

# Risk of Bias in systematic reviews of prognostic factor and model studies

Trusted evidence.  
Informed decisions.  
Better health.



# Conducting a systematic review of prognosis studies

1. Formulate review question (PICOTS)
2. Searching for studies
3. Screening and Selection of articles
4. Extraction of data
5. Risk of Bias assessments
6. Synthesis of data (meta-analysis)
7. Interpretation and conclusions

## RESEARCH METHODS AND REPORTING

### A guide to systematic review and meta-analysis of prognostic factor studies

Richard D Riley,<sup>1\*</sup> Karel G M Moons,<sup>2,4\*</sup> Kym I E Snell,<sup>1</sup> Joie Ensor,<sup>1</sup> Lotty Hooft,<sup>2,4</sup>  
Douglas G Altman,<sup>3</sup> Jill Hayden,<sup>5</sup> Gary S Collins,<sup>3</sup> Thomas P A Debray<sup>2,4</sup>

## RESEARCH METHODS AND REPORTING

### A guide to systematic review and meta-analysis of prediction model performance

Thomas P A Debray,<sup>1,2</sup> Johanna A A G Damen,<sup>1,2</sup> Kym I E Snell,<sup>3</sup> Joie Ensor,<sup>3</sup> Lotty Hooft,<sup>1,2</sup>  
Johannes B Reitsma,<sup>1,2</sup> Richard D Riley,<sup>3</sup> Karel G M Moons<sup>1,2</sup>

# Risk of Bias tools

- Overall prognosis studies
  - RoB-OPS – in preparation
- Prognostic factor/predictor finding studies
  - QUIPS – J Haydn, Ann Int Med 2006 + 2013
- Prognostic (prediction) model studies (development and validation)
  - PROBAST+AI – Moons, BMJ 2025
  - PROBAST-2019 E&E – Moons, Ann Int Med 2019



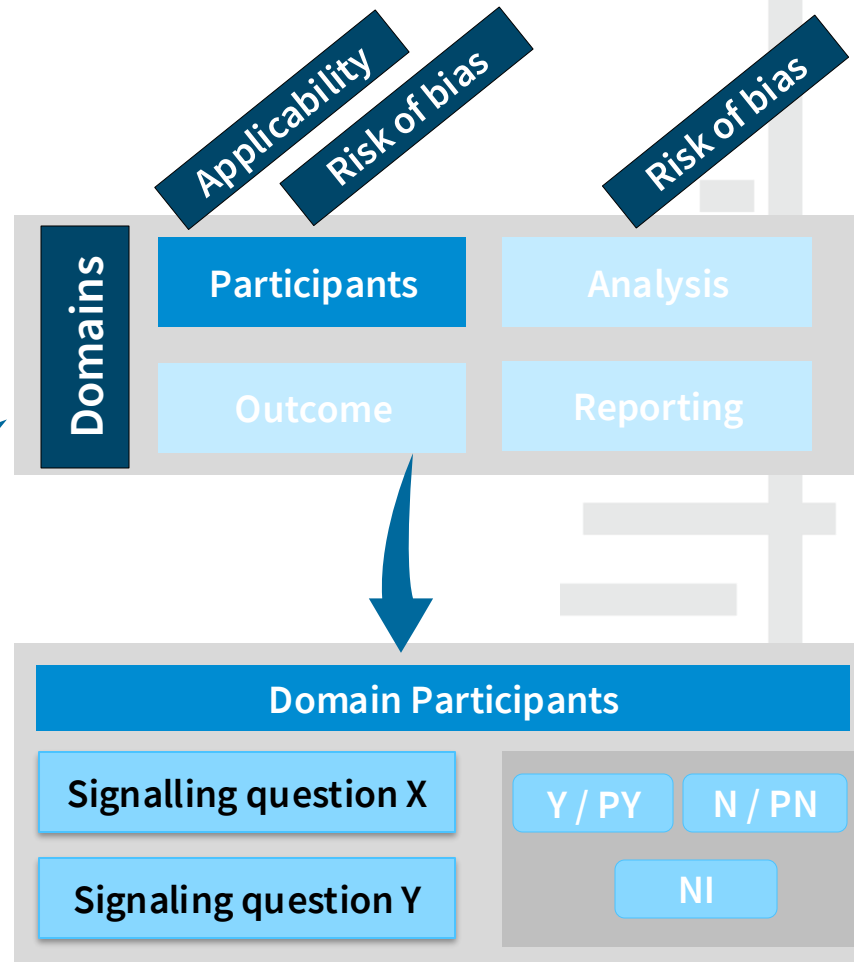
## RoB-OPS draft structure

Step 1: Define the questions

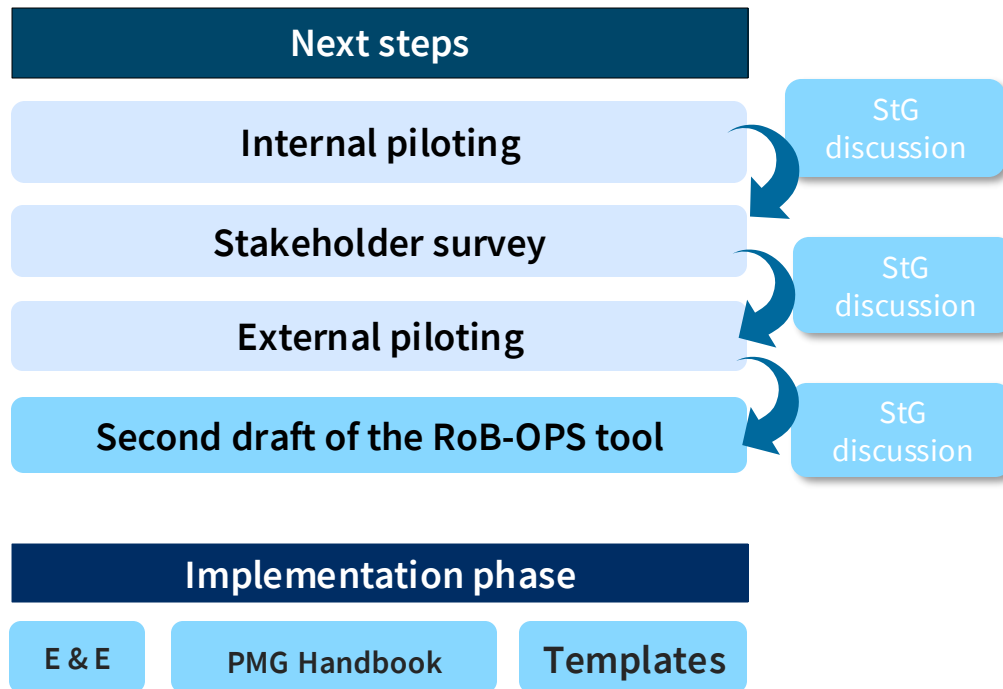
Step 2: Rate applicability

Step 3: Rate risk of bias

Step 4: Present the assessment(s)



# RoB-OPS

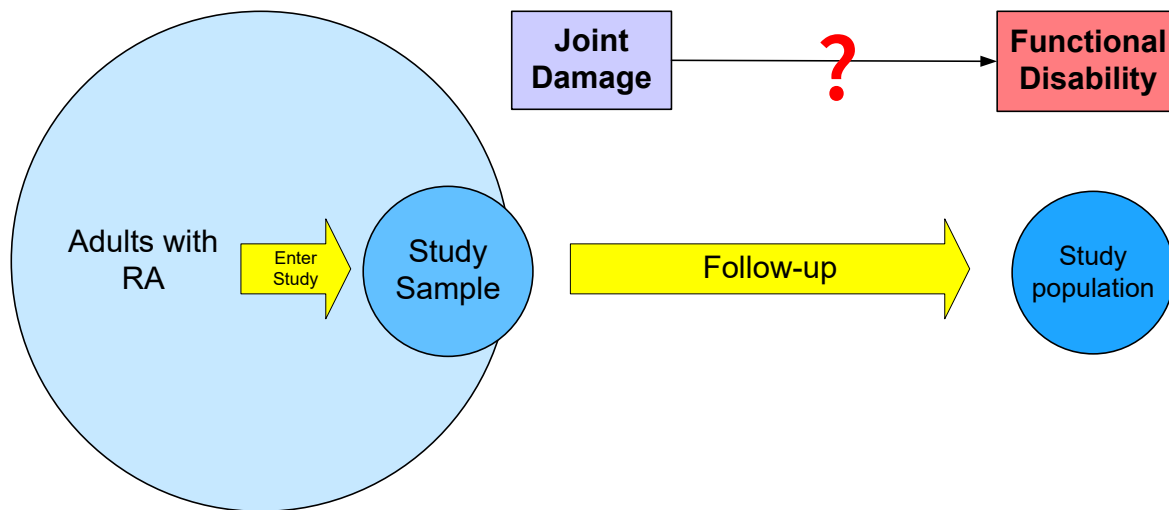


# Risk of Bias tools

- Overall prognosis studies
  - RoB-OPS – in preparation
- Prognostic factor/predictor finding studies
  - QUIPS – J Haydn, Ann Int Med 2006 + 2013
- Prognostic (prediction) model studies (development and validation)
  - PROBAST+AI – Moons, BMJ 2025
  - PROBAST-2019 E&E – Moons, Ann Int Med 2019



# Prognostic Factor Studies



# RoB prognostic factor studies: QUIPS Tool

Domain-based evaluation + signaling questions/items

- Follows QUADAS-2, ROBINS-I, RoB-2
- Assessments made separately for different bias domains
- Domains:
  1. Study participation
  2. Study attrition
  3. Prognostic factor measurement
  4. Outcome measurement
  5. Covariate adjustment
  6. Analysis and presentation





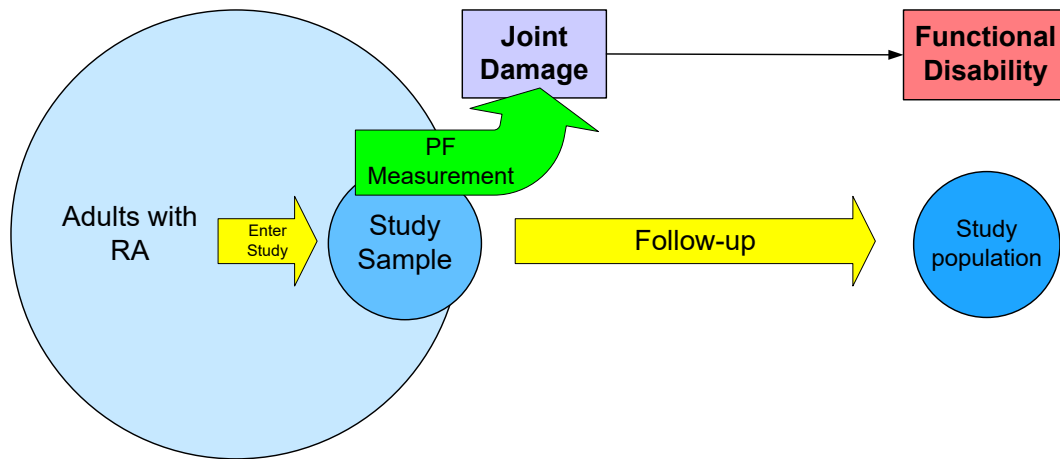
*Table 2. Domains Included in the Framework of Potential Biases and the Proportion of Reviews Assessing the Biases\**

Potential Bias	Studies Adequately Assessing Bias, %†	Domains Addressed	Studies Assessing Domain, %
1. The study sample represents the population of interest on key characteristics, sufficient to limit potential bias to the results (study participation).	55	1. Source population clearly defined 2. Study population described 3. Study population represents source population or population of interest	50 21 50
2. Loss to follow-up (from sample to study population) is not associated with key characteristics, sufficient to limit potential bias (i.e., the study data adequately represent the sample) (study attrition).	42	4. Completeness of follow-up described 5. Completeness of follow-up adequate	19 42
3. The prognostic factor of interest is adequately measured in study participants to sufficiently limit potential bias (prognostic factor measurement).	59	6. Prognostic factors defined 7. Prognostic factors measured appropriately	31 59
4. The outcomes of interest are adequately measured in study participants to sufficiently limit potential bias (outcome measurement).	51	8. Outcome defined 9. Outcome measured appropriately	42 51
5. Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest (confounding measurement and account).	13	10. Confounders defined and measured 11. Confounding accounted for	21 53
6. The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid results (analysis).	33	12. Analysis described 13. Analysis appropriate 14. Analysis provides sufficient presentation of data	8 33 32

\* Data are from 153 prognostic systematic reviews with quality items that could be extracted.

† Adequate assessment included 1) study participation: "source population clearly defined" and "study population described" or "study population represents source

# Opportunities for bias



Adapted from: *Fletcher & Fletcher, Clinical Epidemiology – The Essentials. Chapter 6. Williams & Wilkins, Baltimore. 1996*

# Risk of Bias tools

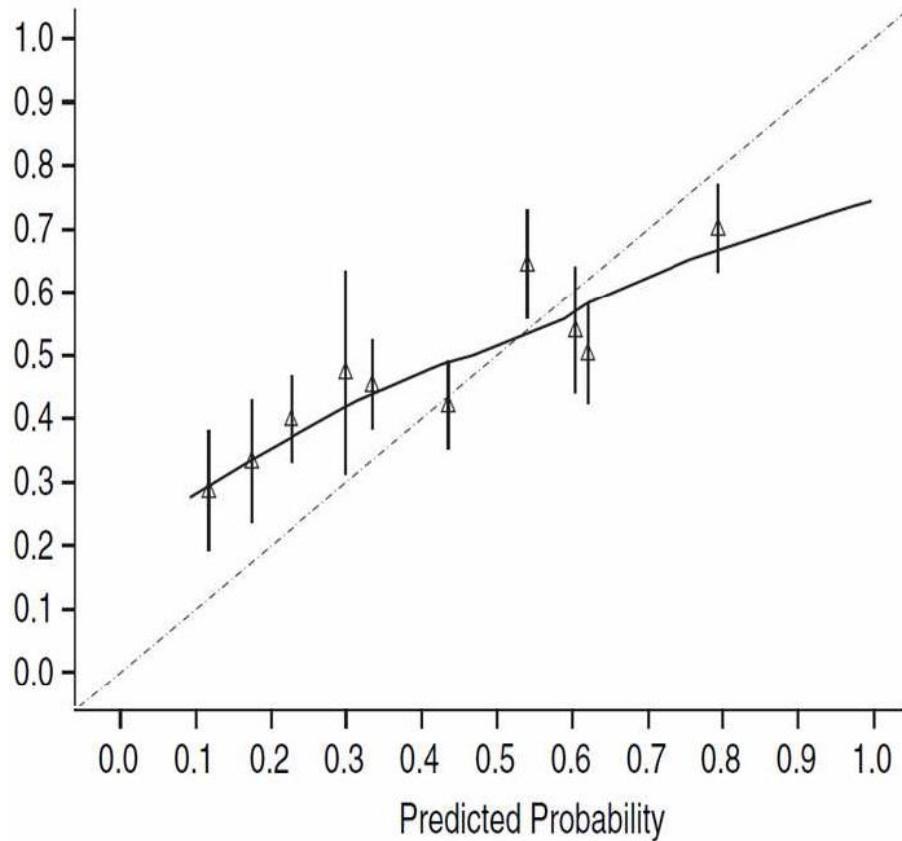
- Overall prognosis studies
  - RoB-OPS – in preparation
- Prognostic factor/predictor finding studies
  - QUIPS – J Haydn, Ann Int Med 2006 + 2013
- Prognostic (prediction) model studies (development and validation)
  - PROBAST+AI – Moons, BMJ 2025
  - PROBAST-2019 E&E – Moons, Ann Int Med 2019



# Specific issue in prediction model studies

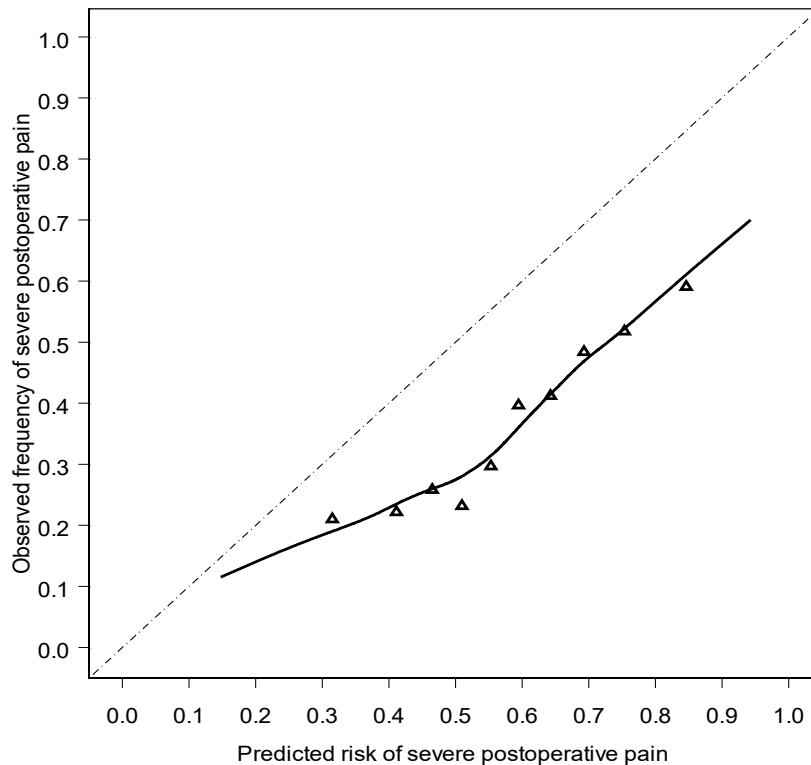
- Quality of prognostic model development
- Overfitted models
  - too large ROC area
  - too optimistic calibration plot or outcome classification
    - Wrong estimated predictor weights
    - Wrong estimated intercept
- Unfortunately: often don't know from development study → only visible until model validation → ideally external





## Slope < 1.0

- Low prob too low
- High prob too high



Systematic  
overestimation  
predicted probabilities

Intercept (outcome  
incidence) development  
study too high!

## PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods

Karel G M Moons,<sup>1</sup> Johanna A A Damen,<sup>1,2</sup> Tabea Kaul,<sup>1</sup> Lotty Hooft,<sup>1,2</sup>  
Constanza Andaur Navarro,<sup>1</sup> Paula Dhiman,<sup>3</sup> Andrew L Beam,<sup>4</sup> Ben Van Calster,<sup>5,6</sup>  
Leo Anthony Celi,<sup>7,8</sup> Spiros Denaxas,<sup>9,10</sup> Alastair K Denniston,<sup>11</sup> Marzyeh Ghassemi,<sup>12</sup>  
Georg Heinze,<sup>13</sup> André Pascal Kengne,<sup>14</sup> Lena Maier-Hein,<sup>15,16</sup> Xiaoxuan Liu,<sup>11,17,18,19</sup>  
Patricia Logullo,<sup>3</sup> Melissa D McCradden,<sup>20</sup> Nan Liu,<sup>21</sup> Lauren Oakden-Rayner,<sup>22</sup>  
Karandeep Singh,<sup>23</sup> Daniel S Ting,<sup>21,24</sup> Laure Wynants,<sup>5,25</sup> Bada Yang,<sup>1,2</sup> Johannes B Reitsma,<sup>1</sup>  
Richard D Riley,<sup>18,19</sup> Gary S Collins,<sup>3</sup> Maarten van Smeden<sup>1</sup>

# Focus

- Diagnostic and prognostic prediction models
- Development and evaluation (validation) studies
- All types of predictors and health outcomes
- All types of techniques
  - Incl. Machine Learning / Artificial Intelligence





# Structure of PROBAST

Also domain-based: each with section **quality / risk of bias + applicability**

*Quality* refers to the methodological quality of the model development or production process.

*Risk of bias* is a systematic error in the estimates of the model's true predictive performance. The predictive performance is ideally evaluated using calibration, discrimination, and clinical utility.

*Applicability* refers to the extent to which the prediction model from the study matches your systematic review question, for example in terms of the population or outcomes of interest.



## PROBAST 4 phases

Step	Task	When to complete
1	Specify the intended purpose of the prediction model assessment or of the prediction model systematic review	Once per assessment or systematic review
2	Classify the type of prediction model study (development or evaluation or both)	Once for each prediction model of interest in each publication assessed, for each relevant outcome
3	Assess quality and applicability to the intended purpose of the prediction model for model development for the separate domains & Assess risk of bias and applicability to the intended purpose of the prediction model for model evaluation for the separate domains	Once for each model development for each distinct prediction model in a publication  Once for each model evaluation for each distinct prediction model in a publication
4	Assess the overall quality, risk of bias, and applicability of the prediction model (study)	Once for each distinct assessment of each prediction model in a publication

## Step 3: Assess quality and/or risk of bias, and applicability

<b>DOMAIN 1: Participants and data sources</b>	
<b>A. Quality</b>	
<i>Describe the sources of data and criteria for participant selection:</i>	
	<b>Y/ PY/ PN/ N/ NI</b>
1.1 Were appropriate data sources used?	
1.2 Was an appropriate study design used?	
1.3 Did the in- and exclusions of study participants result in a representative dataset?	
Concern regarding quality of selection of participants and data sources	<b>QUALITY CONCERN:</b> <i>low/high/unclear</i>
<i>Rationale of quality rating:</i>	
<b>B. Applicability</b>	
<i>Describe included data sources, participants, setting, and dates:</i>	
Concern that the (data of the) included participants do not match the review question or the assessor's intended use of the prediction model	<b>APPLICABILITY CONCERN:</b> <i>low/high/unclear</i>
<i>Rationale of applicability rating:</i>	

# Practical

## Split group in subgroups



## QUIPS Risk of Bias Assessment Instrument for Prognostic Factor Studies

Modified from: Hayden JA, Côté P, Bombardier C. Evaluation of the Quality of Prognosis Studies in Systematic Reviews. *Annals of Internal Medicine*. 2006;144:427-437, with the assistance of the QUIPS-LBP Working Group.

Author and year of publication	Study identifier	Reviewer			
<b>Biases</b>	<b>Issues to consider for judging overall rating of "Risk of bias"</b>	<b>Study Methods &amp; Comments</b>	<b>Rating of reporting</b>	<b>Rating of "Risk of bias"</b>	
Instructions to assess the risk of each potential bias:	These issues will guide your thinking and judgment about the overall risk of bias within each of the 6 domains. Some issues may not be relevant to the specific study or the review research question. These issues are taken together to inform the overall judgment of potential bias for each of the 6 domains.	Provide comments or text excerpts in the white boxes below, as necessary, to facilitate the consensus process that will follow.	Click on each of the blue cells and choose from the drop-down menu to rate the adequacy of reporting as yes, partial, no or unsure.	Click on the green cells; choose from the drop-down menus to rate potential risk of bias for each of the 6 domains as High, Moderate, or Low considering all relevant issues	
<b>1. Study Participation</b>	<b>Goal: To judge the risk of selection bias (likelihood that relationship between PF and outcome is different for participants and eligible non-participants).</b>				
Source of target population	The source population or population of interest is adequately described for key characteristics (LIST).				
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias (number and type used, e.g., referral patterns in health).				
Recruitment period	Period of recruitment is adequately described.				
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described.				
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or "zero time" description).				
Adequate study participation	There is adequate participation in the study by eligible individuals.				
Baseline characteristics	The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics (LIST).				
<b>Summary Study participation</b>	<b>The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between PF and</b>				
<b>2. Study Attrition</b>	<b>Goal: To judge the risk of attrition bias (likelihood that relationship between PF and outcome are different for completing and non-completing).</b>				
Proportion of baseline sample available for analysis	Response rate (i.e., proportion of study sample completing the study and providing outcome data) is adequate.				
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described.				
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided.				
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for key characteristics (LIST). There are no important differences between key characteristics (LIST) and outcomes in participants who completed the study and those who did not.				
<b>Study Attrition Summary</b>	<b>Loss to follow-up (from baseline sample to study population analyzed) is not associated with key characteristics (i.e., the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between PF and outcome.</b>				
<b>3. Prognostic Factor Measurement</b>	<b>Goal: To judge the risk of measurement bias related to how PF was measured (differential measurement of PF related to the level of outcome).</b>				
Definition of the PF	A clear definition or description of PF is provided (e.g., including dose, level, duration of exposure, and clear specification of the method of measurement).				
Valid and Reliable Measurement of PF	Method of PF measurement is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).				
Method and Setting of PF Measurement	Continuous variables are reported or appropriate out-points (i.e., not data-dependent) are used.				
Proportion of data on PF available for analysis	The method and setting of measurement of PF is the same for all study participants.				
	Adequate proportion of the study sample has complete data for PF variable.				

The subsequent steps for use of the PROBAST+AI tool.

Step	Task	When to complete
1	Specify the intended purpose of the prediction model assessment or of the prediction model systematic review	Once per assessment or systematic review
2	Classify the type of prediction model study (development or evaluation or both)	Once for each prediction model of interest in each publication assessed, for each relevant outcome
3	Assess quality and applicability to the intended purpose of the prediction model for model development, for each domain & Assess risk of bias and applicability to the intended purpose of the prediction model for model evaluation, for each domain	Once for each model development for each distinct prediction model in a publication  Once for each model evaluation for each distinct prediction model in a publication
4	Assess the overall quality and applicability for model development, and separately, the risk of bias and applicability for model evaluation	Once for each assessed prediction model in a publication, and separately for model development and for model evaluation.

**Go to:**

**<https://methods.cochrane.org/prognosis/workshops-and-events>**

And download the workshop materials for either QUIPS or PROBAST+AI

Or scan the QR code:



## EXTRA

# What to do with your risk of bias assessments?





# Presentation of Risk of Bias

‘Risk of Bias’ table (transparent reporting)

Judge the specific domains for each study:

- Low risk of bias
- Moderate risk of bias
- High risk of bias

Provide complete descriptions from studies supporting judgments



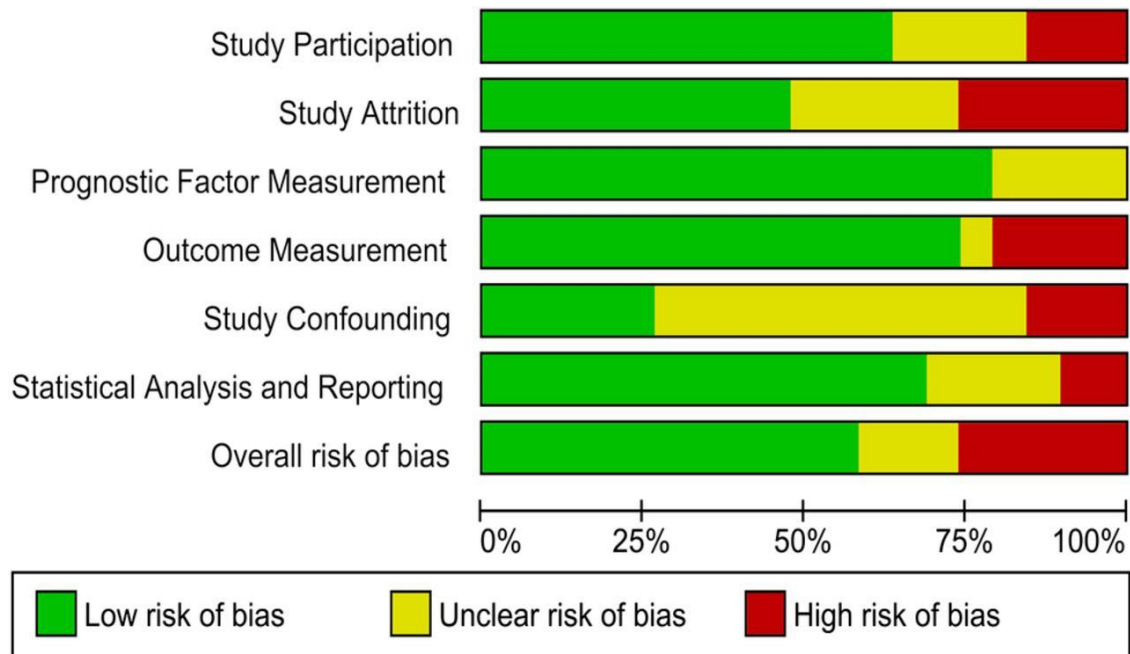
# Quality assessment/Risk of Bias Tool prognostic factor studies Presentation across studies

	Risk of Bias					
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
Jiang 2015	Low	High	Low	Moderate	Moderate	Low
Schwindt* 2015	Moderate	High	Moderate	Low	Moderate	Low
Nunes* 2014	Moderate	High	Moderate	Moderate	Moderate	Low
Schumacher* 2013	Low	Moderate	Moderate	Low	High	High
Hayashi-Kurahashi 2012	Low	High	Moderate	Moderate	Low	Low
Le Bihannic* 2012	Moderate	Low	Low	Moderate	Moderate	Low
Wikström 2012	Moderate	High	Moderate	Low	High	Low
Klebermass 2011	High	High	Moderate	Low	Moderate	Low
West 2011	Low	Moderate	Low	Moderate	Moderate	Low
Kidokoro 2010	High	Low	Low	High	High	Low
Maruyama 2002	Low	Low	Low	Moderate	High	Low
Hellström-Westas* 1991	High	Low	Low	High	High	Low
Tharp 1981	Moderate	Low	Moderate	High	High	Low

High  
Moderate  
Low

# Quality assessment/Risk of Bias Tool prognostic factor studies

## Presentation RoB summary



# Incorporating Assessments into Analyses

Not appropriate to ignore potential biases

Trade-off between bias and precision

- Including all eligible studies will produce a result with high precision
- But results may be biased due to flaws

Cautious analysis and interpretation



## Approaches to Include RoB Results in Analysis

Restrict primary analysis to ONLY studies with low risk of bias (e.g. on all domains)

- Threshold-type of approach (arbitrary)
- Sensitivity analysis including higher risk studies

Explore the impact of individual bias domains

- Graphically according to risk of bias
- Comparison of subgroups



# Take home messages

*Overall prognosis studies* - What is most likely course (outcome) of individuals with certain health condition  
**RoB-OPS (in preparation)**

*Predictor finding studies* - which predictors contribute to prediction of particular prognostic/diagnostic outcome - aim not to develop a model for individualised predictions  
**QUIPS (Hayden, Ann Intern Med 2005)**

*Model development studies* - to develop prediction model from data: identify important predictors; estimate predictor weights; construct model for individualised predictions; quantify predictive performance; internal validation  
**PROBAST+AI (2025) - Formal Risk of Bias tool  
Prognostic and Diagnostic**

*Model validation studies* - test (validate) predictive performance of previously developed model in participant data other than development set

*Model impact studies* - quantify effect/impact actually using model on participant/physician management and health outcomes - relative to not using the model → comparative studies.  
**Comparative, intervention studies - RoB Cochrane (Higgins BMJ 2011)**

# Reporting guideline prediction modeling studies

RESEARCH METHODS AND REPORTING



OPEN ACCESS



Check for updates

## TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods

Gary S Collins,<sup>1</sup> Karel G M Moons,<sup>2</sup> Paula Dhiman,<sup>1</sup> Richard D Riley,<sup>3,4</sup> Andrew L Beam,<sup>5</sup> Ben Van Calster,<sup>6,7</sup> Marzyeh Ghassemi,<sup>8</sup> Xiaoxuan Liu,<sup>9,10</sup> Johannes B Reitsma,<sup>2</sup> Maarten van Smeden,<sup>2</sup> Anne-Laure Boulesteix,<sup>11</sup> Jennifer Catherine Camaradou,<sup>12,13</sup> Leo Anthony Celi,<sup>14,15,16</sup> Spiros Denaxas,<sup>17,18</sup> Alastair K Denniston,<sup>4,9</sup> Ben Glocker,<sup>19</sup> Robert M Golub,<sup>20</sup> Hugh Harvey,<sup>21</sup> Georg Heinze,<sup>22</sup> Michael M Hoffman,<sup>23,24,25,26</sup> André Pascal Kengne,<sup>27</sup> Emily Lam,<sup>12</sup> Naomi Lee,<sup>28</sup> Elizabeth W Loder,<sup>29,30</sup> Lena Maier-Hein,<sup>31</sup> Bilal A Mateen,<sup>17,32,33</sup> Melissa D McCradden,<sup>34,35</sup> Lauren Oakden-Rayner,<sup>36</sup> Johan Ordish,<sup>37</sup> Richard Parnell,<sup>12</sup> Sherri Rose,<sup>38</sup> Karandeep Singh,<sup>39</sup> Laure Wynants,<sup>40</sup> Patricia Logullo<sup>1</sup>

[www.tripod-statement.org](http://www.tripod-statement.org)