Cochrane Training

RoB 2 Domain IV
Bias in measurement of the outcome

Asbjørn Hróbjartsson[1,2] and Isabelle Boutron[1,3,4]

Bias Methods Group[1], University of Southern Denmark[2], Cochrane France[3], Université de Paris[4]

Trusted evidence.
Informed decisions.
Better health.

Outline

- Introduction
- Bias mechanisms and empirical evidence
- Assessing the risk of bias in measurement of the outcome: signalling questions 1-2
- Assessing the risk of bias in measurement of the outcome: Signalling questions 3-5
- Questions

Risk of bias in randomized trials

Bias arising from the randomization process

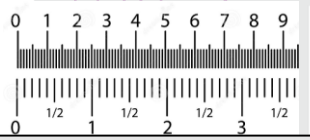
Bias due to deviations from intended intervention

Bias due to missing outcome data

Bias in measurement of the outcome

Experimental

Comparator



Outcome

Outcome

1.02 3.87
2.20 4.32
1.38 5.44

Bias in selection of the reported result

Bias mechanisms

Errors in measuring outcome variables

The measured value ≠ the true value of the outcome.

Terminology

- Measurement error (continuous outcome)
- Misclassification (dichotomous outcome, categorial outcome)
- Under/over-ascertainment (event)

- **Errors**
  - Non-differential
  - Differential

Bias mechanisms

Errors in measuring outcome variables

Non differential errors:

- Errors occurs similarly in both groups
- Errors are not related to the treatment allocated
- Example :
  - blood test to measure haemoglobin level
  - Blood pressure measurement

Example: classification by investigators/endpoint committee

Double blind randomized trials

The outcome of composites of death and myocardial infarction with or without refractory angina

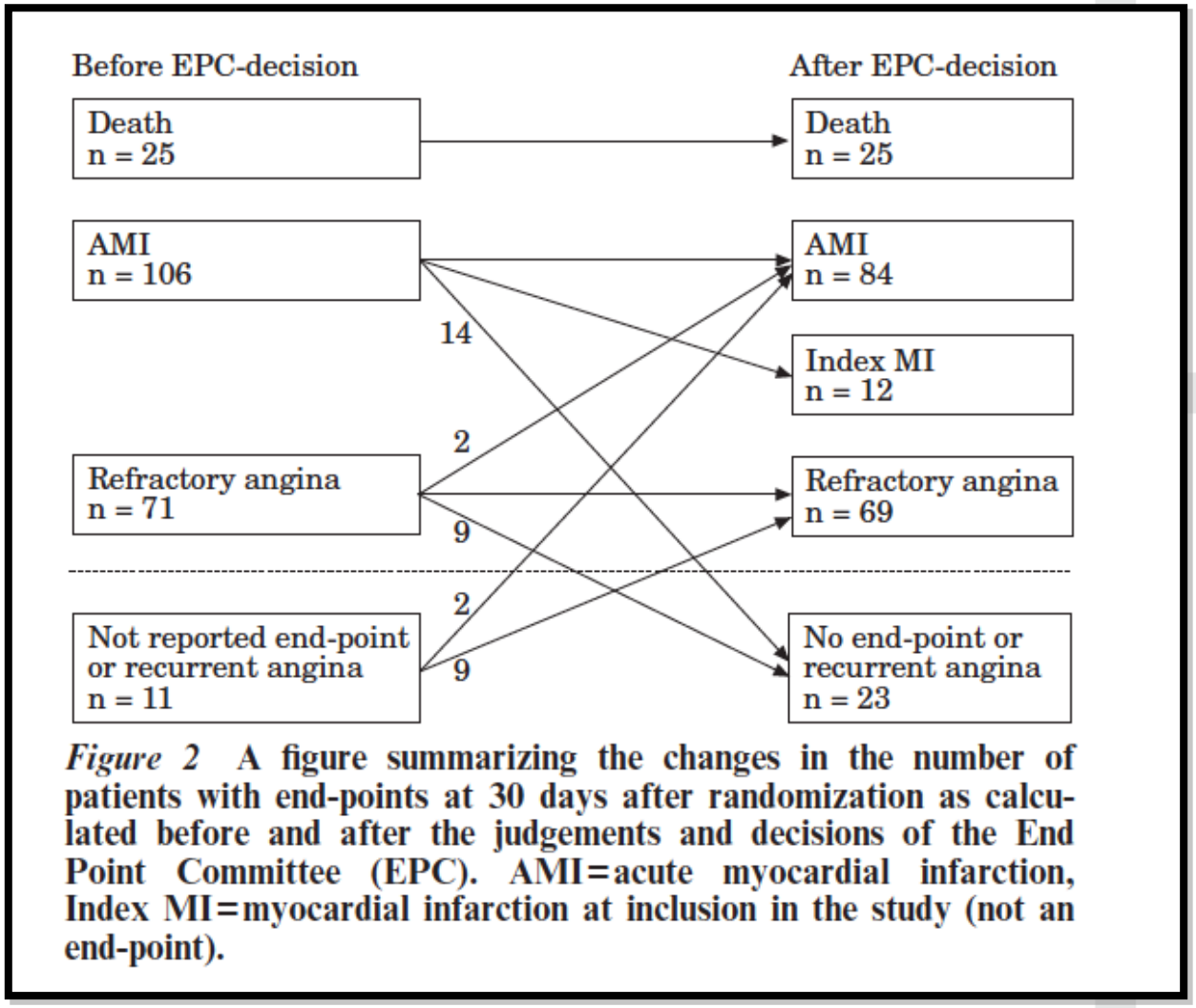


Figure 2 A figure summarizing the changes in the number of patients with end-points at 30 days after randomization as calculated before and after the judgements and decisions of the End Point Committee (EPC). AMI=acute myocardial infarction, Index MI=myocardial infarction at inclusion in the study (not an end-point).

European Heart Journal (1999) 20, 771–777

Bias mechanisms

Errors in measuring outcome variables

Non differential errors:

- **Continuous outcome (mean difference)**
  - **->** usually **no bias**

- **Dichotomous/categorial outcome (OR, RR, HR):**
  - **->** mainly **bias toward the null**

- Situations where non-differential error can bias effect estimates away from the null are **unlikely** to occur in randomized trials

Bias mechanisms

Errors in measuring outcome variables

Differential errors

- Errors are related to the treatment allocated
- Example : assessment of the pain level on a VAS systematically assessed as lower in the intervention arm

⇒Bias

⇒Essential role of blinding

Cochrane Training

Blinding terminology

- ”Single” blind
- ”Double” blind
- ”Triple” blind

	Physicians	Textbooks
– Single blind	10	5
– Double blind	17	9
– Triple blind	15	7

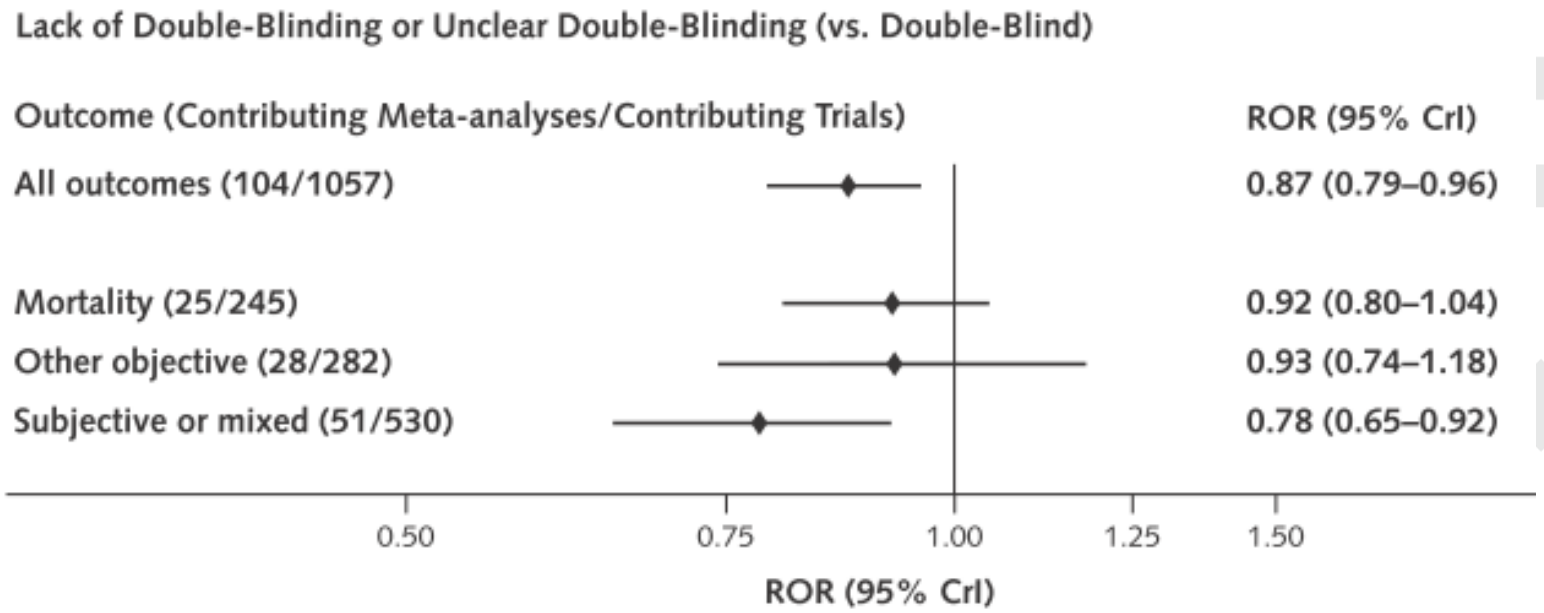
Devereaux, JAMA, 2001

Empirical evidence of bias

Annals of Internal Medicine | Research and Reporting Methods

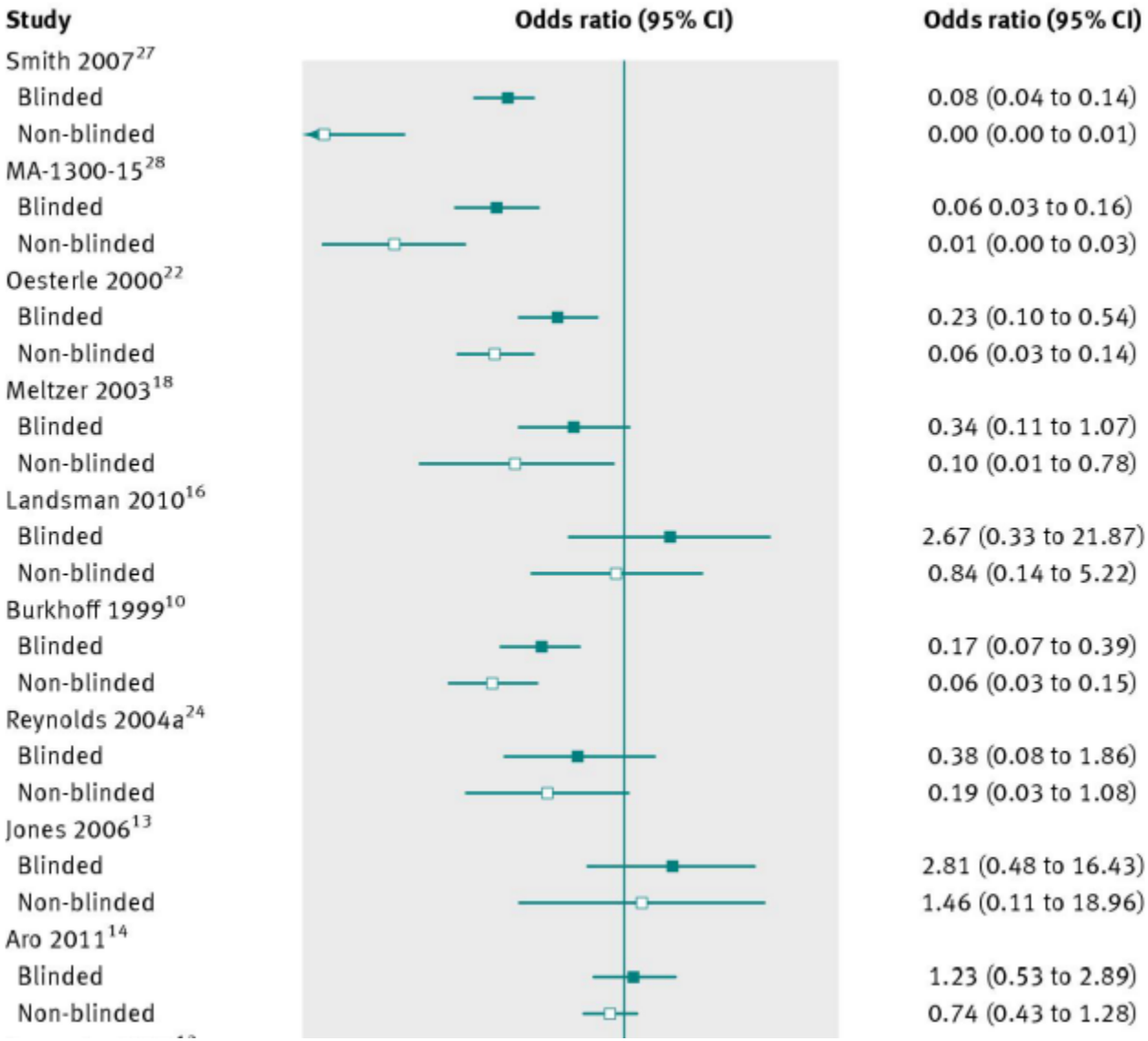
Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials

Jelena Savović, PhD; Hayley E. Jones, PhD; Douglas G. Altman, DSc; Ross J. Harris, MSc; Peter Jüni, MD; Julie Pildal, MD, PhD; Bodil Als-Nielsen, MD, PhD; Ethan M. Balk, MD, MPH; Christian Gluud, DrSciMed; Lise Lotte Gluud, DrSciMed;



Ann Intern Med. 2012;157:429-438.

Estimated intervention effect according to blinded or non-blinded outcome assessor



Hróbjartsson A et al. BMJ 2012;344:bmj.e1119

Is blinding always feasible?

Assessment of zinc treatment for common cold[1,2]

- Specific taste and aftertaste of zinc
- Hunches: « anything tasting as bad as zinc and with as much as aftertaste as zinc must be a good medicine »
- Success of blinding was questionnable

1) Desbiens et al, Annals of Internal Medicine, 2000
2) Fair, J et al.. Chronic Dis., 1987

Is blinding always feasible?

Blinding of outcome assessment

Centralized blinded assessment

- Radiography
- Video
- Audiotape
- Photography
- Blinded adjudication committee

Not always possible

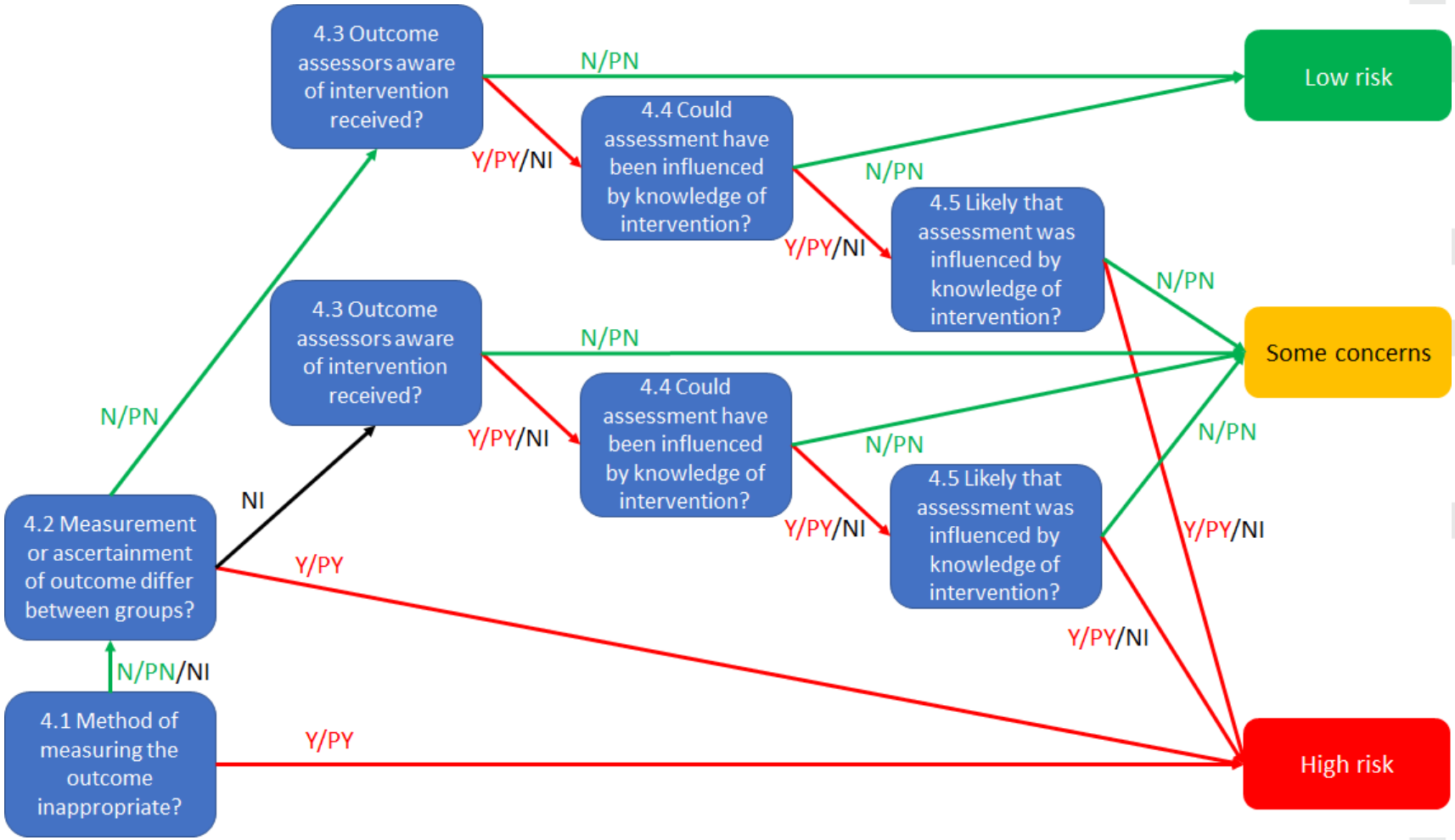
- Patient reported outcome ?

Cochrane Training

Assessing the risk of bias in measurement of the outcome

Signalling questions

4.1 Was the method of measuring the outcome inappropriate



Current version of RoB 2

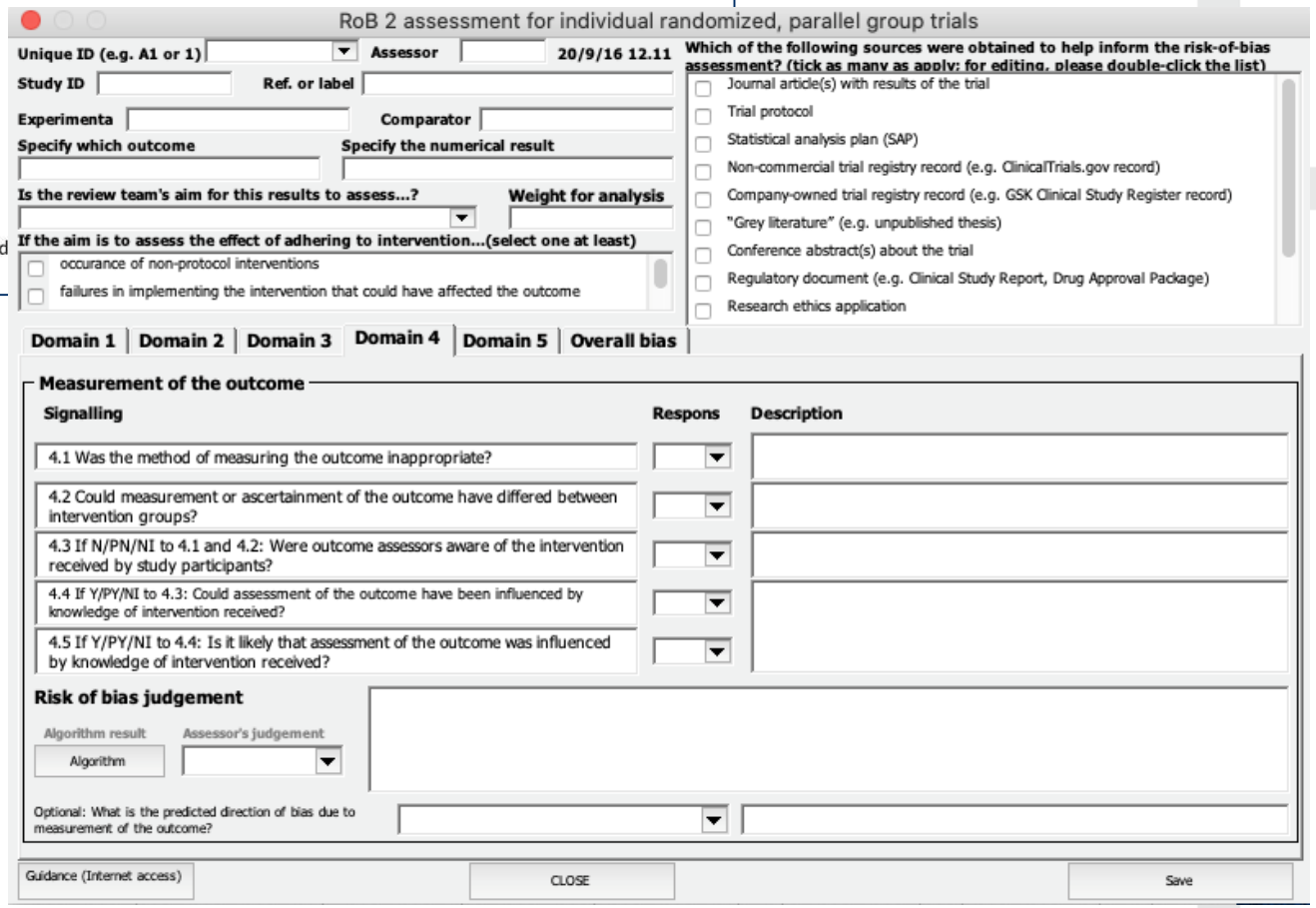
Update!

Download the 22 August 2019 version:

- The full guidance document.
- The cribsheet summarizing the tool.
- A template for completing the assessment.
- An Excel tool to implement RoB 2 (contains macros; d

To download the file

Assessment at the outcome level



Cochrane Training

Assessing the risk of bias in measurement of the outcome

Signalling questions 1-2.

4.1 Was the method of measuring the outcome inappropriate?

- Poor validity of the methods
  - The methods **does not measure** what it is intended to measure
  - Example 1:
    - Event: severe hypoglycemia
    - Measurement: portable blood glucose machine used by patients
    - Issue: does not reliably measure glycemia <3,1 mmol/l
  - Example 2: Self reported physical exercise using the International Physical Activity Questionnaire (IPAQ) (Carvalho FA, *Musculoskelet Sci Pract.* 2017)
- Poor reliability of the methods
  - Example: Four-point rating scale for assessing pain level is less reliable than VAS or numeric rating scale

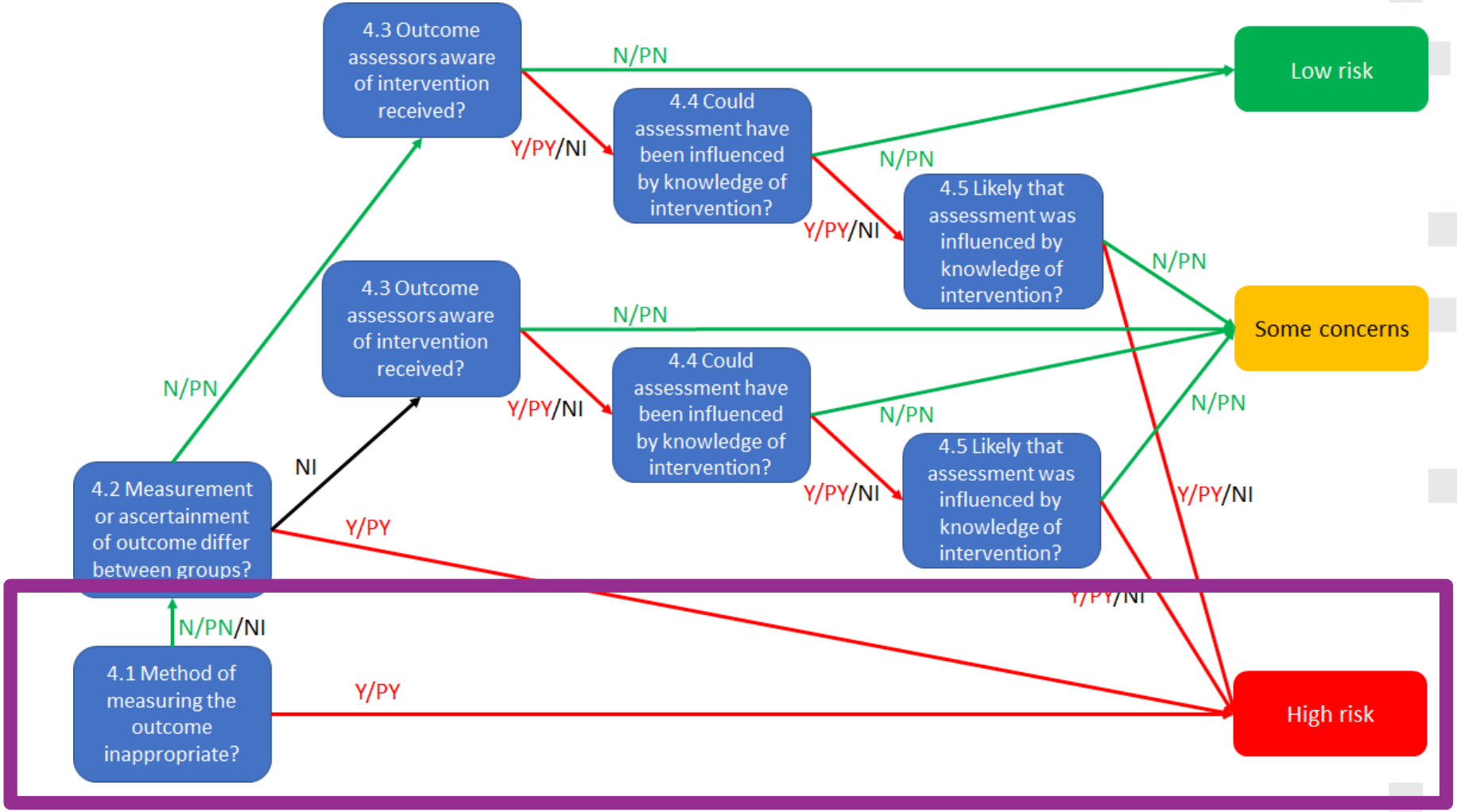
4.1 Was the method of measuring the outcome inappropriate?

The question does **NOT** aim to assess whether the choice of the outcome is relevant

- **NO/Probably NO**
  - In **most trials**, for pre-specified outcomes

- **Yes/probably yes:**
  - Measurement unlikely to identify plausible intervention effect
  - Measurement has been demonstrated to have poor validity

4.1 Was the method of measuring the outcome inappropriate



4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?

In randomized trials outcome measurement is is usually performed similarly in both group.

However, specific situations may arise

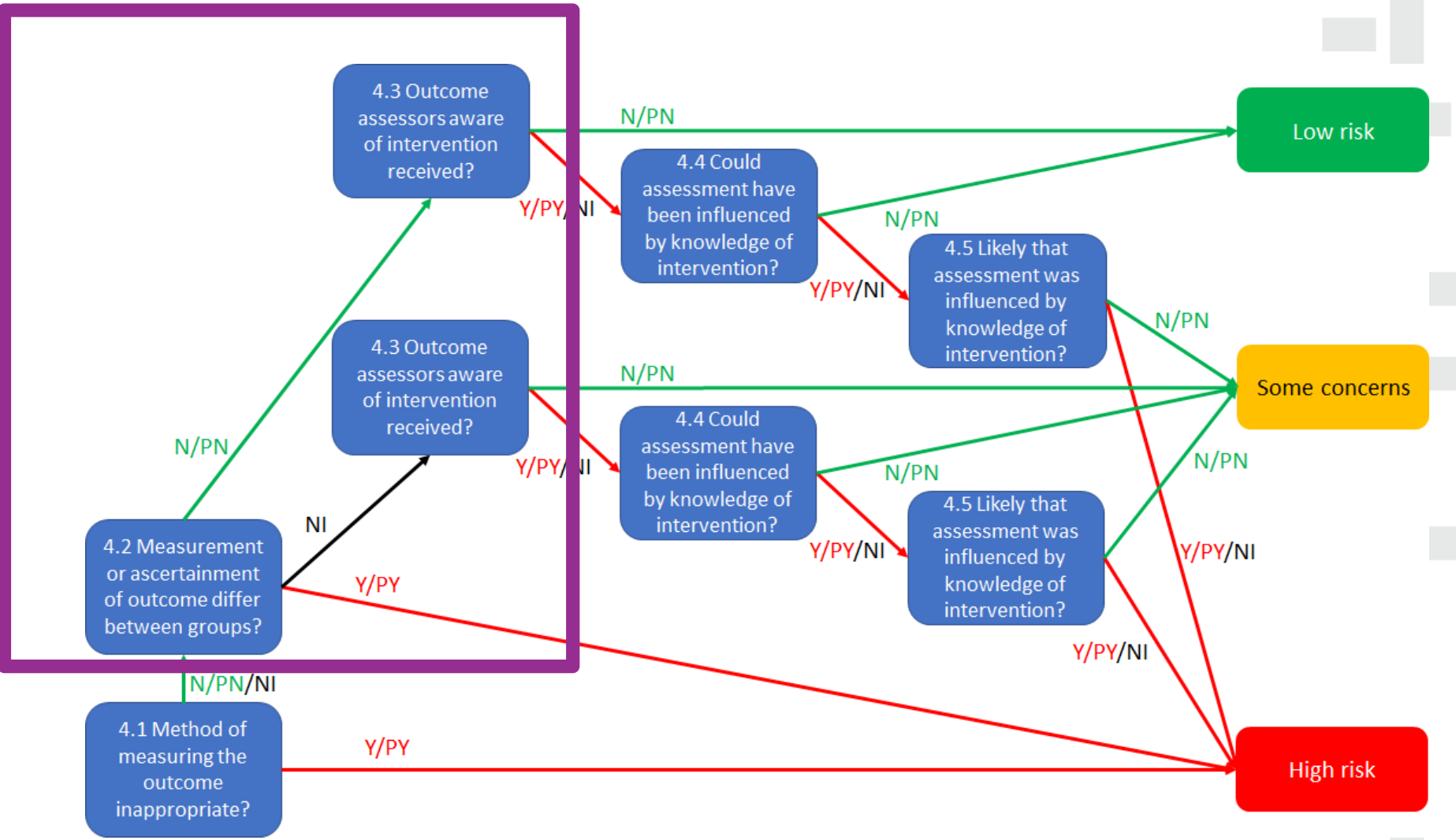
- **different outcome assessors**
  - surgeons/GP
- **'Diagnostic detection bias'**
  - Number of visits differ because of the intervention evaluated -> increasing opportunities to detect outcome events
  - Treatment adverse event → complementary tests more frequently performed on one arm
    - Ex: headache ->MRI -> tumor more likely to be detected

4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?

- **NO/Probably NO**
  - Comparable methods of measurement (same methods, definition, time points, assessors)

- **Yes/probably yes**
  - Context of passive collection of outcome data (adverse event) or additional visits in one group

Cochrane Training



Cochrane Training

Assessing the risk of bias in measurement of the outcome

Signalling questions 3-5

Bias in measurement of the outcome

The role of blinding

- who is assessing the outcome
- whether outcome assessor is blinded to intervention assignment
- whether assessment of outcome is likely to be influenced by knowledge of intervention assignment

Blinding and outcome assessors

A person measuring, ascertaining or recording the outcome is an 'outcome assessor':

i. an observer not directly involved in the intervention provided to the participant, such as an adjudication committee, a biologist performing an automated test, or a health professional recording outcomes from health records or disease registries.

Ii. the participant when the outcome is participant-reported: for example pain, quality of life, or self-completed questionnaire evaluating depression, anxiety or function.

iii. the intervention provider when the outcome is the result of a therapeutic decision such as a decision to offer a surgical intervention or to discharge the patient.

Blinding reporting

Reporting
Often inadequate in trial reports.

‘26% of journal articles reported no information on blinding whatsoever beyond the trial being ‘double blind’.

More details in protocols

Table 2 Reporting on blinding in journal articles (n = 200)

	Double blind		Single blind		Not DB/SB[1]	
	n	%	*n*	%	*n*	%
	156	78	12	6	32	16
Reporting of blinding status of key trial persons						
Complete (all categories[2])	0	0	0	0	0	0
Partial (patients and health care providers and data collectors)	3	2	0	0	7	22
Minimal (at least one category[3])	65	42	3	25	24	75
None (no explicit information)	88	56	9	75	1	3
No information at all beyond trial being ‘blind’, eg, ‘double blind’	41	26	7	58	1	3
Experimental and control treatments appear ‘similar’[4]	72	46	1	8	10	31
Time of unblinding described	14	9	0	0	1	3
Blinding mentioned in discussion	10	6	0	0	4	13

[1]Trials not described as ‘single-blind’ or ‘double-blind’. Typically such trials described blinding with other words (eg, ‘assessor-blind’).
[2]Patients, health care providers, data collectors, assessors of outcome, data analysts, manuscript writers.
[3]Excluding trials with ‘partial reporting’ and ‘complete reporting’.
[4]Including analogue terms, eg, ‘identical’ or ‘indistinguishable’.

Haahr Clin Trials 2006, Hróbjartsson et al. JCE 2009

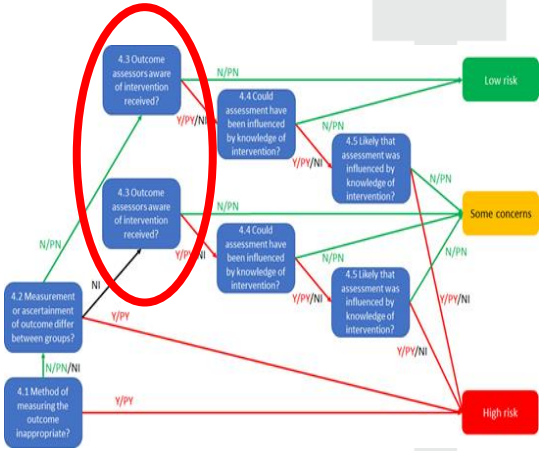
Signalling question 4.3

Whether outcome assessors were aware of the intervention received by study participants?

It is important to determine whether outcome assessments were **made blinded** to intervention assignment. If blinding was **successfully implemented**, then the risk of bias due to differential measurement error is low.

Component 1: were outcome assessors intended to be blind?

Component 2: was intention of blinding successful?



Outcome assessments made blinded to interventions

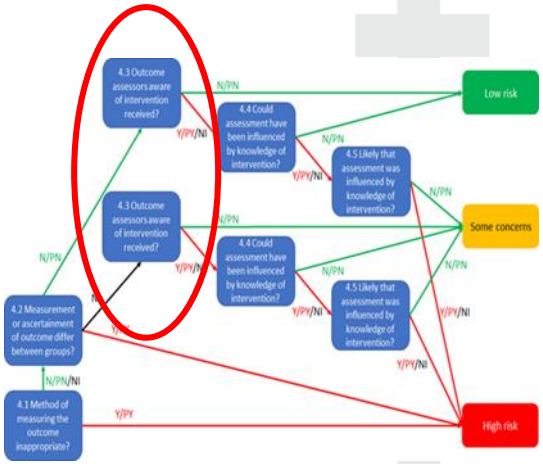
When is blinding of outcome assessors intended?

Green flag

-”Outcome assessors were blinded”
-”Non-blind participants and blind outcome assessor”
-”Double-blind drug trial with no indication of lack of blinding of outcome assessor"

Red flag

-"single blind" or "double-blind" only information
-external assessors not involved in patient care (but blinding not mentioned explicit)
-”Blind assessors interviewed non-blind patients”



Successfully implemented blinding of outcome assessor

When is blinding of outcome assessors successful?

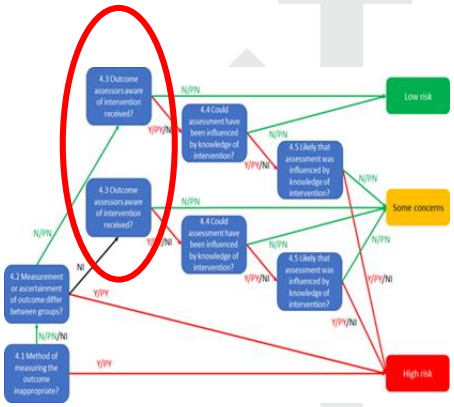
Green flag

Pre-trial testing of matching of compared interventions
Assessor interaction with non-blind patients and description of procedures to handle cases of accidental unblinding
No tell-tale effects

Red flags

Assessor interaction with non-blind patients and no procedures to handle the risk of unblinding
Tell-tale effects (taste of zinc)
Run-in periods (active or placebo)

Probably less important than if blinding was intended.



Bello JCE 2017

Signalling question 4.4

Could assessment have been influenced by knowledge of intervention?

The importance of lack of blinding of the outcome assessor will depend on the extent to which the assessment **can be influenced by** knowledge of the intervention assignment.

Green Flag

Objective outcomes: **all-cause mortality** and (some) automated test procedure, e.g. laboratory measurements

Red flag

Subjective outcomes

Subjective/objective Outcomes

Subjective outcomes
involving judgement
moderate to high inter-observer variation

Objective outcomes
not involving judgement
no or low inter-observer variation

Other uses of subjective/objective not relevant for RoB2
Objective: observer-reported
Subjective: inherently private to a person

Moustgaard JCE 2014

A model for observer bias

Different outcomes

- Objective (e.g. all-cause mortality)
- **Subjective** (e.g. global improvement, clinical function score)

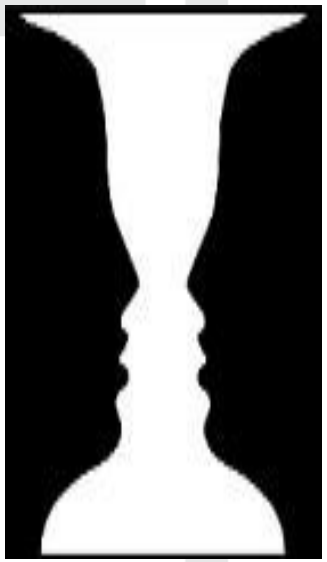
Different persons

- **Preconceptions**: none
- Preconceptions: some or strong.

Red flag

Person with **preconceptions** observing a **subjective outcome**

Lampard's shot, 2010 World Cup. England trailing Germany 1:2. Goal or not?

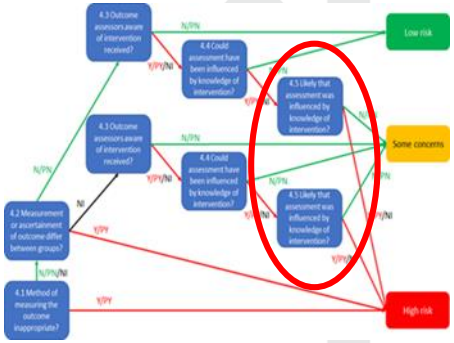


Signalling question 4.5

Was it likely that assessment was influenced by knowledge of intervention?
When the outcome assessor could have been influenced by knowledge of intervention received, users should assess whether it is likely that such influence occurred.

Considerations: trial context
Preconceptions
Hope
Hunches

Conflicts of interest



SQ 4.5: Likely influence?

Red flags: high risk of bias

Experimental intervention vs no-treatment or usual care control
Outcome assessors strongly engaged in other parts of the trial
Outcome with high degree of subjectivity

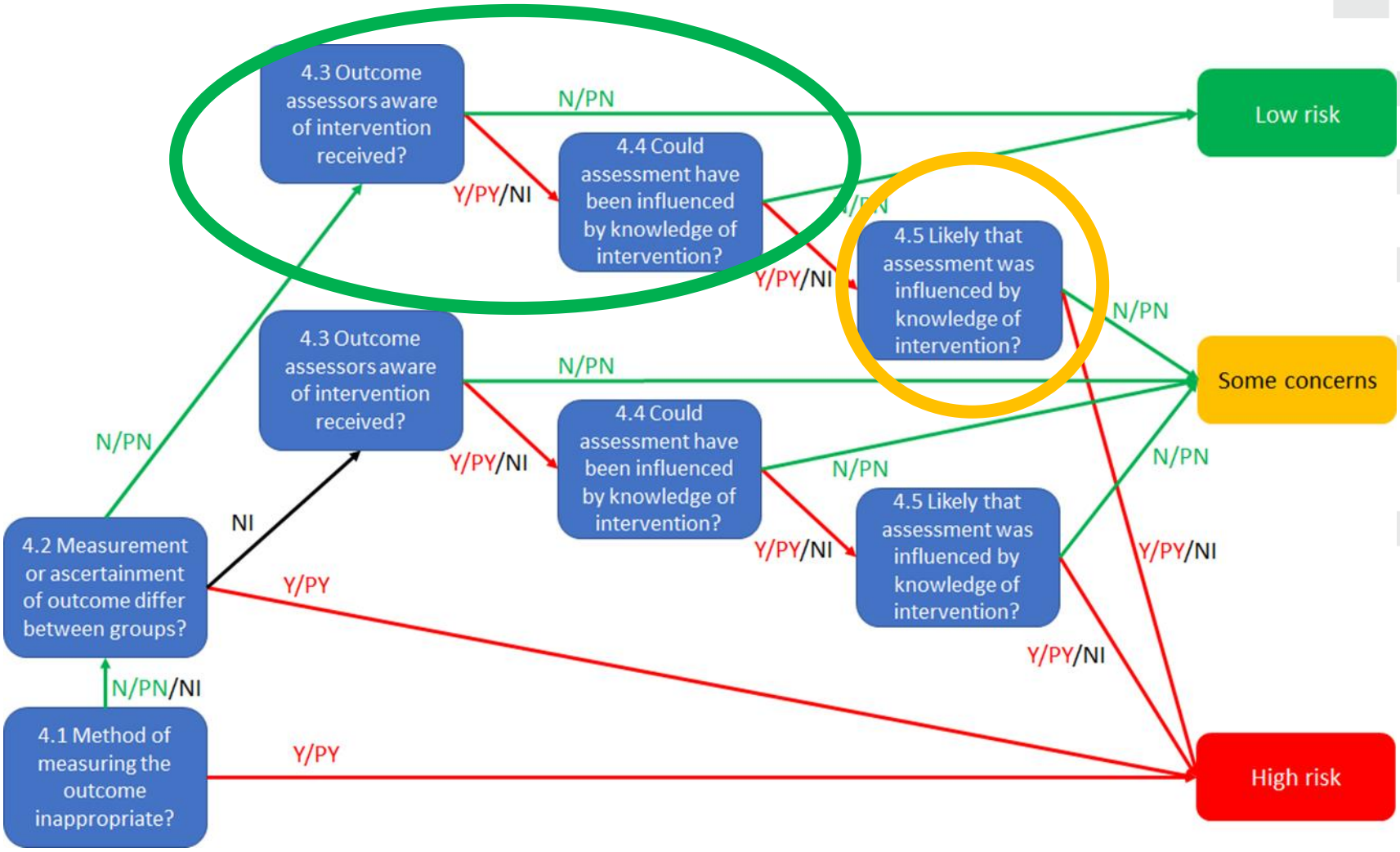
Click to add text

Green Flags: some concern

Active control group
External outcome assessor not otherwise engaged in the trial
Low degree of outcome subjectivity

Cochrane Training

Algorithm



Challenges

Blinding terminology in flux
”double-blind” carry different meanings to different authors

Look for direct descriptions

Reporting of blinding often inadequate
in publications

Use supplemental sources of information

Information on risk of unblinding often missing
assessment informal, absent and not reported

If suspected, contact authors

RoB2 involves judgements based on imperfect information

Example: observer is outcome assessor

Recombinant Human Bone Morphogenetic Protein-2: A Randomized Trial in Open Tibial Fractures Treated with Reamed Nail Fixation

By Hannu T. Aro, MD, PhD, Shunmugam Govender, MBBS, MD, FRCS, Amratlal D. Patel, FRCS, Philippe Hernigou, MD, Arturo Perera de Gregorio, MD, Gheorghe Ion Popescu, MD, Jane Davis Golden, MHP, Jared Christensen, PhD, and Alexandre Valentin, MD

Investigation performed at Turku University Hospital, Turku, Finland; University of Kwazulu-Natal, Durban, South Africa; Norfolk and Norwich University Hospital, Norwich, United Kingdom; Hôpital Henri Mondor, Créteil, France; Hospital Principe de Asturias, Madrid, Spain; Emergency Hospital, Bucharest, Romania; and Pfizer, Inc., Cambridge, Massachusetts

277 patients randomised to usual care + rHBMP-2 vs. usual care

"This was a multicenter **single-blinded** randomized study conducted at twenty-eight European and South African sites".

"The primary efficacy end point was the proportion of subjects with a **healed fracture** as demonstrated by radiographic and clinical assessment thirteen and twenty weeks after definitive wound closure."

"This study was limited by its single-blind design. Given the nature of the intervention under study, it was **not possible to blind the investigators** to the study group."

Outcome: radiographic union

Non-blinded surgeons, reported in paper: OR 0.74 (0.43 to 1.23)

Blinded radiologists, not reported in paper: OR 1.23 (0.53 to 2.89)

Example: patient is outcome assessor

Annals of Internal Medicine | Original Research

Echinacea for Treating the Common Cold

A Randomized Trial

Bruce Barrett, MD, PhD; Roger Brown, PhD; Dave Rakel, MD; Marlon Mundt, PhD; Kerry Bone, Dip Phyto; Shari Barlow, BA; and Tola Ewers, MS

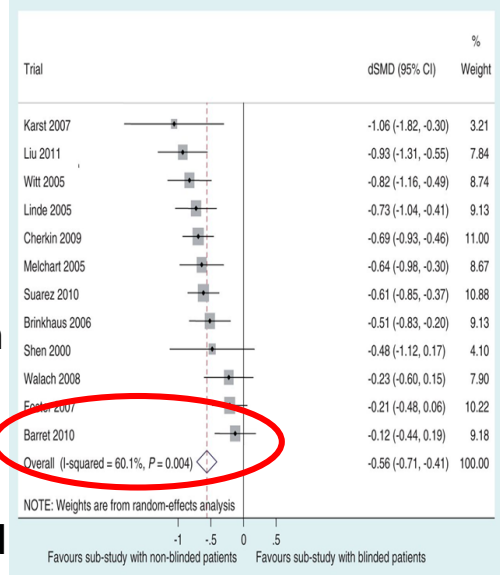
719 patients randomised to echinacea tablets vs placebo VS echinacea tablets vs no-treatment.

"Patients were assigned to 1 of 4 parallel groups: no pills, placebo pills (blinded), echinacea pills (blinded), or Echinacea pills (unblinded, open-label)."

"Placebo and echinacea tablets contained the same proportions of inert ingredients and were covered with identical digestible coatings".

"The primary outcome was the area under the curve for global severity, with severity assessed twice daily by self-report using the Wisconsin Upper Respiratory Symptom Survey, short version".

"Blinding seemed to be intact. Of the 363 participants who received pills and were blinded, 141 (39%) guessed their assignment correctly …"



Hróbjartsson IJE 2014

Conclusion

Important to differentiate between trials with low and high risk of bias in measurement of the outcome

Low risk of bias

- blinding implemented, and unlikely that the blinding could have been broken
- no blinding, but measurement unlikely to be influenced by knowledge of the intervention assignment

High risk of bias

- no blinding or broken blinding, and measurement likely to be influenced

Acknowledgements

- Slides in part based on material from Miranda Cumpston, Julian Higgins and Jonathan Sterne, Cochrane Bias Methods Group and the Australasian Cochrane Centre

Cochrane Training

Questions