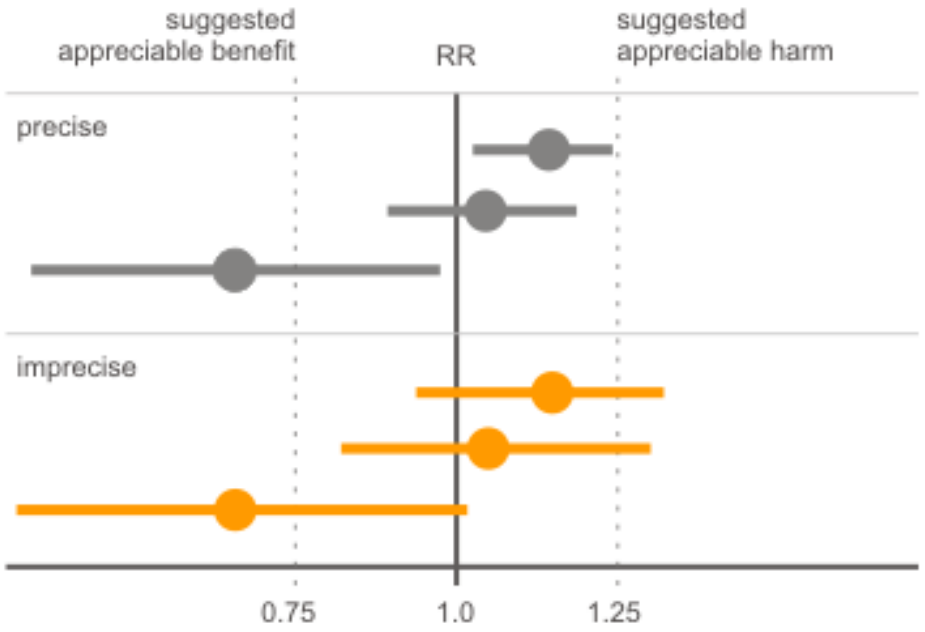


Siegfried & Mbuagbaw https://academic.oup.com/book/36249/chapter-abstract/316163832?redirectedFrom=fulltext

Risk of bias	Inconsistency of results	Indirectness of evidence	Imprecision of results	Other considerations

- We are interested in head-to-head comparisons
- Do the studies assess the PICO?
  - Population
  - Intervention
  - Comparisons
  - Outcomes
- Are there sufficient similarities in the indirect data to inform the recommendation?

- Small sample sizes and low event rates can drive imprecision
- Wide confidence intervals which include appreciable benefit or harm and cross the line of no effect



- Publication bias should always be suspected
  - Small, negative or inconclusive results not published
  - Less important in the era of trial registration
- For profit interest
- Selective outcome reporting bias

✓ Let’s GRADE together

2019 WHO Guidelines on HIV self-testing

Systematic review and meta-analysis

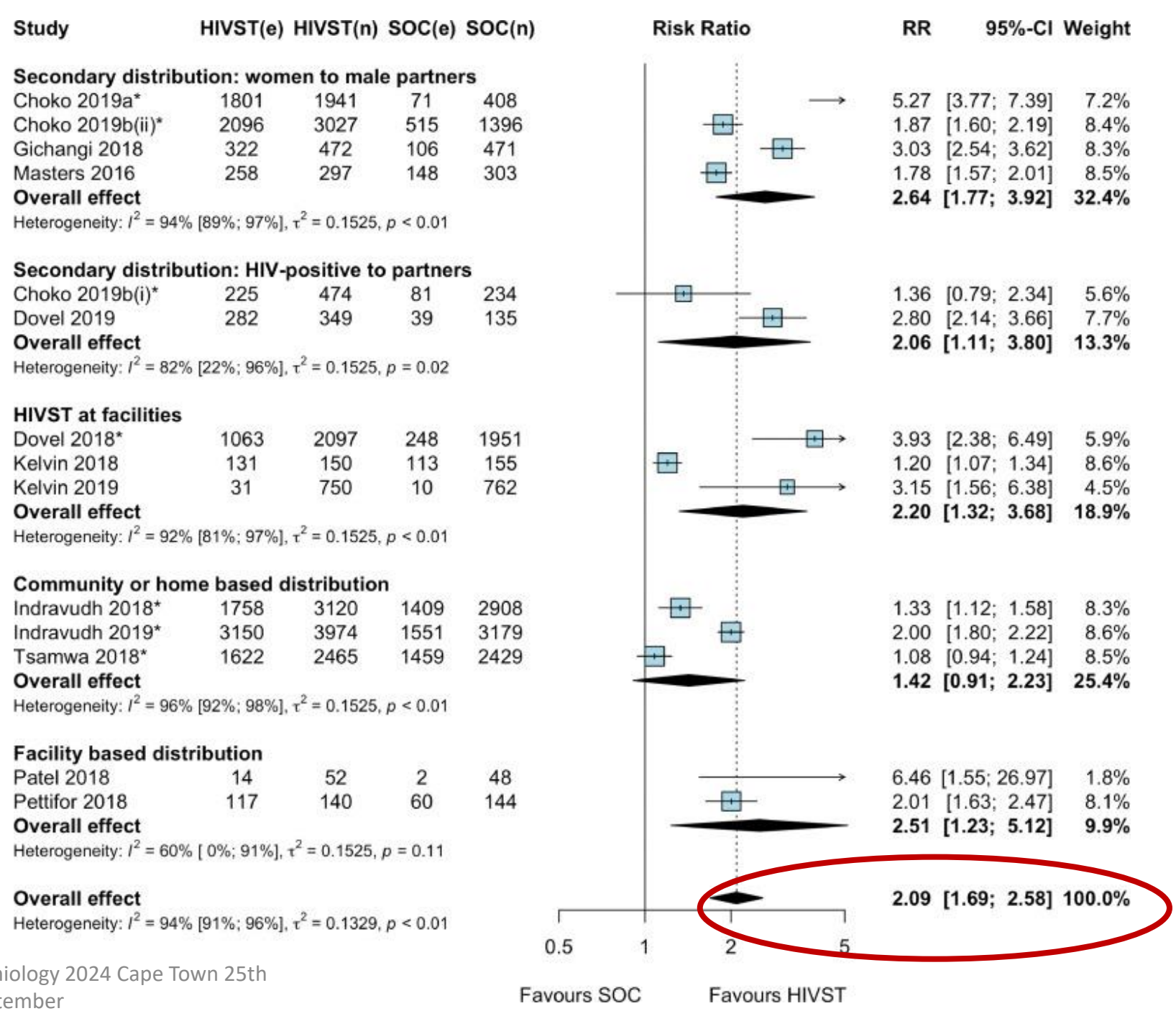
Jamil MS, Eshun-Wilson I, Witzel TC, Siegfried N, Figueroa C, Chitembo L, Msimanga-Radebe B, Pasha MS, Hatzold K, Corbett E, Barr-DiChiara M, Rodger AJ, Weatherburn P, Geng E, Baggaley R, Johnson C. Examining the effects of HIV self-testing compared to standard HIV testing services in the general population: A systematic review and meta-analysis. EClinicalMedicine. 2021 Jul 7;38:100991. doi: 10.1016/j.eclinm.2021.100991. PMID: 34278282; PMCID: PMC8271120.

WHO Guidelines

https://www.who.int/publications/i/item/978-92-4-155058-1

Consolidated guidelines on HIV testing services, 2019. Web Annex B. GRADE table: should HIV self-testing be offered as an additional HIV testing approach?

EFFECT ESTIMATE



ADD TITLE with ? HTS written out… wasn't sure what the outcome is?

Table B1. GRADE table for HIVST compared to SOC for HTS

Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]								**RR 2.09** (1.69 to 2.58)	**436 more per 1,000** (from 276 more to 632 more)		CRITICAL

RISK OF BIAS

Uptake of HIV testing

Study	Random sequence generation (selection bias)	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting (reporting bias)	Other bias	Recruitment bias	Cluster imbalance	Loss of cluster	Incorrect analysis	Overall risk of bias
	ROB domains							cROB domains				
Choko 2019a	+	+	−	−	+	+	+	−	+	+	+	−
Choko 2019b	+	+	−	−	?	+	+	−	+	?	?	−
Dovel 2018	+	?	−	−	+	?	+	−	?	+	+	−
Dovel 2019	?	?	−	−	+	+	?					−
Gichangi 2018	+	+	−	−	+	+	+					−
Indravudh 2018	+	+	−	−	+	+	+	+	?	+	+	−
Indravudh 2019	+	+	−	−	?	+	+	+	?	+	+	−
Kelvin 2018	?	+	−	+	+	+	+					−
Kelvin 2019	?	+	−	?	+	+	+					−
Masters 2016	+	+	−	−	+	+	+					−
Patel 2018	?	+	−	−	−	+	+					−
Pettifor 2018	?	?	−	?	+	?	+					−
Tsanwa 2018	+	?	−	−	+	+	?	+	+	+	+	−

i. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 10 trials and attrition bias in one trial (Patel, 2018: 36% LTFU overall, 44% in the intervention and 27% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. 9 of 13 trials had more than three high risk or unclear risk of bias domains.

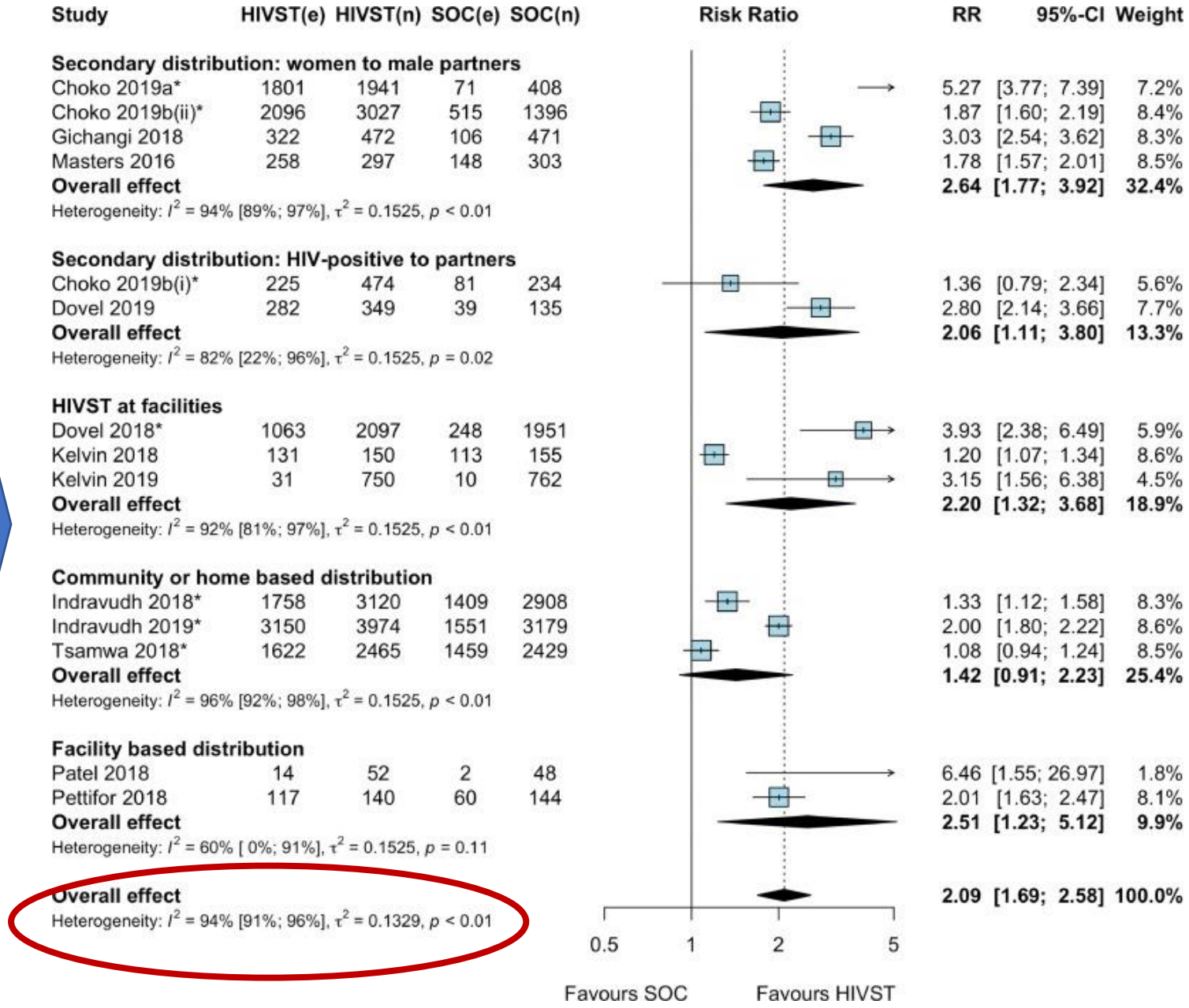
Table B1. GRADE table for HIVST compared to SOC for HTS

Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]	very serious [i]							**RR 2.09** (1.69 to 2.58)		CRITICAL	

VERY SERIOUS RISK OF BIAS

INCONSISTENCY



j. There was a high statistical heterogeneity (Heterogeneity: Tau² = 0.133; Chi² = 213.31, df = 13, p < 0.01; I² = 94%, 91% - 96%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary We did not downgrade.

Table B1. GRADE table for HIVST compared to SOC for HTS

Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]	very serious [i]	not serious [j]						**RR 2.09** (1.69 to 2.58)		CRITICAL	

INCONSISTENCY WAS NOT SERIOUS

INDIRECTENESS

all but one trial were conducted in Africa (6 in Malawi, 4 in Kenya, one in Zambia, one in South Africa, one in the US).

Table B1. GRADE table for HIVST compared to SOC for HTS

Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]	very serious [i]	not serious [j]	not serious [k]					**RR 2.09** (1.69 to 2.58)		CRITICAL	

INDIRECTNESS WAS NOT SERIOUS

IMPRECISION

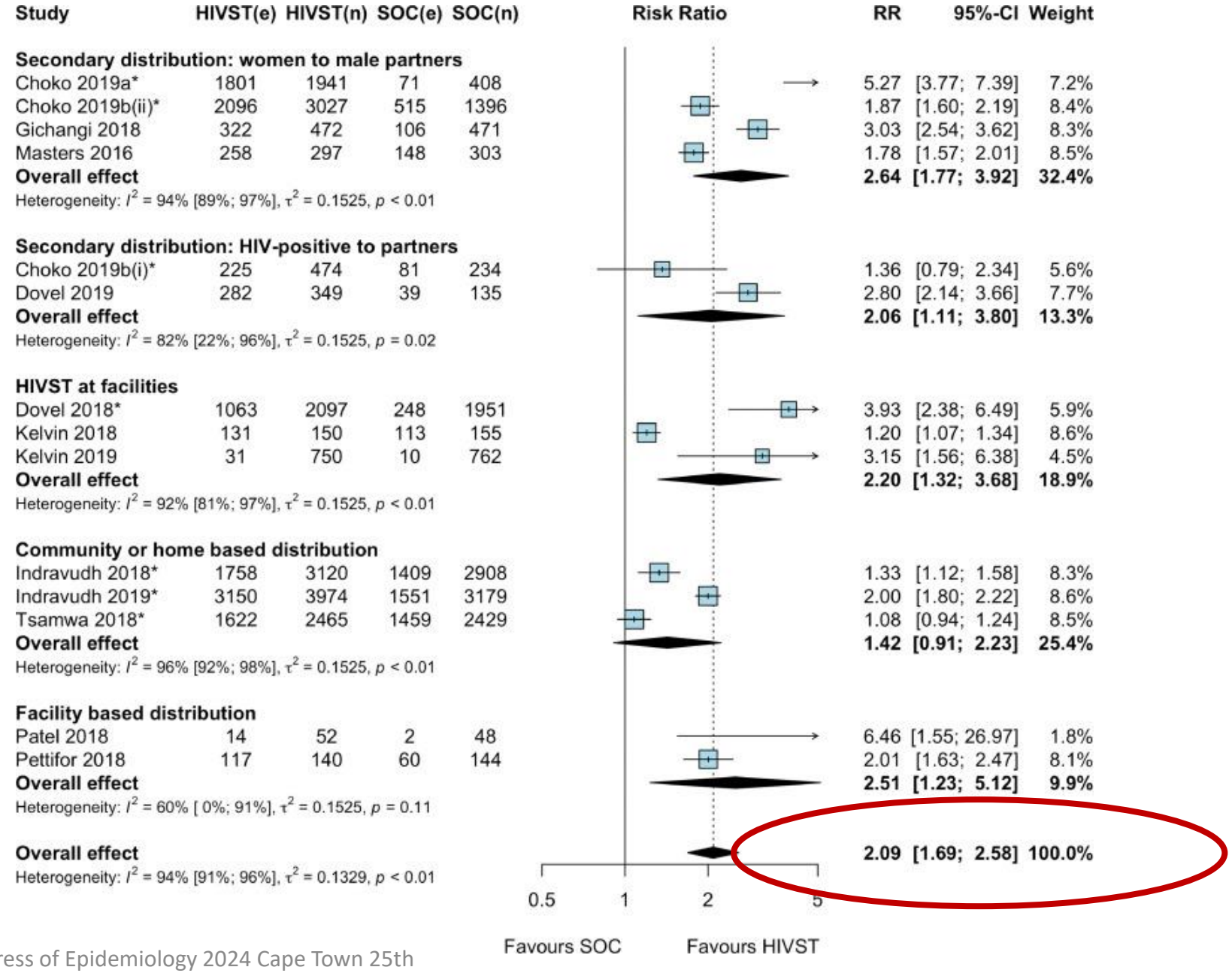


Table B1. GRADE table for HIVST compared to SOC for HTS

Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]	very serious [i]	not serious [j]	not serious [k]	not serious	none	12870/19308 (66.7%) [l]	5812/14523 (40.0%) [l]	**RR 2.09** (1.69 to 2.58)	**436 more per 1,000** (from 276 more to 632 more)		CRITICAL

IMPRECISION WAS NOT SERIOUS

Table B1. GRADE table for HIVST compared to SOC for HTS

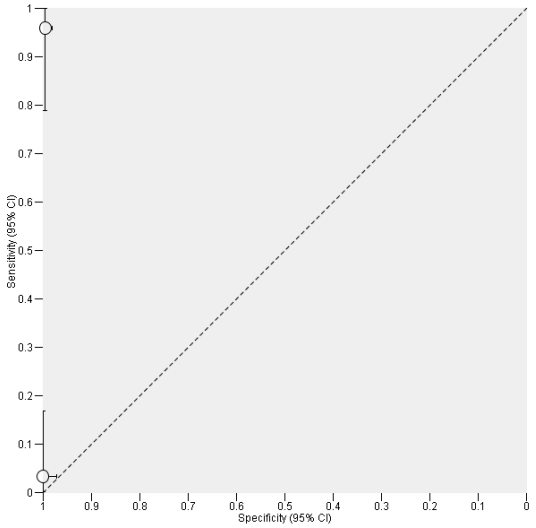
Author(s): Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson
Question: HIVST compared to SOC for HTS

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	HIVST	SOC	Relative (95% CI)	Absolute (95% CI)		
Uptake of HIV testing (general population)												
13 [g]	randomised trials [h]	very serious [i]	not serious [j]	not serious [k]	not serious	none	12870/19308 (66.7%) [l]	5812/14523 (40.0%) [l]	**RR 2.09** (1.69 to 2.58)	**436 more per 1,000** (from 276 more to 632 more)	⊕⊕◯◯ LOW	CRITICAL

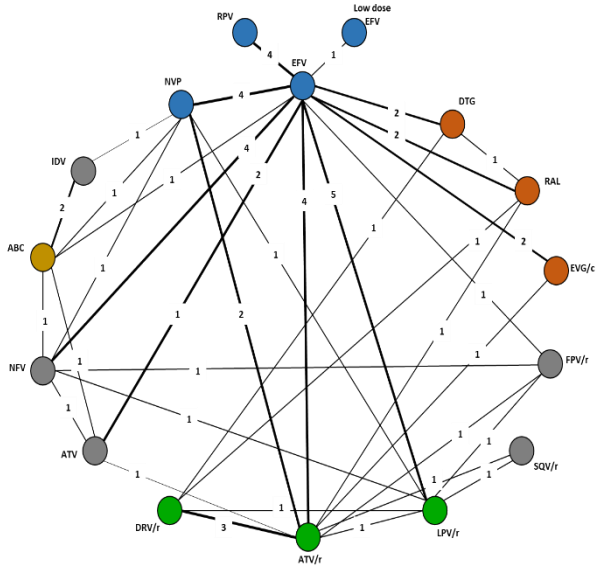
LOW CERTAINTY EVIDENCE

GRADE extensions

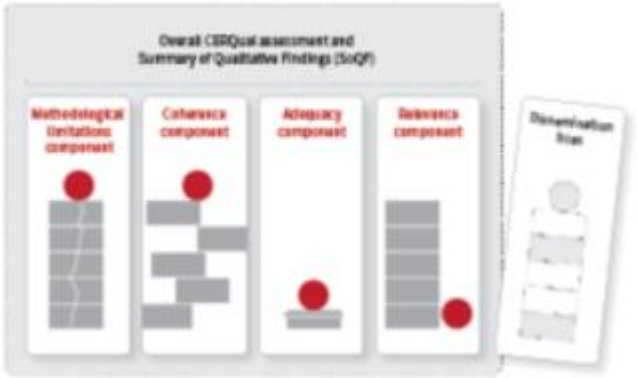
Diagnostic Test Accuracy Studies



Network Meta-analysis



GRADE CERQual



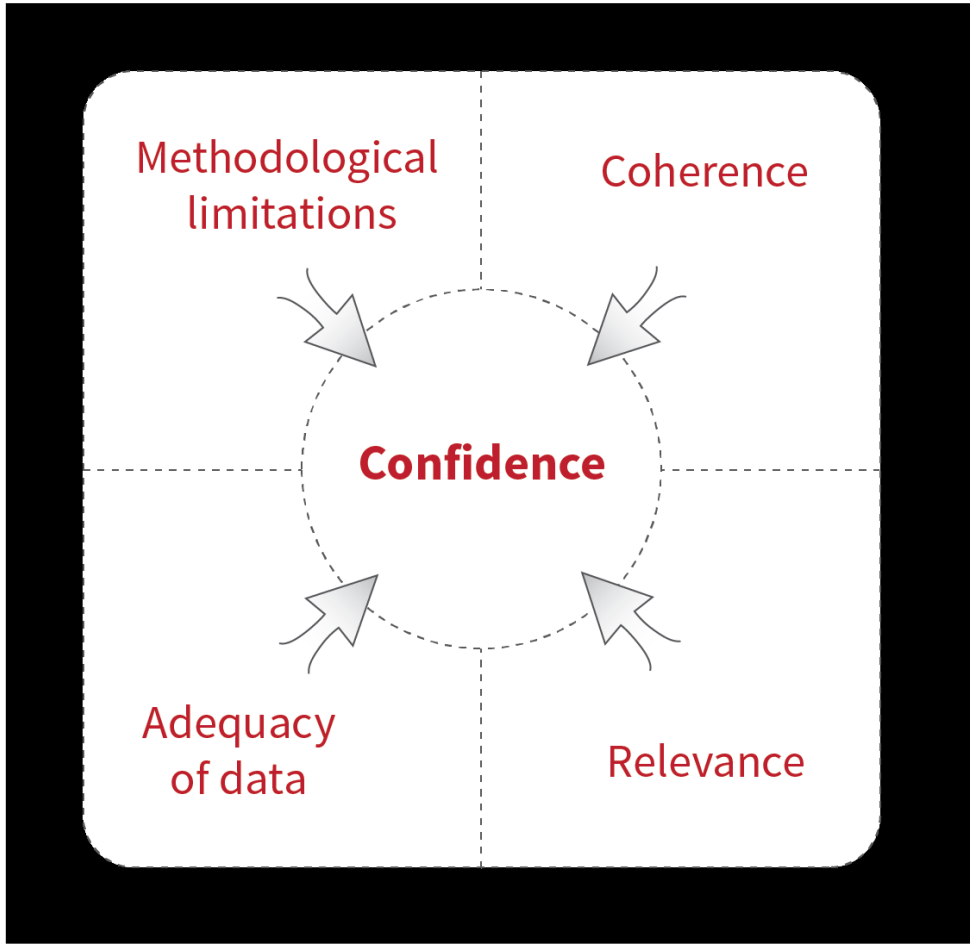
- ✓ Qualitative evidence and GRADE CERQual

What is qualitative evidence?

- ✓ Qualitative research aims to describe the social world; understand people's views, experiences and motivations; and often to explain the social world by developing hypotheses, theories or models
- ✓ Common methods for qualitative research:
  - Focus groups
  - Individual, semi-structured interviews
  - (Participant) observation
  - Document analysis
- ✓ **GRADE CERQual** synthesizes evidence from qualitative studies

GRADE CERQual Approach

- ✓ GRADE-CERQual aims to transparently assess and describe how much confidence to place in findings from qualitative evidence syntheses
- ✓ Confidence can be high, moderate, low or very low based on four criteria



Interpretation of CERQual ratings

Rating	Interpretation
High	It is **highly likely** that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is **likely** that the review finding is a reasonable representation of the phenomenon of interest
Low	It is **possible** that the review finding is a reasonable representation of the phenomenon of interest
Very Low	It is **not clear** whether the review finding is a reasonable representation of the phenomenon of interest

https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001895

- ✓ Formulating recommendations with GRADE Evidence-to-Decision-making

HOW

GRADE Evidence to Decision-making

- Balance of benefits versus harms
- Certainty of evidence
- Values & Preferences (cultural, social, special populations)
- Resource use
- Feasibility
- Equity & Human Rights
- Acceptability

Systematic reviews

What does quantitative evidence inform?

- ✓ **Effectiveness of an intervention**
  - Whether a tested intervention is effective, is neither effective nor ineffective, or results in harms
- ✓ **Certainty of the evidence**
  - The extent to which we are confident in results arising from studies where an intervention is tested. Certainty can be classified as high, moderate, low or very low
- ✓ **Costs and Cost-effectiveness**
  - What is the cost of an intervention and the delivery thereof? Cost-effectiveness is a relative concept and compares interventions by estimating how much it costs to gain a unit of health outcome

What does qualitative evidence inform?

- ✓ **How people value the outcomes**
  - Differences in, or uncertainties about how stakeholders value the outcomes
- ✓ **Acceptability of the intervention**
  - The extent to which an intervention is considered to be reasonable, satisfactory or adequate to relevant stakeholders
- ✓ **Feasibility of the intervention**
  - Extent to which an intervention is capable of being accomplished or implemented
- ✓ **Gender, equity and Human Rights**
  - Which certain groups are likely to benefit more or less than others from the intervention in ways that could be corrected

GRADE Evidence to Decision-making Table

- ✓ Summarises the criteria that determine the direction and strength for each recommendation
- ✓ Transparent and provides rationale for decision
- ✓ Factors are not weighted
- ✓ Process identifies where uncertainty or variability is present
- ✓ Use **GRADE ETD** for clinical and public health recommendations and **WHO INTEGRATE** for complex, public health, multi-systems recommendations

GRADE Evidence to Decision Table

GRADE Domain	Judgement	
Benefits vs Harms	Benefits > Harms; Benefits = Harms; Harms > Benefits	
Certainty of Evidence	High, Moderate, Low, Very Low	
Values & Preferences	No Major Variability OR Major Variability	
Resource use	More or Less resources required	
Feasibility	Yes OR No OR Uncertain	
Equity & Human Rights	Does it contribute to realization of human rights	
Acceptability	No Major Variability OR Major Variability	
RECOMMENDATION	In favour or Against or No Recommendation	
Strength	**STRONG** OR **CONDITIONAL**	
Research Gaps		

Structure of a recommendation

- A statement addressing the elements of the PICO
- A grade of the strength of the recommendation
- A rating of the certainty of the supporting evidence

STRENGTH

Strength of a recommendation

The extent to which the GDG is confident that the desirable effects of an intervention outweigh the undesirable effects

STRONG GDG is *confident* that the desirable effects of adherence to the recommendation outweigh the undesirable effects (or vice versa)

CONDITIONAL GDG concludes that the desirable effects of adherence to the recommendation *probably* outweigh the undesirable effects (or vice versa), but is not confident

Factors to determine strength

- The higher the certainty of evidence the more likely a strong recommendation
- The lower the certainty of evidence the more likely a conditional recommendation
- The greater the variability or uncertainty in values and preferences, acceptability, feasibility and costs, the more likely a conditional recommendation is warranted

Recommendation EXAMPLES from WHO

STRONG	Self-administered injectable contraception <u>should</u> be made available as an additional approach to deliver injectable contraception for individuals of reproductive age. *(Strong recommendation; moderate certainty evidence)*
CONDITIONAL	Self-collection of samples for *Treponema pallidum* (syphilis) and *Trichomonas vaginalis* <u>may</u> be considered as an additional approach to deliver STI testing services. *(Conditional recommendation; low certainty evidence)*

Guidelines

Main messages

- GRADE Evidence Profiles summarise the certainty of evidence across selected critical outcomes for each PICO
- Evidence-to-Decision-making Tables summarise the judgments of a guideline group across essential GRADE domains transparently
- The direction and strength of recommendations are formulated by consensus

In conclusion….

- ✓ The merit of the GRADE approach is not that it ensures agreement between reasonable individuals, but the **explicitness** of the judgments being made
- ✓ **GRADE Evidence Profiles** synthesize **quantitative data** with a confidence rating to inform effectiveness, certainty and cost-effectiveness
- ✓ **GRADE CERQual** synthesize **qualitative data** with a confidence rating to inform values, acceptability, feasibility, equity and human rights
- ✓ **GRADE Evidence-to-Decision-making Tables** summarise the judgments of the guideline group across essential GRADE domains **transparently**

Useful GRADE resources and key readings

- GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. Guyatt, Gordon H. et al. Journal of Clinical Epidemiology, Volume 64, Issue 4, 380 – 382 https://www.jclinepi.com/article/S0895-4356(10)00329-X/fulltext
- Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. *GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines*. BMJ. 2016;353:i2089.
- GRADE working group website: http://www.gradeworkinggroup.org/
- Grade software for systematic reviews and guidelines: https://gradepro.org/
- GRADE Handbook: https://gdt.gradepro.org/app/handbook/handbook.html
- WHO handbook for guideline development, 2nd ed: https://apps.who.int/iris/handle/10665/145714