PROBAST: Prediction model risk of bias assessment tool

On behalf of PROBAST Steering Group

• Robert Wolff, Kleijnen Systematic Review
• Karl Moons, University of Utrecht
• Penny Whiting, University of Bristol
• Richard Riley, University of Keele
• Marie Westwood, Kleijnen Systematic Reviews
• Gary Collins, University of Oxford
• Hans Reitsma, University of Utrecht
• Sue Mallett, University of Birmingham
Prediction

• Prediction = foreseeing / foretelling
  ➢ (probability) of something that is yet unknown

• Largely two situations in medicine:
  ➢ Probability of future conditions/situations = prognosis
  ➢ Probability of result of test that is not yet done (a more invasive/costly reference standard) = diagnosis
Prediction models

Daily Mail 15th July 2016

• Sprinkling cinnamon on cereal may help us learn new information quicker

• Could your skin colour kill you? White people with anger problems are 40% more likely to die from a heart attack
Different types of prognostic studies

• Prognostic factor studies
  ➢ identify risk factors of a population, but do not attempt to develop a model of risks for individuals

• Prediction of risk in an individual
  ➢ Model converts observed values of two or more characteristics of an individual to risk probability in an individual
### Conducting systematic reviews of prediction model studies

<table>
<thead>
<tr>
<th>Step</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting of primary study</td>
<td>transparent reporting of prediction models for prognosis and diagnosis (TRIPOD) – Collins et al. 2015 Ann Intern Med; Moons et al. 2015 Ann Intern Med</td>
</tr>
<tr>
<td>Searching for studies</td>
<td>Search filters for prediction studies – Geersing et al. 2012 PLOS One; Ingui et al. 2002 J Am Med Inform Assoc; Wong et al. 2003 AMIA Annual Symp Proc</td>
</tr>
<tr>
<td>Selecting studies and collecting data</td>
<td>Guidance for defining review question, design of the review and checklist for critical appraisal and data extraction (CHARMS) – Moons et al 2014 PLOS Med</td>
</tr>
<tr>
<td>Reporting of systematic reviews</td>
<td>Transparent reporting of systematic reviews and meta-analysis (PRISMA) Moher et al. PLOS Med 2009</td>
</tr>
<tr>
<td>Assessing risk of bias of systematic reviews</td>
<td>Risk of bias in systematic reviews (ROBIS) Whiting et al. J Clin Epid 2015</td>
</tr>
</tbody>
</table>

---

Popularity of prediction models

- Thousands of clinical prediction models
- Also for same outcome / target population
  - ≥ 300 CVD; ≥ 100 brain injury; diabetes; breast cancer
Systematic reviews of prediction model studies

• Numerous methodology reviews in recent years:
  ➢ Mallett et al. BMC Med 2010
  ➢ Steyerberg et al. Epidemiology 2010

• Conclusions from methodology reviews:
  ➢ (Very) poor reporting
  ➢ (Very) poor methods
  ➢ Prediction model systematic reviews: Hardly ever assess risk of bias
Development and structure of PROBAST

- **Development:**
  - Delphi procedure with 42 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, ACROBAT-NRS and new Cochrane ROB
  - Five domains
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

→ (New) Cochrane Risk of Bias tool

Risk of bias and applicability

• Bias occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results.
  ➢ For PROBAST defined as “likelihood that a prediction model leads to distorted predictive performance for its intended use in the targeted individuals”. Predictive performance typically evaluated using calibration, discrimination, and (re)classification

• Applicability
  ➢ extent model from the primary study matches the systematic review question (participants, predictors or outcomes of interest).
**DOMAIN 1: Participant selection**

**A. Risk of Bias**

*Describe the sources of data and criteria for participant selection:*

<table>
<thead>
<tr>
<th></th>
<th>Dev</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Were participants defined and assessed in a similar way to participants in the model development study?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Were all inclusions and exclusions of participants appropriate?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 judgement

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

*Complete for each evaluation of a distinct model.*

<table>
<thead>
<tr>
<th><strong>Reaching an overall judgement about risk of bias of the prediction model evaluation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low risk of bias</strong></td>
</tr>
<tr>
<td><strong>High risk of bias</strong></td>
</tr>
<tr>
<td><strong>Unclear risk of bias</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Reaching an overall judgement about applicability of the prediction model evaluation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low concerns regarding applicability</strong></td>
</tr>
<tr>
<td><strong>High concerns regarding applicability</strong></td>
</tr>
<tr>
<td><strong>Unclear concerns regarding applicability</strong></td>
</tr>
</tbody>
</table>
Domain 1 (Participant selection)

1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?
2. Were all inclusions and exclusions of participants appropriate?
3. Were participants defined and assessed in a similar way to participants in the model development study?
Domain 2 (Predictors)

1. Were predictors defined and assessed in a similar way for all participants in the study?

2. Were predictors defined and assessed in a similar way to predictors in the development model? [VALIDATION ONLY]

3. Were predictor assessments made without knowledge of outcome data?

4. Are all predictors available at the time the model is intended to be used?
Domain 3 (Outcome)

1. Was the outcome determined appropriately?
2. Was a pre-specified or standard outcome definition used?
3. Were predictors excluded from the outcome definition?
4. Was the outcome defined and determined in a similar way for all participants?
5. Was the outcome defined and determined in a similar way to the outcome in the development model? [VALIDATION ONLY]
6. Was the outcome determined without knowledge of predictor information?
Domain 4 (Sample size and participant flow)

1. Were there a reasonable number of outcome events?
2. Was the time interval between predictor assessment and outcome determination appropriate?
3. Were all enrolled participants included in the analysis?
4. Were participants with missing data handled appropriately?
1. Were continuous and categorical predictors handled appropriately? [DEVELOPMENT ONLY]

2. Was selection of predictors based on univariable analysis avoided? [DEVELOPMENT ONLY]

3. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?

4. Were relevant model performance measures evaluated, e.g. (re)calibration and discrimination?

5. Was model overfitting and optimism in model performance accounted for? [DEVELOPMENT ONLY]

6. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? [DEVELOPMENT ONLY]
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

• Publications (planned for second half of 2016)
  ➢ Background paper with the tool
  ➢ Explanation and Elaboration (E&E)
Development of PROBAST (2)
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 Birmingham
- 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 Keele
- 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*, University of Bristol
- Robert Wolff*, Kleijnen Systematic Reviews
- Angela Wood, University of Cambridge

* denotes steering group members
Thank you and questions
Growth in clinical prediction model articles (Tufts CPM database)
Prediction models

• Combination of more than two predictors (variables, covariates, determinants)
• Model converts observed values in individual to probability in an individual of...
  ➢ having a particular disease \(\rightarrow\) diagnosis
  ➢ developing particular health state within a certain time (days? Weeks?) \(\rightarrow\) prognosis

• Possible outcomes:
  ➢ Death, complication, disease progression, pain, quality of life,
  ➢ hospitalisation, therapy response etc