PROBAST Example: prognostic model development and validation study.

Perel P, Prieto-Merino D, Shakur H, Clayton T, Lecky F, Bouamra O, et al. Predicting early death in patients with traumatic bleeding: development and validation of prognostic model. BMJ. 2012;345:e5166

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 identify patients with traumatic bleeding who are at risk of early death and hence to inform timely care decisions
***Participants** including selection criteria and setting:*	Trauma patients presenting at accident and emergency (hospital setting)
***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):*	Age, sex, type of injury, time since injury, blood pressure, heart rate, respiratory rate, capillary refill time, Glasgow coma score. Predictors measured at entry to the emergency department.
***Outcome** to be predicted:*	Early all-cause mortality

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	Perel P et al. Predicting early death in patients with traumatic bleeding: development and validation of prognostic model. BMJ 2012 Aug 15;345:e5166. doi: 10.1136/bmj.e5166
Models of interest	CRASH-2
Outcome of interest	All-cause mortality within 4 weeks of injury

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: Patients from the CRASH-2 trial – 20,127 trauma patients with or at risk of significant bleeding within eight hours of injury. Took place in 274 hospitals in 40 countries. Randomized trial comparing tranexamic vs. placebo. VAL: Data from Trauma Audit and Research Network (TARN) including 60% of relevant hospitals in England and Wales and some hospitals in Europe, collected between 2000 and 2008, were used. TARN collects data on patients who arrive at hospital alive and who meet any of the following criteria: death from injury at any point during admission; hospital stay >3 days; need for intensive or high dependency care; need for between hospital transfer for specialist care. The validation data set excluded patients with isolated limb injuries and those over 65 years with isolated femoral neck or pubic ramus fracture.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		PY	Y
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*	**low**	**low**
Rationale of bias rating: DEV: No information on inclusions and exclusions from the CRASH-2 trial (development dataset), unlikely to have a major impact. The CRASH-2 trial is referenced (Lancet 2010;376:23-32). VAL: (risk of) significant bleeding was not a specific entry criterion in the validation dataset, so researchers selected only patients with an estimated blood loss of at least 20%. This sounds like a reasonable and sensible approach.			
B. Applicability			
Describe included participants, setting and dates: As before, no dates reported for the CRASH-2 trial (development dataset) but the trial is referenced.			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**
Rationale of applicability rating: Participants and setting seem to fit review question			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: DEV: Demographic characteristics (age and sex), characteristics of injury (type of injury and time from injury to randomisation), and physiological variables (Glasgow coma score, systolic blood pressure, heart rate, respiratory rate and central capillary refill time). Predictors seem to have been pre-specified (within CRASH-2), all predictors were measured at baseline (taken from patients' entry forms completed before randomisation). VAL: The physiological data available in TARN are identical to those in CRASH-2, in that for every patient the heart rate, systolic blood pressure, Glasgow coma score, respiratory rate, and capillary refill time on arrival are entered by the hospital data coordinators. All predictors were measured on arrival.			
		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?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*	**low**	**low**
Rationale of bias rating: No major issues identified.			
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: All predictors were measured at baseline (taken from patients' entry forms completed before randomisation, for CRASH-2 and by hospital data co-ordinators at time of arrival for TARN).			

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: The primary outcome was all cause mortality within four weeks of injury. The development dataset was part of an RCT so some patients received the trial intervention between predictor measurement and outcome measurement. This was accounted for in the model by including treatment as a covariate in the model. For model development (CRASH-2 data), the outcome was defined as all-cause mortality, in-hospital or within 28 days of discharge. Validation uses TARN data, which records mortality at any point during admission as an outcome.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		PN	PN
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*	**low**	**low**
Rationale of bias rating: Potential for the outcome ("all-cause mortality") to be determined with predictor information available and the difference in follow-up definition (in-hospital or within 28 days of discharge – development, versus at any point during admission [duration not defined] – validation) are unlikely to have a major influence.			
B. Applicability			
At what time point was the outcome determined: Patients' outcomes were recorded at discharge, at death in hospital, or 28 days after injury, whichever occurred first. *If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:* N/A			
Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**
Rationale of applicability rating: Early all-cause mortality was the outcome of interest for the review. The outcome of the primary study matches the outcome of interest of the review.			

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: 20,127 participants, 9 candidate predictors, 3076 deaths

VAL: 14,220 trauma patients and 1,765 deaths

Describe how the model was developed (predictor selection, optimism, risk groups, model performance):

All candidate predictors were initially included in multivariable logistic regression. Analyses were adjusted for treatment, by including treatment arm as a co-variable. A variable for economic region (low, middle, or high-income country, as defined by the World Bank) was also included. The final model was developed using a backwards stepwise approach. Optimism was assessed by bootstrap re-sampling.

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 validation used the same variables as were included in the derivation model except hours since injury, as this variable had a very large number of patients with missing data.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:

Discrimination (c-statistic) and calibration.

Describe any participants who were excluded from the analysis:

DEV: "Complete case analysis as amount of missing data was very low"

Describe missing data on predictors and outcomes as well as methods used for missing data:

DEV: Data were missing for <1% of participants for age, systolic blood pressure, heart rate, respiratory rate and Glasgow coma score. Data for capillary refill time were missing for 3% of patients. No information on missing data for sex, type of injury and time from injury to randomisation.

VAL: Authors conducted "multiple imputations to substitute the missing values of the predictors included in the prognostic model by using the procedure of imputation by chained equations in Stata Release 11".

		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?		Y	Y
4.3 Were all enrolled participants included in the analysis?		N	Y
4.4 Were participants with missing data handled appropriately?		NI	Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	Y
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
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)*	**low**	**low**

Rationale of bias rating:

In the development of the model complete case analysis was used, however, the frequency of missing data was low (RCT data), therefore the risk of bias was still considered low. No other problematic issues are present with respect to this domain.

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)*	**low (Dev)** **low (Val)**
Summary of sources of potential bias: DEV: Model was developed on a large prospective, pragmatic, randomised controlled trial. Treatment was accounted for in the model. Large dataset with a large number of events. Complete case analysis was performed, but the frequency of missing was low. Backward selection after the inclusion of all predictors is an accepted approach. VAL: No major problems.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**low (Dev)** **unclear (Val)**
Summary of applicability concerns: DEV: The model has been developed on data from a randomized clinical trial. The concern is that RCT may include a selective population, typical younger, less comorbidity and medication use. In this case, these concerns may be limited given the acute onset of the condition (trauma) and it was a large pragmatic trial.		

VAL: Data from a large ongoing registry of trauma patients with similar data collection. However, one predictor is missing in the validation dataset.