PROBAST Example: prognostic model development study

Aslibekyan S, Campos H, Loucks EB, Linkletter CD, Ordovas JM, Baylin A. Development of a cardiovascular risk score for use in low- and middle-income countries. Journal of Nutrition. 2011;141(7):1375-80.

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. *The following table should be completed once per systematic review.*

Criteria	Specify your systematic review question
Intended use of model:	To predict the occurrence of myocardial infarction (MI) in a general population aged over 40 years.
***Participants** including selection criteria and setting:*	Healthy individuals from the general population, aged less than 75 years.
***Predictors** (used in prediction modelling), including types of predictors (e.g. history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/ prohibitions for specialized equipment):*	Traditional cardiovascular factors with additional predictors such as nutritional intake, physical activity, and social-economic status.
***Outcome** to be predicted:*	Myocardial infarction (MI) occurring during follow-up. MI presence is based on clinical criteria by at least two cardiologists.

Step 2: Classify the type of prediction model evaluation

Use the following table to classify the evaluation as model development, model validation or both. Different signalling questions apply for different types of prediction model evaluation. If the evaluation does not fit one of these classifications then PROBAST should not be used.

Classify the evaluation based on its aim			
Type of prediction study	**PROBAST boxes to complete**	**Tick as appropriate**	**Definition for type of prediction model study**
Development only	Development	×	Prediction model development without external validation. These studies may include internal validation methods, such as bootstrapping and cross-validation techniques.
Development and validation	Development and validation	✓	Prediction model development combined with external validation in other participants in the same article.
Validation only	Validation	×	External validation of existing (previously developed) model in other participants.

This table should be completed once for each publication being assessed and for each relevant outcome in your review.

Publication reference	Aslibekyan S et al. Development of a Cardiovascular Risk Score for Use in Low- and Middle-Income Countries. J Nutrition 2011
Models of interest	Cardiovascular Risk Score, based on WHO nutrient intake and physical activity recommendations and poverty line standards [score 1 in the paper; a second sore is based on additional biomarkers and socioeconomic markers]
Outcome of interest	Myocardial infarction

Step 3: Assess risk of bias and applicability

PROBAST is structured as four key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signalling questions to help make judgements. Signalling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signalling questions are phrased so that "yes" indicates absence of bias. Any signalling question rated as "no" or "probably no" flags the potential for bias; you will need to use your judgement to determine whether the domain should be rated as "high", "low" or "unclear" risk of bias. The guidance document contains further instructions and examples on rating signalling questions and risk of bias for each domain.

The first three domains are also rated for concerns for applicability (low/ high/ unclear) to your review question defined above.

Complete all domains separately for each evaluation of a distinct model. Shaded boxes indicate where signalling questions do not apply and should not be answered.

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: DEV: "The population of the Costa Rica Study included 4547 Hispanics who resided in the Central Valley of Costa Rica between 1994 and 2004. Cases of first nonfatal acute MI were ascertained by 2 independent cardiologists in the participating hospitals and deemed eligible if they met the WHO criteria, survived hospitalization, were under 75 y of age on the day of their first MI, and able to answer the questionnaire. Participants with a self-reported history of diabetes, hypertension, or current use of medication for chronic conditions were excluded from the derivation data set to avoid reverse causation. Eligible cases (n = 2273) were matched by 5-y age group, sex, and area of residence to population controls (n = 2274), identified randomly using data from the National Census and Statistics Bureau of Costa Rica. Women comprised 27% of all study participants (1209 total, 605 controls, and 604 cases)." VAL: The validation data set comprised all study participants excluded from the derivation data set, i.e. participants with a self-reported history of hypertension, diabetes, and/or hypercholesterolemia, thus less healthy individuals.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		N	N
1.2 Were all inclusions and exclusions of participants appropriate?		PN	PN
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*	**high**	**high**
Rationale of bias rating: Development: Participants who had died of fatal-MI were excluded as retrospective self-reported data could not be collected from these patients. The prediction model for non-fatal MI was based on selected healthier participants, including only those who survived a MI or did not develop a MI (controls). This is likely to have introduced bias as the study participants represent a selected lower risk sample of the original 'at risk of MI population'. Stating that the developed prediction model only predicts non-fatal MI does not really solve the issue since at the moment of prediction it is not possible to identify participants who will develop fatal-MI and so would be inappropriate for the model to be used on. Validation: The validation was also performed likely within a case-control framework, such that the same considerations as for the development set apply. Validation set, however, comprised of participants excluded from the derivation set, making it even more unclear what the exact case-control sampling scheme was.			
B. Applicability			
Describe included participants, setting and dates: As the prediction model was developed and validated with exclusion of fatal MI events (see above), and was then stated to predict only non-fatal MI, the model is not applicable for the intended review question.			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**high**	**high**
Rationale of applicability rating: Neither the development nor validation set included a general population. The development set excluded patients with comorbidity; the validation set excluded healthy patients.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: "Cardiovascular risk score components were selected based on a prior analysis of modifiable MI risk factors in our study population as well as international guidelines for healthy lifestyle. The selected risk score components showed expected associations with the risk of MI in our study population (Supplemental Table 1). Score 1 was derived to ensure a simple, low-cost risk estimation algorithm that could be adapted to a variety of populations, whereas score 2 was designed to include the most reliable measures of risk factors available for the Costa Rica Study population (Supplemental Table 2). After the cases were discharged from the hospital, all cases and controls received home visits, during which trained study workers collected lifestyle and medical history data, anthropometric measurements, and biological specimens. Information on diet, physical activity, smoking, alcohol intake, socioeconomic status, and medical history was collected using questionnaires. Dietary exposures were ascertained both via FFQ and biological markers, specifically adipose tissue concentrations of selected fatty acids. To avoid reverse causation and recall bias, data on exposures among cases were recorded as close to the time of MI as possible."			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		PY	PY
2.2 Were predictor assessments made without knowledge of outcome data?		N	PN
2.3 Are all predictors available at the time the model is intended to be used?		PN	PN
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*	**high**	**high**
Rationale of bias rating: The cases and controls were asked to fill in a questionnaire on, e.g. nutrition and physical activity (predictors), and thus participants and researchers collecting the information were aware of their outcome status. Data on predictors was collected after the outcome had occurred. For most predictors, participants were asked about retrospective exposure and so this information would be available at the time the model is intended to be used. Also, for some measurements (e.g. waist: hip ratio) these were measured after the outcome. Values of these are likely to have changed from when the model is intended to be used (although the intended time frame for use of the model is not clear).			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**
Rationale of applicability rating: The predictors considered for use in the model were the same as those defined in the review question. The review question did not specify timing or assessment of predictors and so there is no concern that these do not match.			

DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: DEV: "Cases of first nonfatal acute MI were ascertained by 2 independent cardiologists in the participating hospitals and deemed eligible if they met the WHO criteria, survived hospitalization, were under 75 y of age on the day of their first MI, and able to answer the questionnaire." VAL: All patients with non-fatal MI that were excluded in the development dataset had some type of comorbidity, e.g. diabetes or hypertension, in the past.			
		Dev	Val
3.1 Was the outcome determined appropriately?		PY	NI
3.2 Was a pre-specified or standard outcome definition used?		N	NI
3.3 Were predictors excluded from the outcome definition?		Y	PY
3.4 Was the outcome defined and determined in a similar way for all participants?		PY	PY
3.5 Was the outcome determined without knowledge of predictor information?		PY	PY
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		NI	NI
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*	**unclear**	**unclear**
Rationale of bias rating: There was insufficient information on how the outcome was defined and timing between predictor assessment and outcome determination to be able to rate the risk of bias for this domain.			
B. Applicability			
At what time point was the outcome determined: The study used a case-control design selecting patient who had experienced the outcome (non-fatal MI) and those who had not; the time frame between predictor assessment and outcome occurrence is not clear. *If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:* No composite outcome			
Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**high**	**high**
Rationale of applicability rating: The review question is interested in prediction of MI. This study only considered non-fatal MI and so findings are not directly applicable to the review question.			

DOMAIN 4: Analysis
Risk of Bias
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): DEV: Numbers of participants: 4,547 total source population, 839 cases and 839 controls included Number of candidate predictors: Unclear Number of outcome events: 839 Number of events per candidate predictor: unclear
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Predictor selection: "Cardiovascular risk score components were selected based on a prior analysis of modifiable MI risk factors in our study population as well as international guidelines for healthy lifestyle. The selected risk score components showed expected associations with the risk of MI in our study population." Model development: "The derivation data sets were developed using the complete case method, i.e. participants with missing values for any of the covariates were excluded from the analysis. Additionally, participants with a self-reported history of diabetes, hypertension, or current use of medication for chronic conditions were excluded from the derivation data set to avoid reverse causation. After restriction, remaining cases were rematched to controls on age, sex, and area of residence to preserve the study design." "For each version of the score, conditional logistic regression models were fit with MI as the outcome and cardiovascular risk score components as predictors, while matching on age, sex, and area of residence to control for potential confounding by these demographic factors. The obtained regression coefficients for each score component were then multiplied by the values of cardiovascular risk score components and summed across components to produce the final value of the cardiovascular risk score. Thus, the final cardiovascular risk score value represents a weighted sum of individual risk score components. Two regression models, each adjusted for age, sex, and area of residence, were fit to assess the discriminatory ability of score 1 with indicator variables corresponding to quintiles of each score's distribution. A test for linear trend was performed on categorical models, using the median value of each quintile as a continuous predictor." Risk groups: "All dietary variables included in score 1 (trans fats, polyunsaturated fats, saturated fats, cholesterol fiber, and folate) were included as categorical variables according to the quintile of intake." "Being physically active was defined as expending >10% of daily energy in the performance of moderate- and high-intensity physical activities (at least 4 times the basal metabolism rate). (…) smoking was defined as a dichotomous variable (currently smoking vs. not) and alcohol intake was measured in grams per day and defined as a categorical variable with the following cutoffs: 0 (not drinkers), 0.1–5.0, 5.1–10, over 10. For both scores, participants were classified as healthy if their waist: hip ratio value lay below the cutoff of ≤0.90 for men and ≤0.85 for women as per WHO guidelines. (…) Socioeconomic status was classified as low if a participant's self-reported annual income fell below the threshold of twice the national poverty line for the year of recruitment into the study." Model performance: Model performance was not quantified. Optimism: Optimism was not considered.
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): External VAL: "ROC curves were constructed to evaluate the performance of both scores in the validation data set (Fig. 2A,B)."

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: No measures of performance of the derivation model were presented. Discrimination (AUC) was reported for the validation data set only.			
Describe any participants who were excluded from the analysis: DEV: To derive score 1, 456 of the original 4547 participants were excluded due to missing covariate values, 2167 were excluded due to history of chronic disease, and 246 were lost to the rematching process, yielding the final sample size of 1678 (839 cases and 839 controls). VAL: Not reported			
Describe missing data on predictors and outcomes as well as methods used for missing data: Not reported			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		N	NI
4.3 Were all enrolled participants included in the analysis?		N	NI
4.4 Were participants with missing data handled appropriately?		N	NI
4.5 Was selection of predictors based on univariable analysis avoided?		PY	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		N	NI
4.7 Were relevant model performance measures evaluated appropriately?		N	N
4.8 Were model overfitting and optimism in model performance accounted for?		PN	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		PY	
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*	**high**	**high**
Rationale of bias rating: Many continuous predictors were recoded to categorical or binary variables. A large proportion of the potentially eligible population was excluded from the analysis due to missing covariate values or rematching. Participants with missing data were excluded from the analysis. Model performance was not evaluated appropriately (no data for development and performance only for validation), overfitting and optimism was not accounted for and complexities in the data were not considered.			

Step 4: Overall assessment

Use the following tables to reach overall judgements about risk of bias and concerns for 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 for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have **low concerns for applicability**.
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have **high concerns for applicability**.
Unclear concerns for applicability	If unclear concerns (but no "high concern") for applicability for at least one domain, the prediction model evaluation is judged to have **unclear concerns for 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)*	**high (Dev)** **high (Val)**
Summary of sources of potential bias: Both development and validation studies appear to use a (non-nested) case-control design which is likely to have introduced bias. Data on predictors were collected retrospectively with knowledge of the outcome. There was insufficient information on how the outcome was classified. There were multiple limitations with the analysis including a large proportion of missing data that was excluded from the analysis, inappropriate categorisation of continuous variables, and lack of results on model performance.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**high (Dev)** **high (Val)**
Summary of applicability concerns: There were concerns for the applicability of both the population and outcome. The population of interest was a general population aged 40 and over; the study included restricted populations for both development and validation data. The review question is interested in prediction of MI. This study only considered non-fatal MI.		