PROBAST

Prediction model risk of bias assessment tool

25th Cochrane Colloquium – 18th September 2018

Robert Wolff[1], Penny Whiting[2], Susan Mallett[3], Richard Riley[4], Marie Westwood[1], Jos Kleijnen[1,5], Karel Moons[6]

Kleijnen Systematic Reviews Ltd (1), Universities of Bristol (2), Birmingham (3), Keele (4), Maastricht (5), Utrecht (6)

I have no actual or potential conflict of interest in relation to this presentation.

www.systematic-reviews.com ○ robert@systematic-reviews.com

Prediction

Prediction = foreseeing / foretelling

- (probability) of something that is yet unknown

Largely two situations in medicine:

1. Probability of future conditions/situations = **prognosis**
2. Probability of result of a more invasive/costly reference (gold) standard test that is not yet done = **diagnosis**

"Prediction is difficult, especially about the future." (Piet Hein)

Prediction model

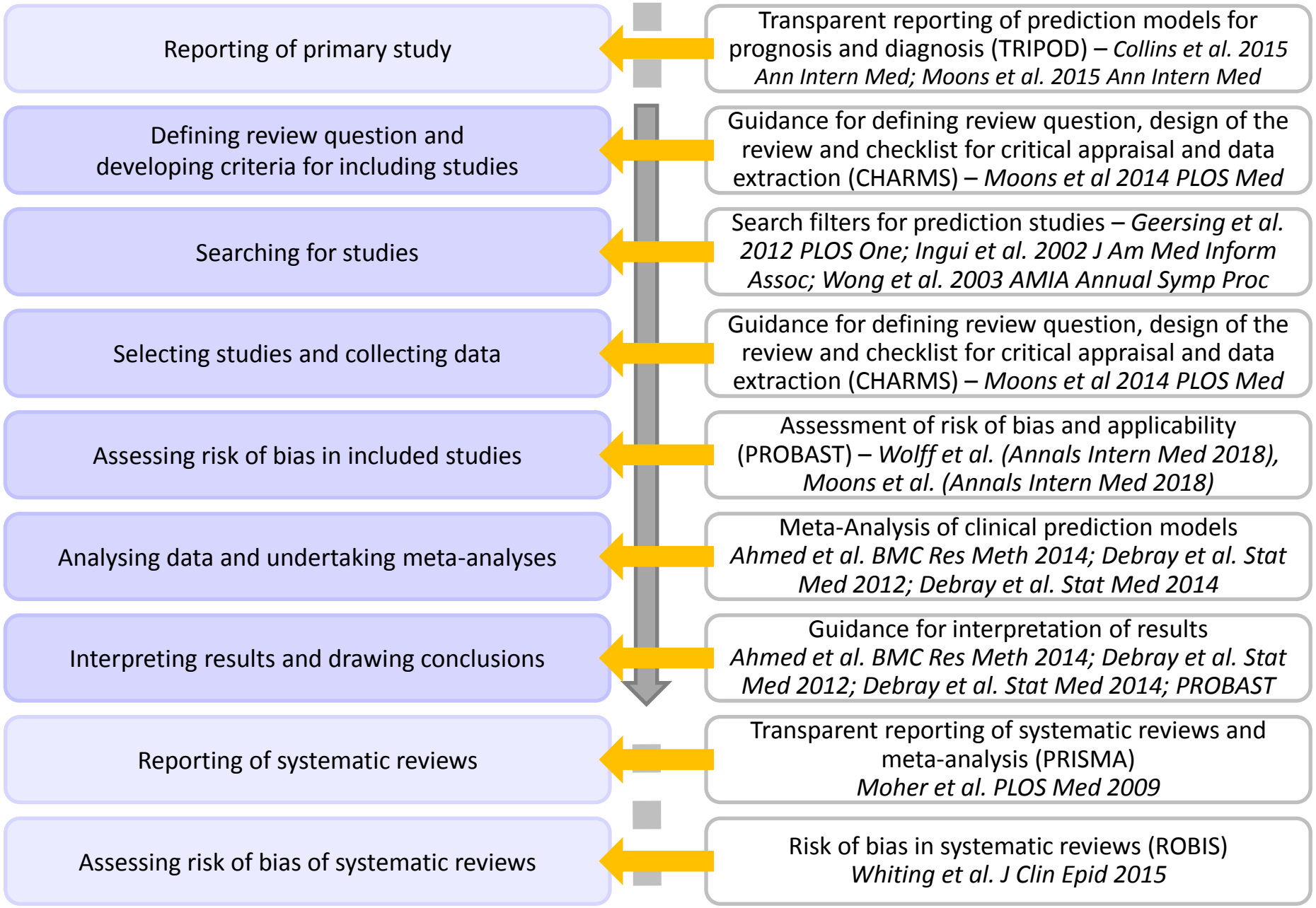
Combination of more than two predictors which convert observed values in individual to absolute probability…

- … of having a particular disease → **diagnosis**
- … of developing particular health state within a certain time (hours, days, weeks, years) → **prognosis**

Possible outcomes:

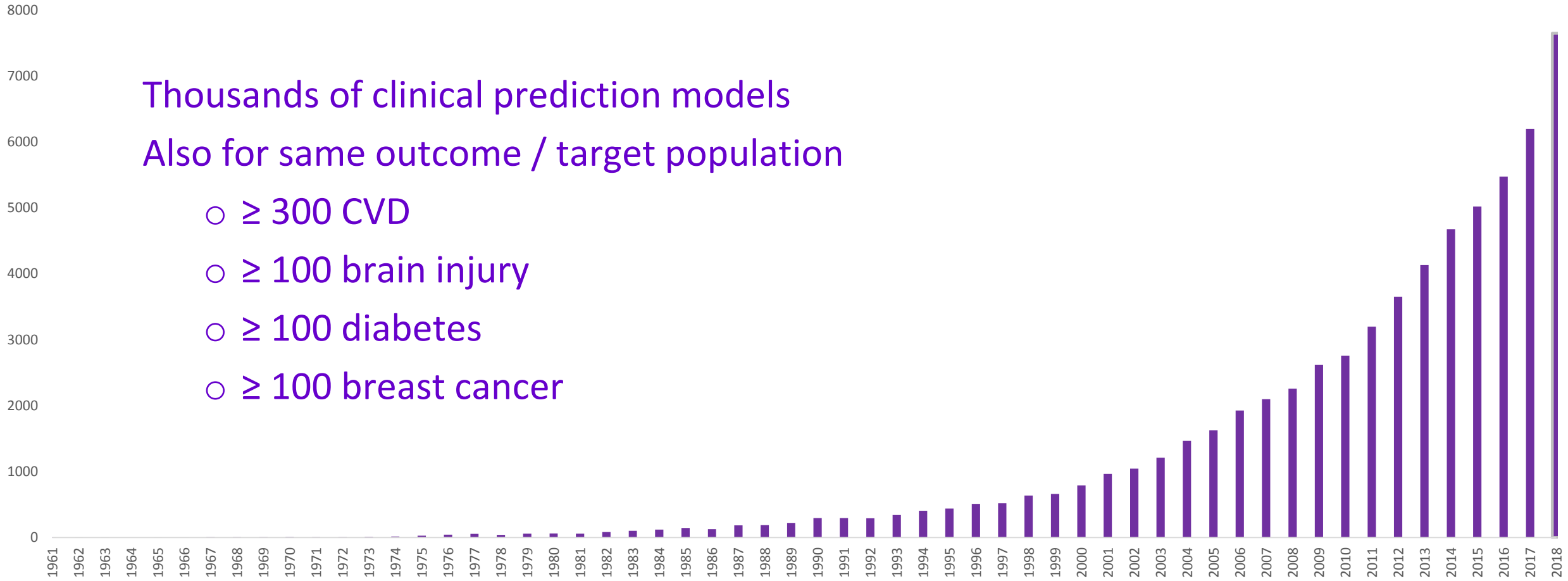
Death, complication, disease progression, pain, quality of life, hospitalisation, therapy response etc.

Conducting systematic reviews of prediction modelling studies



Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 - http://handbook.cochrane.org/

Popularity of prediction models



Thousands of clinical prediction models

Also for same outcome / target population

- ≥ 300 CVD
- ≥ 100 brain injury
- ≥ 100 diabetes
- ≥ 100 breast cancer

Search for "prediction model" on pubmed.gov (17/09/2018)

Systematic reviews of prediction model studies

Numerous methodology reviews:

- Mallett et al. BMC Med 2010
- Collins et al. J Clin Epidemiol. 2013
- Steyerberg et al. Epidemiology 2010
- Bouwmeester et al. PLoS Med 2012

- Conclusions from methodology reviews:
  - (Very) poor reporting
  - (Very) poor methods
  - Each SR: own search strategy, own checklist data extraction. Hardly ever risk of bias assessment

Development and structure of PROBAST

Development:

- Delphi procedure with 40 panel members
- Seven rounds
- Seven steering group members from six institutions
- Feedback from piloting

Structure:

- Assessment of risk of bias and applicability
- Follows QUADAS-2, ROBIS, ROBINS-i and ROB 2.0
- Four domains with 20 signalling questions

Which prediction studies?

Predictive factor studies - which predictors contribute to prediction of particular prognostic/diagnostic outcome – often using multivariable modelling – aim not to develop a prediction model for individualised predictions

QUIPS 2 – assessing bias in studies of prognostic factors
(Hayden et al. 2013 Ann Intern Med)

Model development studies – to develop prediction model(s) from data at hand: identify important predictors; estimate multivariable predictor weights; construct model for individualised predictions; quantify predictive performance in development set; internal validation.

PROBAST
(Diagnostic and prognostic models)

Model validation studies – test (validate) predictive performance of previously developed model in participant data other than development set – sometimes combined in development study – sometimes followed by updating/revision model

Model impact studies – quantify effect/impact actually using model on participant/physician behaviour and management, on health outcomes or cost-effectiveness of care – relative to not using the model → comparative studies.

Comparative, intervention studies – different risk assessment
→ ROB 2.0, ROBINS-i

Bouwmeester et al. PLoS Med 2012

Risk of bias / applicability

Risk of bias defined as any flaw or shortcoming in the design, conduct or analysis of a primary study that is likely to distort the predictive performance of a model. The predictive performance is typically evaluated using calibration, discrimination and sometimes classification measures, and these are likely to be overestimated in studies with high risk of bias.

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

DOMAIN 1: Participant selection			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection:			

		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			
1.2 Were all inclusions and exclusions of participants appropriate?			
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*		
Rationale of bias rating:			

B. Applicability			
Describe included participants, setting and dates:			

Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*		
Rationale of applicability rating:			

Step 4: Overall assessment

Use the following tables to reach overall judgements about risk of bias and concerns regarding applicability of the prediction model evaluation (development and/ or validation) across all assessed domains.
Complete for each evaluation of a distinct model.

Reaching an overall judgement about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to **high risk of bias**. Such model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at **high risk of bias**.
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Reaching an overall judgement about applicability of the prediction model evaluation	
Low concerns regarding applicability	If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have **low concerns regarding applicability.**
High concerns regarding applicability	If high concerns regarding applicability for at least one domain, the prediction model evaluation is judged to have **high concerns regarding applicability.**
Unclear concerns regarding applicability	If unclear concerns (but no "high concern") regarding applicability for at least one domain, the prediction model evaluation is judged to have **unclear concerns regarding applicability** overall.

Overall judgement about risk of bias and applicability of the prediction model evaluation		
Overall judgement of risk of bias	**RISK:** *(low/ high/ unclear)*	
Summary of sources of potential bias:		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	
Summary of applicability concerns:		

Domain 1 (Participant selection)

1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?

1.2 Were all inclusions and exclusions of participants appropriate?

Domain 2 (Predictors)

2.1 Were predictors defined and assessed in a similar way for all participants?

2.2 Were predictor assessments made without knowledge of outcome data?

2.3 Are all predictors available at the time the model is intended to be used?

Domain 3 (Outcome)

3.1 Was the outcome determined appropriately?

3.2 Was a pre-specified or standard outcome definition used?

3.3 Were predictors excluded from the outcome definition?

3.4 Was the outcome defined and determined in a similar way for all participants?

3.5 Was the outcome determined without knowledge of predictor information?

3.6 Was the time interval between predictor assessment and outcome determination appropriate?

Domain 4 (Analysis)

4.1 Were there a reasonable number of participants with the outcome?

4.2 Were continuous and categorical predictors handled appropriately?

4.3 Were all enrolled participants included in the analysis?

4.4 Were participants with missing data handled appropriately?

4.5 Was selection of predictors based on univariable analysis avoided?

4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?

4.7 Were relevant model performance measures evaluated appropriately?

4.8 Was model overfitting and optimism in model performance accounted for?

4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?

Development of PROBAST

- Ongoing piloting
  - Various settings, e.g. Cochrane authors, MSc students, guideline developers
  - Feedback positive. However, guidance needed
  - Please get in touch if you would like to use PROBAST or if you have any feedback

- Publications (submissions in 2018)
  - Background paper with the tool
  - Explanation and Elaboration (E&E)
  - Website

PROBAST group

Doug Altman, University of Oxford
Patrick Bossuyt, University of Amsterdam
Gary Collins, University of Oxford
Nancy Cook, Harvard University
Gennaro D´Amico, Ospedale V Cervello
Thomas Debray, University of Utrecht
Jon Deeks, University of Birmingham
Joris de Groot, University of Utrecht
Emanuele di Angelantonio, University of Cambridge
Tom Fahey, Royal College of Surgeons in Ireland
Paul Glasziou, Bond University
Frank Harrell, Vanderbilt University
Jill Hayden, Dalhousie University
Martin Heymans, University of Amsterdam
Lotty Hooft, University of Utrecht
Chris Hyde, Peninsula Technology Assessment Group
John Ioannidis, Stanford University
Alfonso Iorio, McMaster University
Stephen Kaptoge, University of Cambridge
Jos Kleijnen*, Kleijnen Systematic Reviews
Andre Knottnerus, Maastricht University
Mariska Leeflang, University of Amsterdam
Susan Mallett*, University of Oxford
Carl Moons*, University of Utrecht
Frances Nixon, NICE
Michael Pencina, University of Boston
Pablo Perel, London School of Hygiene and Tropical Medicine
Bob Philips, CRD
Heike Raatz, University of Basel
Hans Reitsma, University of Utrecht
Rob Riemsma, Kleijnen Systematic Reviews
Richard Riley*, University of Birmingham
Maroeska Rovers, University of Utrecht
Anne Rutjes, University of Bern
Willi Sauerbrei, University of Freiburg
Stefan Sauerland, IQWiG
Fülöp Scheibler, IQWiG
Rob Scholten, University of Utrecht
Ewoud Schuit, University of Utrecht
Ewout Steyerberg, University of Rotterdam
Toni Tan, NICE
Gerben ter Riet, University of Amsterdam,
Danielle van der Windt, Keele University
Yvonne Vergouwe, University of Rotterdam
Andrew Vickers, Memorial Sloan-Kettering CC
Marie Westwood*, Kleijnen Systematic Reviews
Penny Whiting*, Kleijnen Systematic Reviews
Robert Wolff,* Kleijnen Systematic Reviews
Angela Wood, University of Cambridge

* denotes steering group members