# Guideline development in TB diagnostics

## McGill Summer Institute, Montreal June 2017

Karen R Steingart, MD, MPH
Cochrane Infectious Diseases Group
Liverpool School of Tropical Medicine

karen.steingart@gmail.com

# Declarations of Interest

- Editor, Cochrane Infectious Diseases Group and Cochrane Diagnostic Test Accuracy Editorial Team
- Member, GRADE Working Group
- No financial disclosures

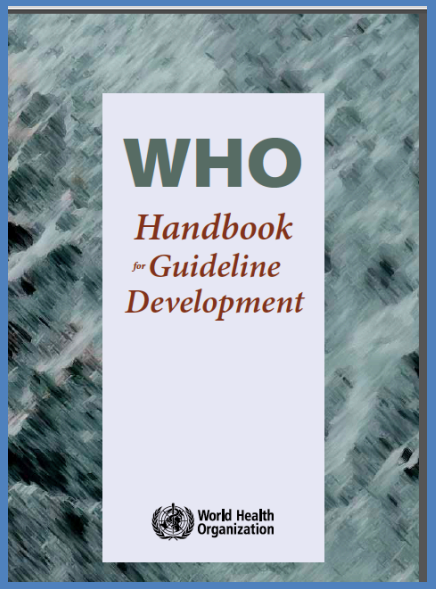
# Many factors enter into healthcare decisions

- What options are available?
- What does the evidence suggest about potential benefits and harms?
- What is the quality of the evidence?
- Might there be different options based on age, gender, comorbidities?
- Have patients' values and preferences been considered?
- Are there any social or economic considerations?

# What are guidelines?

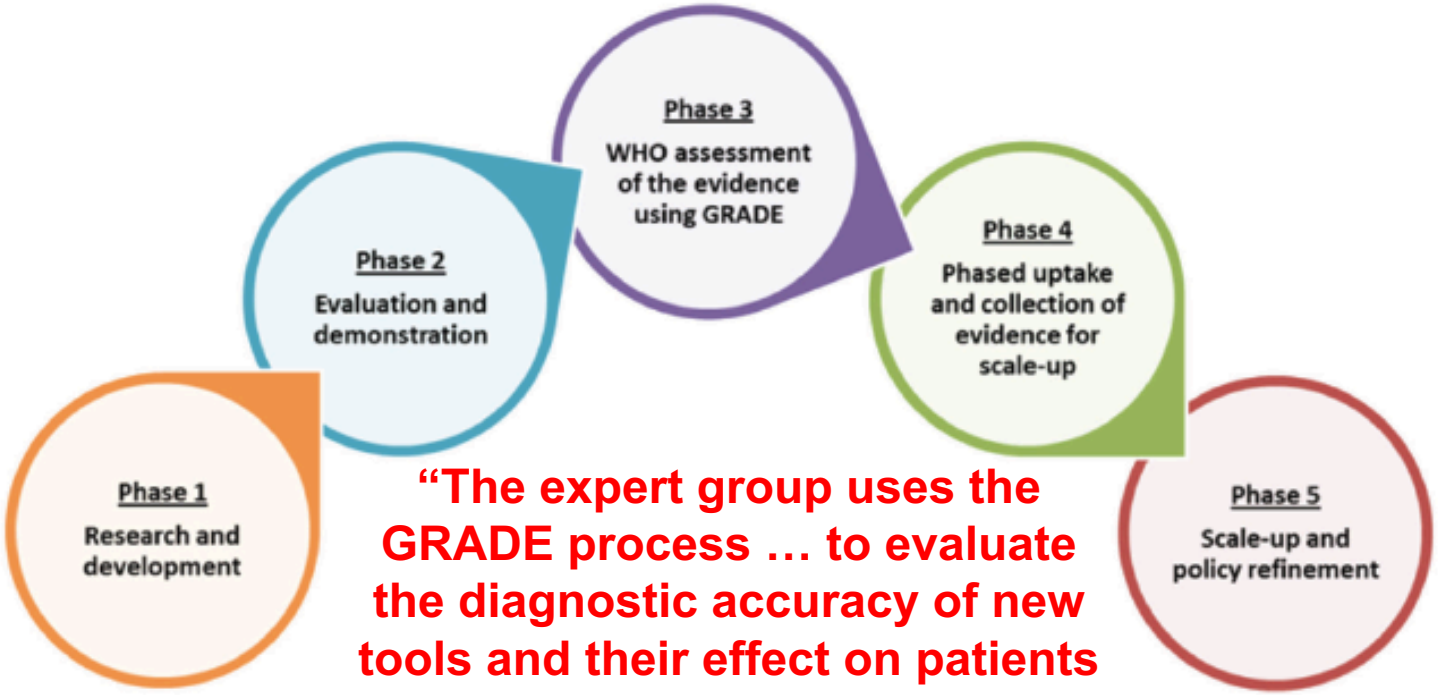
- Guidelines are recommendations intended to assist providers and recipients of health care and other stakeholders to make informed decisions.

- Recommendations may relate to clinical interventions, public health activities, or government policies.



# TB diagnostic development and assessment for WHO recommendations

Figure 1. The phases of TB diagnostics development and assessment for WHO recommendation using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) process




**"The expert group uses the GRADE process … to evaluate the diagnostic accuracy of new tools and their effect on patients and public health."**

http://www.who.int/tb/publications/implementing_TB_diagnostics/en/

# Grading of Recommendations Assessment, Development and Evaluation

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

**GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies**

The GRADE system can be used to grade the quality of evidence and strength of recommendations

Schünemann British Med J 2008



ELSEVIER

CrossMark

Journal of Clinical Epidemiology 76 (2016) 89–98

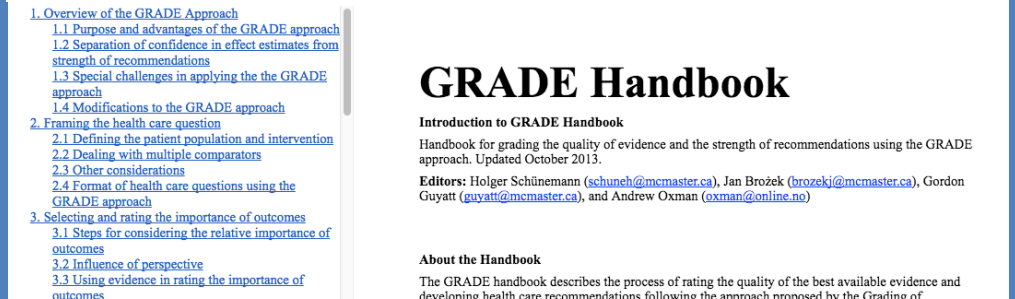
Journal of Clinical Epidemiology

SERIES: GRADING OF RECOMMENDATIONS ASSESSMENT, DEVELOPMENT AND EVALUATION (GRADE)

GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health

Holger J. Schünemann[a,b,c,*], Reem Mustafa[a,c,d], Jan Brozek[a,b,c], Nancy Santesso[a,c], Pablo Alonso-Coello[a,c,e], Gordon Guyatt[a,b,c], Rob Scholten[f], Miranda Langendam[c,g], Mariska M. Leeflang[g], Elie A. Akl[a,c,h], Jasvinder A. Singh[c,i], Joerg Meerpohl[c,j],

Schünemann J Clinical Epi 2016



1. Overview of the GRADE Approach
   1.1 Purpose and advantages of the GRADE approach
   1.2 Separation of confidence in effect estimates from strength of recommendations
   1.3 Special challenges in applying the the GRADE approach
   1.4 Modifications to the GRADE approach
2. Framing the health care question
   2.1 Defining the patient population and intervention
   2.2 Dealing with multiple comparators
   2.3 Other considerations
   2.4 Format of health care questions using the GRADE approach
3. Selecting and rating the importance of outcomes
   3.1 Steps for considering the relative importance of outcomes
   3.2 Influence of perspective
   3.3 Using evidence in rating the importance of outcomes

**GRADE Handbook**

**Introduction to GRADE Handbook**

Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013.

**Editors:** Holger Schünemann (schuneh@mcmaster.ca), Jan Brożek (brozekj@mcmaster.ca), Gordon Guyatt (guyatt@mcmaster.ca), and Andrew Oxman (oxman@online.no)

**About the Handbook**

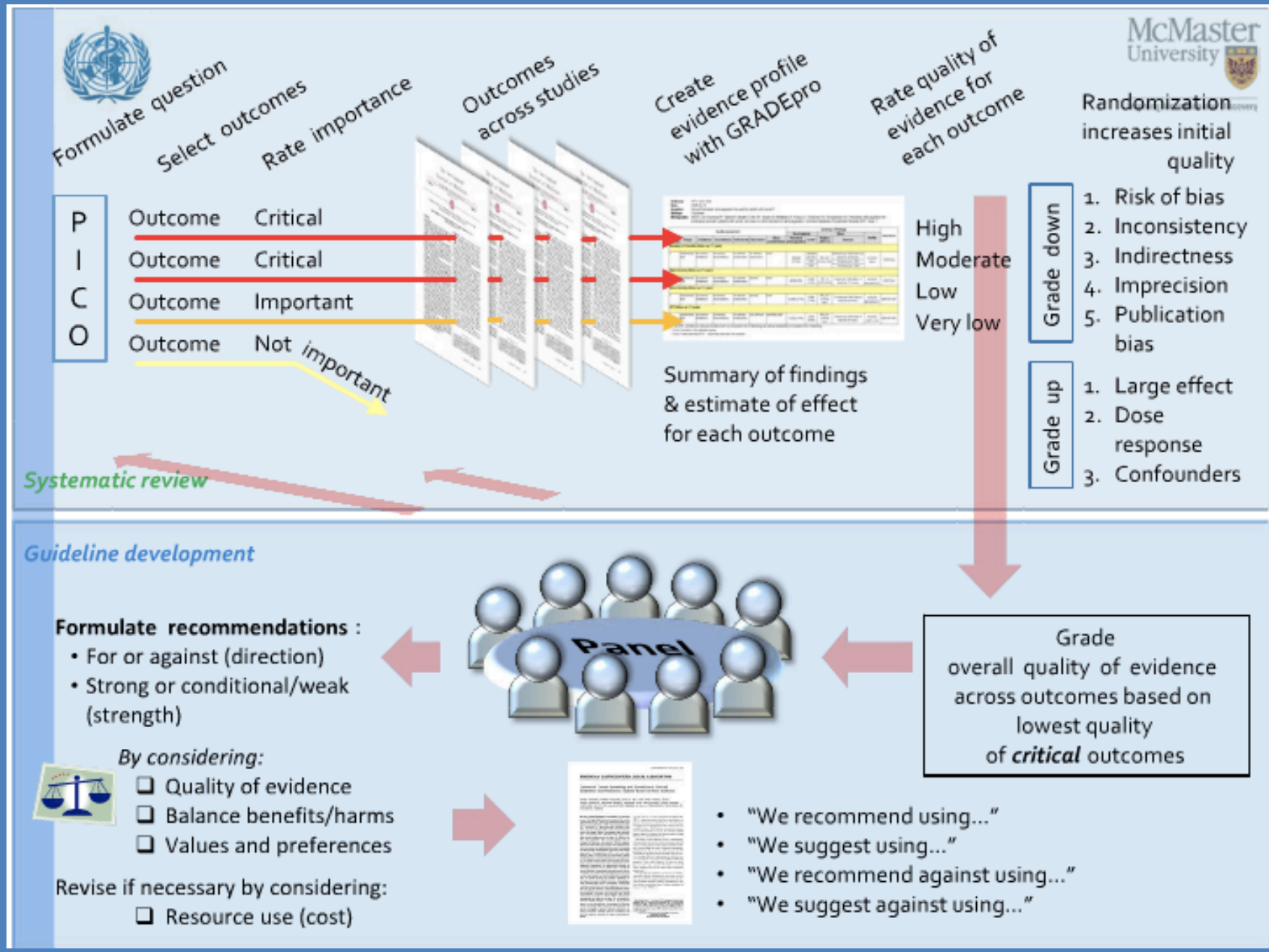
The GRADE handbook describes the process of rating the quality of the best available evidence and developing health care recommendations following the approach proposed by the Grading of

https://gradepro.org/

# GRADE

- Developed by wide group of international guideline developers
- Separates judging confidence in effect estimates (quality of evidence) and rating strength of recommendations
- Provides explicit, comprehensive criteria for downgrading and upgrading quality of evidence
- Involves a transparent process from evidence to recommendations
- Acknowledges values and preferences
- Provides a clear interpretation of strong versus weak recommendations for clinicians, patients, and policy makers

- GRADE ≠ a system for performing systematic reviews


Formulate question
Select outcomes
Rate importance
Outcomes across studies
Create evidence profile with GRADEpro
Rate quality of evidence for each outcome
McMaster University
Randomization increases initial quality
P
I
C
O
Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important
Summary of findings & estimate of effect for each outcome
High
Moderate
Low
Very low
Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias
Grade up
1. Large effect
2. Dose response
3. Confounders
Systematic review
Guideline development
Panel
Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes
Formulate recommendations :
• For or against (direction)
• Strong or conditional/weak (strength)
By considering:
❑ Quality of evidence
❑ Balance benefits/harms
❑ Values and preferences
Revise if necessary by considering:
❑ Resource use (cost)
• "We recommend using…"
• "We suggest using…"
• "We recommend against using…"
• "We suggest against using…"

# Checklist for guideline development

1. Organization, budget, planning and training
2. Priority setting
3. Guideline group membership
4. Establishing guideline group processes
5. Identifying target audience and topic selection
6. Consumer and stakeholder involvement
7. Conflict of interest considerations
8. Question generation
9. Importance of outcomes, interventions, values, preferences, utilities
10. Deciding what evidence to include and searching for evidence
11. Summarizing evidence and considering additional information
12. Judging quality of evidence
13. Developing recommendations and determining their strength
14. Wording of recommendations
15. Reporting and peer review
16. Dissemination and implementation
17. Evaluation and use
18. Updating

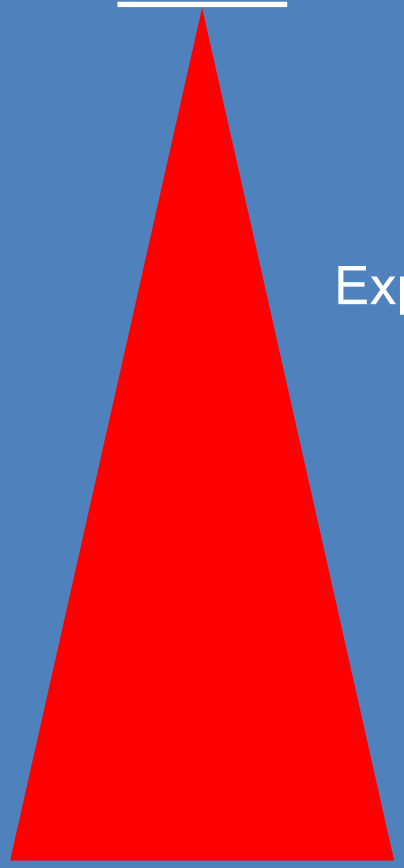
# What is Evidence?

- Evidence is the available body of facts or information indicating whether a belief or proposition is true or valid (Oxford Dictionary)
- Historically, medical training and common sense were considered sufficient for evaluating new tests and treatments
- The first RCT in the health sciences was published in 1948; demonstrated the efficacy of streptomycin for TB treatment
- “Understanding certain rules of evidence is necessary to correctly interpret literature on causation, prognosis, diagnostic tests, and treatment strategy….” EBM Working Group 1992

# Hierarchy of evidence

STUDY DESIGN

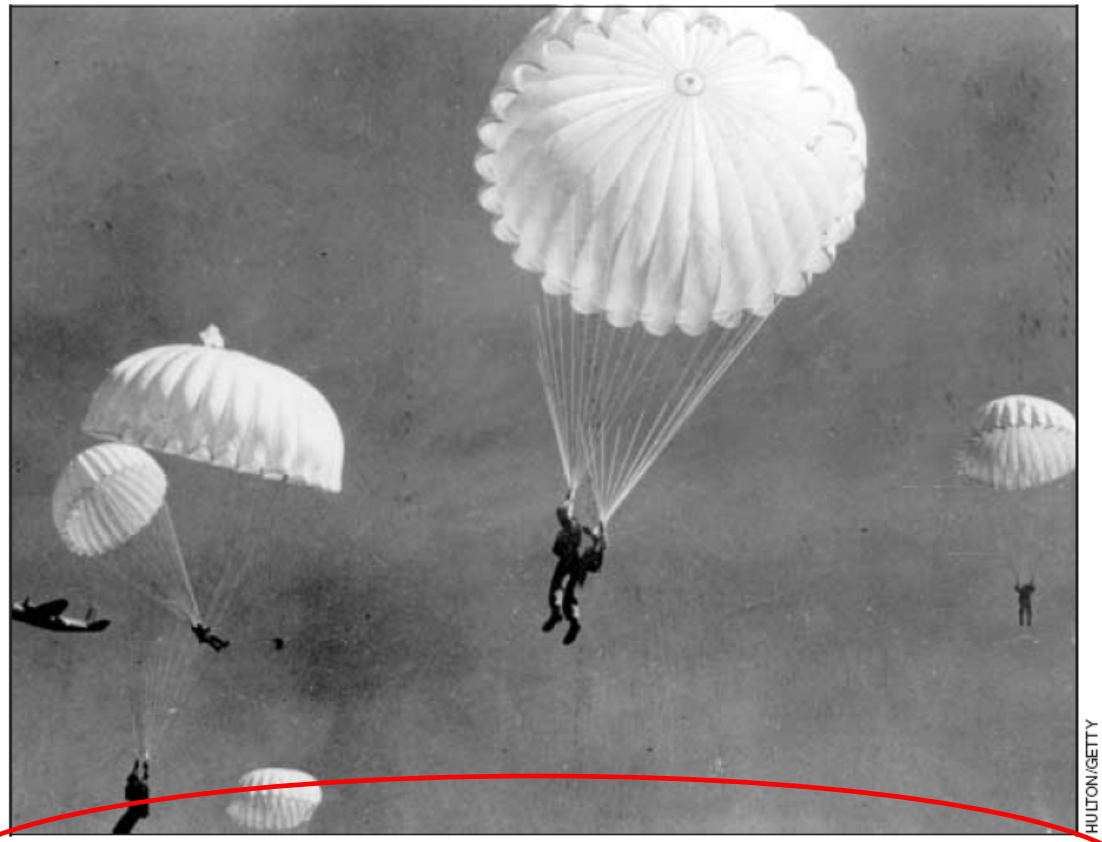
- Randomized Controlled Trials
- Cohort, Cross-Sectional, and Case-Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert opinion



# 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

BMJ VOLUME 327 20–27 DECEMBER 2003




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

## What this study adds

No randomised controlled trials of parachute use have been undertaken

---

The basis for parachute use is purely observational, and its apparent efficacy could potentially be explained by a "healthy cohort" effect

---

Individuals who insist that all interventions need to be validated by a randomised controlled trial need to come down to earth with a bump

# Parachute use and risk of death

US Parachute Association reported 21 deaths in 2015

0.001% of all jumps (1 fatality in every 165,172 skydives)

There are no RCTs - how confident are you in the certainty of the evidence?

http://www.uspa.org/

# Certainty of evidence (previously referred to as quality of evidence)

*In the context of making recommendations, certainty of evidence reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation*

Hultcrantz J Clin Epi 2017

# GRADE levels of certainty of evidence

| Level | DEFINITION |
|---|---|
| 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 effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different |
| Low ⨁⨁◯◯ | Our confidence in the effect estimate 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: The true effect is likely to be substantially different from the estimate of effect |

# What information in a systematic review may affect certainty of the evidence

- Were randomized controlled trials or diagnostic cross-sectional studies included?
- How many studies were pooled to get the summary estimate?
- How many participants were included?
- How wide was the confidence interval around the sensitivity or specificity estimate?
- Did the studies have important limitations, such as lack of blinding?

# Five domains for downgrading certainty of the evidence

| | |
|---|---|
| • Risk of bias | Were the studies unbiased? |
| • Indirectness | Does the study PICO address our question? |
| • Inconsistency | Are the study findings consistent or can we explain the inconsistency? |
| • Imprecision | Are the confidence intervals wide? |
| • Publication bias | Is the result overestimated due to publication bias? |

GRADE

- Risk of bias
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

Certainty of Evidence = HIGH


- Risk of bias

No serious risk of bias

Serious risk of bias (-1)

Very serious risk of bias (-2)

- Indirectness
- Inconsistency
- Imprecision
- Publication bias

Certainty of Evidence = HIGH


- Risk of bias — No serious risk of bias ▽
- Indirectness — No serious indirectness

Serious indirectness (-1) ▼

Very serious indirectness (-2)

- Inconsistency ▽
- Imprecision ▽
- Publication bias ▽

Certainty of Evidence = MODERATE

Footnotes

1 We downgraded one level for serious indirectness because…….

GRADE

- Risk of bias — No serious risk of bias
- Indirectness — Serious indirectness (-1) ▼
- Inconsistency — No serious inconsistency / Serious inconsistency (-1) / Very serious inconsistency (-2) ▼
- Imprecision
- Publication bias

Certainty of Evidence = VERY LOW

Footnotes

1 We downgraded one level for serious indirectness because…….

2 We downgraded two levels for very serious inconsistency because…….

# Three factors for upgrading certainty of the evidence

- Strong association
- Confounders all act to reduce observed effect
- Dose-response gradient

# Selected WHO policy statements on TB diagnostics

- Reduction of number of smears for diagnosis of pulmonary TB (2007)
- Commercial serodiagnostic tests (2011)
- Interferon-gamma release assays in low- and middle-income countries (2011)
- Xpert MTB/RIF update (2013)
- Molecular line probe assay for the detection of resistance to second-line anti-TB drugs (2016)
- **Urine lateral flow lipoarabinomannan assay for TB in people living with HIV (2015)**

http://www.who.int/tb/areas-of-work/laboratory/policy_statements/en/

# Should LF-LAM be used to diagnose TB in HIV-positive patients with CD4 ≤ 100 cells per µL?

# GRADE evidence to decision framework

- Is the problem a priority?
- How accurate is the test?
- What is the certainty of the evidence of test accuracy?
- Is there similarity in how much people value the main outcomes?
- What is the certainty of the evidence for any adverse effects of the test?
- What is the certainty of the evidence of effects of the consequences of management (including treatment) that is guided by the test results?
- How certain is the link between test results and management decisions?
- Is the incremental cost (or resource use) small relative to the benefits?
- Are there concerns about equity, acceptability and feasibility?

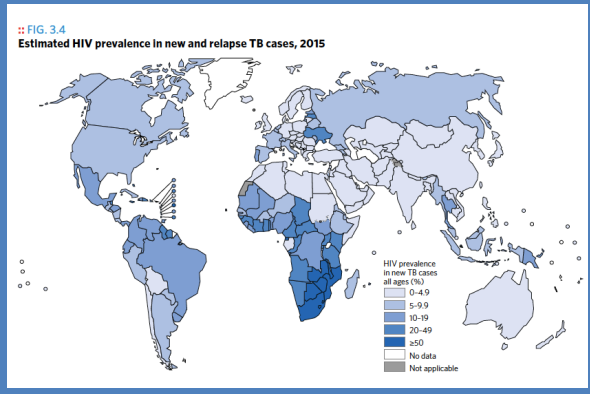
# Is the problem a priority?

- Judgement may be

No, Probably No, Uncertain, Probably Yes, Yes, Varies

- In 2013, there were 1.1 million HIV-positive people who developed TB. There were 360,000 deaths from HIV-associated TB (WHO Global Tuberculosis Report).

- *Guideline Development Group: YES*



# How accurate is the test?

# Summary of Findings Table

| **Should LF-LAM be used to diagnose tuberculosis in HIV-positive adults with CD4 ≤ 100?** |
|---|
| What is the accuracy of LF-LAM for diagnosing tuberculosis (TB) in HIV-positive adults? |
| Participants: HIV-positive adults with symptoms of TB |
| Index test: LF-LAM (grade 2) |
| Role of the test: replacement test or additional test along with sputum microscopy or Xpert® MTB/ RIF |
| Reference standard: microbiological (mainly mycobacterial culture) |
| Studies: cross-sectional |
| Limitations: use of a lower quality reference standard in most studies; small number of studies and participants included in the analyses |
| **Pooled sensitivity: 56% (95% CrI: 41, 70)**<br>**Pooled specificity: 90% (95% CrI: 81, 95)** |

# Summary of Findings Table

| Test result | Number of results per 1000 patients tested (95% CrI) | | | Number of participants (studies) | Quality of the Evidence (GRADE) |
|---|---|---|---|---|---|
| | Prevalence 1% | Prevalence 10% | Prevalence 30% | | |
| **True positives** (patients correctly classified with TB) | 6 (4 to 7) | 56 (41 to 70) | 168 (123 to 210) | 402 (5) | ⊕⊕⊕⊕ HIGH [1 2] |
| **False negatives** (patients incorrectly classified as not having TB) | 4 (6 to 3) | 44 (59 to 30) | 132 (177 to 90) | | |
| **True negatives** (patients correctly classified without TB) | 891 (802 to 941) | 810 (729 to 855) | 630 (567 to 665) | 457 (5) | ⊕⊕◯◯ LOW [3 4] |
| **False positives** (patients incorrectly classified as having TB) | 99 (188 to 49) | 90 (171 to 45) | 70 (133 to 35) | | |
| Inconclusive | Infrequent | | | | - |
| Complications | Non-invasive urine test, no known complications | | | | |

Pooled sensitivity: 56% (95% CrI: 41, 70)
Pooled specificity: 90% (95% CrI: 81, 95)

# Footnotes, Summary of Findings Table

1. QUADAS-2 to assess risk of bias. One study excluded patients who could not produce sputum. We did not downgrade the quality of evidence.
2. The wide 95% CrI for true positives and false negatives may lead to different decisions depending on which credible limits are assumed. We downgraded the quality of evidence one level.
3. QUADAS-2 to assess risk of bias. We considered three studies to be at high risk of bias because they used a lower quality reference standard. We downgraded the quality of evidence one level.
4. The wide 95% CrI for true negatives and false positives may lead to different decisions depending on which credible limits are assumed. We downgraded the quality of evidence one level.

# How accurate is the test?

- Sensitivity 56% (41,70); Specificity 90% (81,95)
- In a subgroup analysis restricted to HIV-positive **inpatients** with CD4 < 100, the sensitivity increased to 61% (48, 75); specificity decreased to 89% (75,95).

- *Guideline Development Group: Accuracy varies. While sensitivity is good, specificity is low (90%). The test detects a substantial number of false positive cases.*
- *The scenario involving HIV-positive inpatients with CD4 < 100 (sensitivity 61%) is the best possible use of the test.*

# Is there similarity in how much people value the main outcomes?

- Global Health Delivery online forum, 25 to 29 May 2015
  - Ideal test is accurate, least invasive, rapid, affordable, simple to handle, and can be used in the field
  - Should have higher sensitivity compared with specificity
  - Tests need to have acceptable levels of false positive and negative results, the benefits should outweigh the harm
  - Patients want evidence from a test before starting empiric treatment

- *Guideline Development Group: Varies for sensitivity = 56%; Probably similar for sensitivity = 61%*

# What is the certainty of the evidence of test accuracy?

| Test result | Number of results per 1000 patients tested (95% CrI) | | | Number of participants (studies) | Quality of the Evidence (GRADE) |
|---|---|---|---|---|---|
| | Prevalence 1% | Prevalence 10% | Prevalence 30% | | |
| **True positives** (patients correctly classified with TB) | 6 (4 to 7) | 56 (41 to 70) | 168 (123 to 210) | 402 (5) | ⊕⊕⊕⊕ HIGH [1 2] |
| **False negatives** (patients incorrectly classified as not having TB) | 4 (6 to 3) | 44 (59 to 30) | 132 (177 to 90) | | |
| **True negatives** (patients correctly classified without TB) | 891 (802 to 941) | 810 (729 to 855) | 630 (567 to 665) | 457 (5) | ⊕⊕◯◯ LOW [3 4] |
| **False positives** (patients incorrectly classified as having TB) | 99 (188 to 49) | 90 (171 to 45) | 70 (133 to 35) | | |
| Inconclusive | Infrequent | | | | - |
| Complications | Non-invasive urine test, no known complications | | | | |

*Guideline Development Group: Low certainty of the evidence*

# What is the certainty of the evidence for any adverse effects or burden of the test?

- Judgement may be Very Low, Low, Moderate, High
- LF-LAM is a non-invasive urine test, no known complications.
- *Guideline Development Group: High certainty of the evidence. The burden and adverse effects of the test are low.*

## What is the certainty of the evidence of effects of the consequences of management (including treatment) that is guided by the test results?

- One multi-country (South Africa, Zimbabwe, Zambia, and Tanzania) randomized controlled trial in HIV-positive inpatients (Peter. Lancet 2016).

- LF-LAM in combination with routine tests to guide the rapid initiation of TB treatment was associated with a relative risk reduction of 17% (4, 28) in eight-week mortality compared with routine diagnostic tests alone (no LF-LAM).

- *Guideline Development Group: Moderate certainty of evidence.*

# How certain is the link between test results and management decisions?

- The randomized controlled trial showed > 95% of clinicians acted on a positive test (data from inpatients).

- *Guideline Development Group: Low certainty of evidence. In many settings, clinicians use empirical treatment for TB.*

# Is the incremental cost (or resource use) small relative to the benefits?

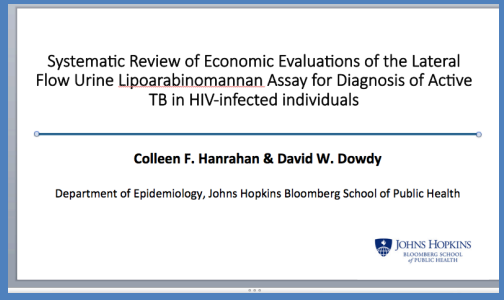
| Study Characteristics | Sun et al | Shah et al |
|---|---|---|
| **Country setting** | **South Africa**, Uganda | Uganda |
| **Clinical setting** | Inpatient | Inpatient **& Outpatient** |
| **Population** | **Hospitalized** HIV-infected adults with presumptive TB **and CD4 <100** | HIV-Infected adults with presumptive TB |
| **Analysis perspective** | **Public-sector TB program (no HIV costs)** | **Health system** |

**Hanrahan and Dowdy**

# Is the incremental cost (or resource use) small relative to the benefits?

- The incremental cost of adding LF-LAM was estimated at $21 by Sun, versus $9 by Shah (reflected different assumptions about LF-LAM specificity).

- *Guideline Development Group: Probably Yes. Both studies found LF-LAM is likely to be highly cost-effective when added to existing diagnostic algorithms. However, results may not be applicable to all HIV-positive people.*


Systematic Review of Economic Evaluations of the Lateral Flow Urine Lipoarabinomannan Assay for Diagnosis of Active TB in HIV-infected individuals

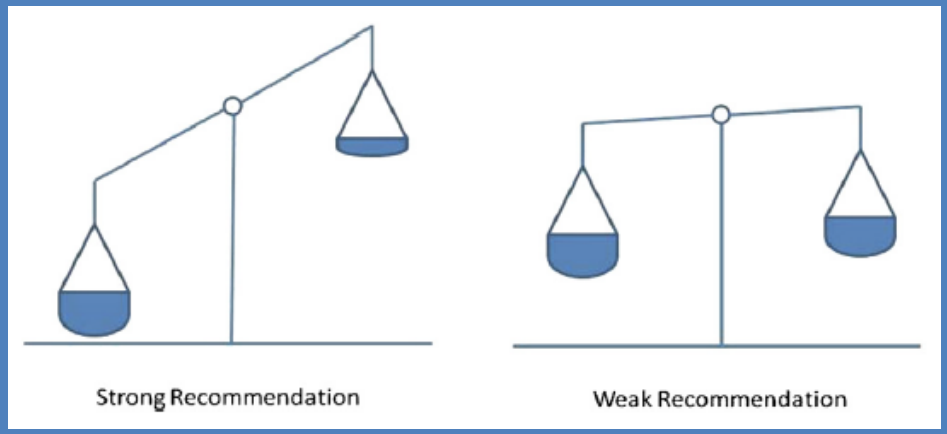
Colleen F. Hanrahan & David W. Dowdy

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

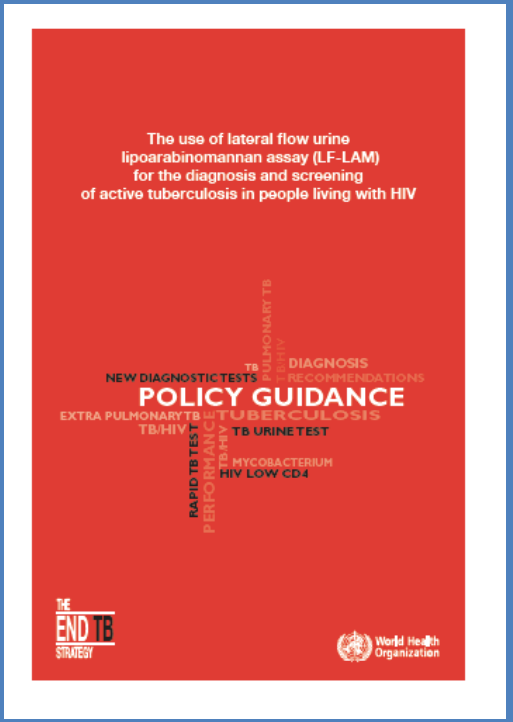
# Should LF-LAM be used for TB diagnosis in HIV-positive people?

- Yes
  - recommend for use
    - strong or weak
- No
  - recommend against use
    - strong or weak



Except as specifically described below for persons with HIV infection with low CD4 counts or who are seriously ill, LF-LAM should not be used for the diagnosis of TB (*strong recommendation, low quality of evidence*).

LF-LAM may be used to assist in the diagnosis of TB in HIV-positive adult in-patients with signs and symptoms of TB (pulmonary and/or extrapulmonary) who have a CD4 cell count less than or equal to 100 cells/μL, or HIV-positive patients who are seriously ill regardless of CD4 count or with unknown CD4 count (*conditional recommendation, low quality of evidence*).




http://www.who.int/tb/areas-of-work/laboratory/policy_statements/en/

# Implications of strong and weak recommendations

| | Strong | Weak (conditional) |
|---|---|---|
| For patients | Most individuals in this situation would want the recommended course of action and only a small proportion would not | The majority of individuals in this situation would want the suggested course of action, but many would not |
| For clinicians | Most individuals should receive the recommended course of action | Different choices will be appropriate for different patients…consistent with values/ preferences |
| For policy makers | Can be adopted as policy in most situations | Policy making will require debate and involvement of many stakeholders |

Formulate question

Select outcomes

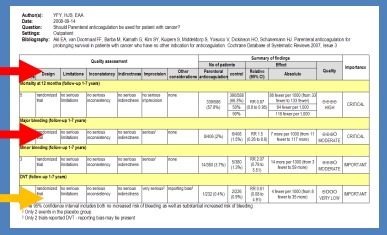
Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

| | | |
|---|---|---|
| P | Outcome | Critical |
| I | Outcome | Critical |
| C | Outcome | Important |
| O | Outcome | Not important |



Summary of findings & estimate of effect for each outcome

High
Moderate
Low
Very low

Grade down

1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

RCT start high, obs. data start low

Grade up

1. Large effect
2. Dose response
3. Confounders

*Systematic review*

*Guideline development*

**Formulate recommendations**:

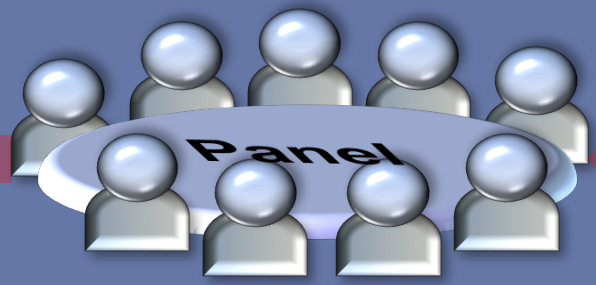
- For or against (direction)
- Strong or weak (strength)

*By considering:*

- ❑ Quality of evidence
- ❑ Balance benefits/harms
- ❑ Values and preferences

Revise if necessary by considering:

- ❑ Resource use (cost)




Rate overall quality of evidence across outcomes based on lowest quality of ***critical*** outcomes

- "We recommend using…"
- "We suggest using…"
- "We recommend against using…"
- "We suggest against using…"

Getty Image: The President's Committee on the Arts and the Humanities, May 2014.


*"That's just great. I discover the cure for the common cold and all you can do is criticize."*

# Challenges applying GRADE to diagnostic tests

- Tests by themselves usually affect outcomes indirectly rather than directly
- Downstream clinical management, guided by test results, affects patient outcomes

J Clin Epidemiol. 2016 Feb 27. pii: S0895-4356(16)00136-0. doi: 10.1016/j.jclinepi.2016.01.032. [Epub ahead of print]

**Development of the GRADE Evidence to Decision (EtD) frameworks for tests in clinical practice and public health.**

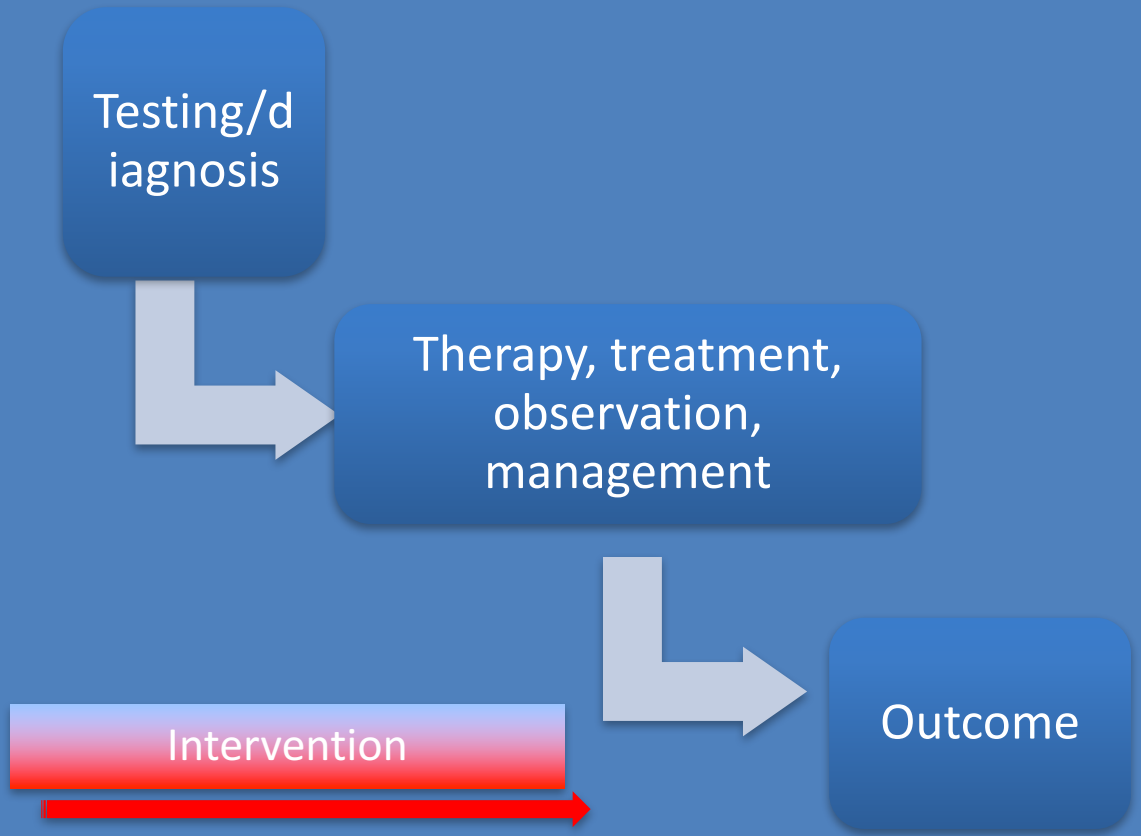
Schünemann HJ[1], Mustafa R[2], Brozek J[3], Santesso N[4], Alonso-Coello P[5], Guyatt G[3], Scholten R[6], Langendam M[7], Leeflang MM[8], Akl EA[9], Singh J[10], Meerpohl J[11], Hultcrantz M[12], Bossuyt P[8], Oxman AD[13]; GRADE Working Group.

⊕ **Author information**

**Abstract**

**OBJECTIVE:** To describe the Grading of Recommendations Assessment, Development and Evaluation (GRADE) interactive Evidence to Decision (EtD) frameworks for tests and test strategies for clinical, public health or coverage decisions.

# What evidence is needed to make deductions about impact on patient outcomes ?




Slide courtesy Holger Schünemann

# References


1. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011 Apr;64(4):401-6
2. GRADEpro GDT https://gradepro.org/
3. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, et al. (2008) Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 336: 1106-1110.
4. Schünemann HJ, Mustafa R, Brozek J, Santesso N, et al. GRADE Working Group. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. J Clin Epidemiol. 2016 Aug;76:89-98.
5. Sun D, Dorman S, Shah M, Manabe YC, Moodley VM, Nicol MP, et al. Cost utility of lateral-flow urine lipoarabinomannan for tuberculosis diagnosis in HIV-infected African adults. Int J Tuberc Lung Dis. 2013;17(4):552-8.
6. Shah M, Dowdy D, Joloba M, Ssengooba W, Manabe YC, Ellner J, et al. Cost-effectiveness of novel algorithms for rapid diagnosis of tuberculosis in HIV-infected individuals in Uganda. AIDS. 2013;27(18):2883-92.

# Acknowledgements

- Holger Schünemann

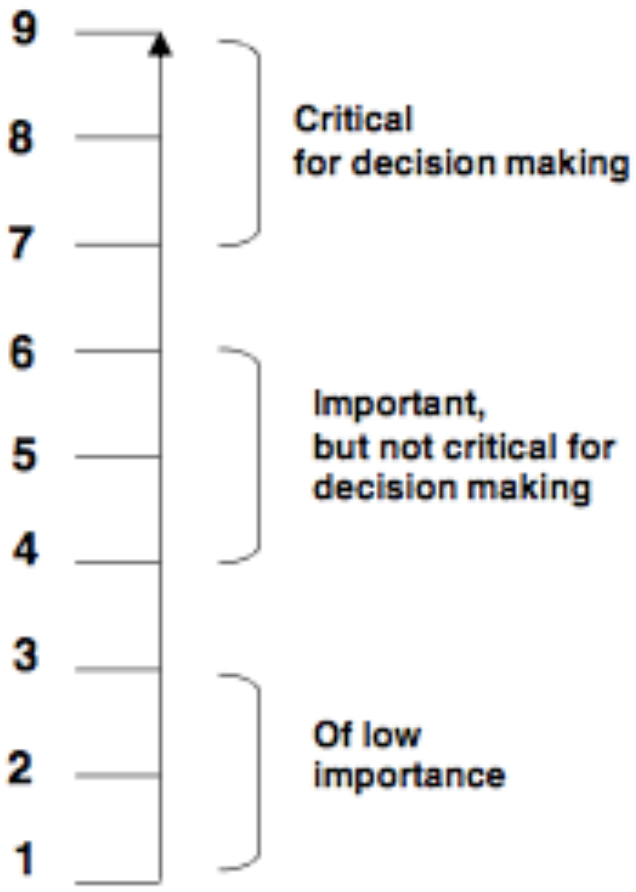
- Nancy Santesso

# Footnotes, Summary of Findings Table

1. We used QUADAS-2 to assess risk of bias. One study excluded patients who could not produce sputa. We did not downgrade.
2. The wide 95% CI for true positives and false negatives may lead to different decisions depending on which confidence limits are assumed. We downgraded one point.
3. We used QUADAS-2 to assess risk of bias. We considered three studies to be at high risk of bias for the reference standard. We downgraded one point.
4. The wide 95% CI for true negatives and false positives may lead to different decisions depending on which confidence limits are assumed. We downgraded one point.

| Domains that contribute to the strength of a recommendation | |
|---|---|
| Factor | Comment |
| Balance between desirable and undesirable outcomes (trade-offs) | The larger the differences between desirable and undesirable consequences, the more likely a strong recommendation is warranted |
| Overall quality of evidence for outcomes | The higher the quality of evidence, the more likely a strong recommendation is warranted |
| Confidence in values and preferences and their variability | The greater the variability (or uncertainty) in values and preferences, the more likely a weak recommendation is warranted |
| Resource use | The higher the cost and the more resources consumed, the less likely a strong recommendation is warranted |

# 3 categories for the relative importance of outcomes with GRADE




Journal of Clinical Epidemiology 64 (2011) 395e400