PROBAST Example: prognostic model development and validation study

Han J, King NK, Neilson SJ, Gandhi MP, Ng I. External validation of the CRASH and IMPACT prognostic models in severe traumatic brain injury. Journal of Neurotrauma. 2014;31(13):1146-52.

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 outcomes in patients with severe traumatic brain injury (TBI) on admission to intensive care unit (ICU)
***Participants** including selection criteria and setting:*	Patients with severe TBI (Glasgow Coma Scale (GCS) ≤8) admitted to ICU
***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):*	Patient demographics such as age and sex; aetiology and severity of brain injury; CT results; physiological data, e.g. blood pressure. Measurements taken prior to any surgical intervention or outcome occurring.
***Outcome** to be predicted:*	*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	Han J et al. External Validation of the CRASH and IMPACT Prognostic Models in Severe Traumatic Brain Injury. J Neurotrauma 2014. 31:1146-1152
Models of interest	CRASH CT Model
Outcome of interest	Unfavourable outcome after six months

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: "Prospective data were collected from a consecutive series of 300 patients with severe TBI (GCS of ≤ 8) admitted between February 2006 and December 2009 to a single neurosurgical intensive care unit in the National Neuroscience Institute (NNI), Singapore. The NNI database comprises patient demographics (age, sex, and race), various etiologies of brain injury (road traffic accidents, falls, assaults, and others), severity of injury (GCS, pupillary reactivity, and other major extracranial injury), brain CT characteristics and physiologic data such as hypoxia, hypotension, blood results, and glucose levels."			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			Y
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*		**low**
Rationale of bias rating: Consecutive series of patients included. No concerns on eligibility. Severe TBI is defined using validated score.			
B. Applicability			
Describe included participants, setting and dates: Consecutive series of patients of 300 patients with severe TBI (GCS≤8) admitted between February 2006 and December 2009 to a single neurosurgical intensive care unit in the National Neuroscience Institute Singapore.			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*		**low**
Rationale of applicability rating: Included patients appear representative of the population specified in the review question. Applicability may be affected by geographical and healthcare setting as patients were all from a single neurosurgical intensive care unit in the National Neuroscience Institute (NNI), Singapore.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: "[CRASH] had two types of models: the basic model, which had four predictors – age, GCS, pupil reactivity, and the presence of major extracranial injury – and the CT model, which added on results of CT derived from the first CT scan performed post-injury. The selected CT criteria included petechial hemorrhages, obliteration of the third ventricle or basal cisterns, subarachnoid bleed, midline shift, and unevacuated hematoma." "The CRASH models were externally validated in the IMPACT data set of 8509 patients with moderate and severe TBI from 11 studies conducted only in high-income countries, and also did not include the variables 'major extracranial injury' and 'petechial hemorrhages', as these were not available in the validation sample."			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			PY
2.2 Were predictor assessments made without knowledge of outcome data?			PY
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*		**low**
Rationale of bias rating: In the absence of content experts, we have assigned PY for 2.1 as GCS and Marshall CT classification use pre-defined scores. Pupil reactivity is a standardised measure.			
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**
Rationale of applicability rating: Included predictors appear representative of the predictors specified in the review question.			

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: "There were two defined outcomes for each of the models: one was mortality at 14 days, and the other was unfavorable outcome at 6 months, defined by the authors based on the Glasgow Outcome Scale (GOS) as severe disability, vegetative state, or death." "Data were prospectively collected for both admission characteristics prior to any surgical intervention and outcome at 14 days and 6 months. Additional radiological information that was necessary to calculate the CRASH and IMPACT prognostic models, as will be described, were retrospectively obtained."			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			PY
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*		**low**
Rationale of bias rating: Signalling question 3.5 rated as PY as some retrospective information for required for model predictors.			
B. Applicability			
At what time point was the outcome determined: Mortality after 14 days *If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:* Not a composite outcome			
Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*		**low**
Rationale of applicability rating: Good match to the pre-defined 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): 300 participants. 143 deaths within 14 days.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Not applicable – model validation study.			
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 of CRASH CT model including temporal and geographical validation. CRASH-2 model development dataset included only 2 patients from Singapore.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: "Validation was based on measures of discrimination and calibration as recommended for validation of prognostic models. Discrimination was assessed using the AUC and 95% confidence interval (CI). An AUC ≤ 0.5 is considered to represent a non-discriminative model, whereas an AUC ≥ 0.8 is considered to represent adequate discriminative ability by the model. Calibration was assessed using the H-L goodness-of-fit test, which tests the null hypothesis that the model's estimates fit the observed data. This test divides subjects into deciles based on predicted probabilities, and then computes a $\chi 2$ from observed and expected frequencies. A p value is then calculated from the $\chi 2$ distribution, to test the fit of the logistical model. If the p value of H-L goodness-of-fit test is > 0.05, we fail to reject the null hypothesis that there is no difference, implying that the model's estimates fit the observed data at an acceptable level. Calibration in this study was also assessed with a Cox calibration analysis using a logistical regression model to obtain two derivatives: the slope and the intercept. The slope, as a coefficient of the logit of predicted probability, reflects the degree of variation in predictions. In an ideal model, the slope is equal to 1. The intercept, as a measure of overall calibration, indicates whether the predictions are systematically too low or too high, and this should ideally be zero. All the analyses were done on the NNI database using the Stata statistical software. (Release 11, StataCorp LP, USA)." No adjustment for optimism as external validation.			
Describe any participants who were excluded from the analysis: "The NNI data were complete for all the predictors required to calculate the CRASH and IMPACT models, except for missing blood glucose levels in 36 patients." Blood glucose not included in CRASH CT model.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Blood glucose not included in CRASH CT model, so no missing data.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			NI
4.2 Were continuous and categorical predictors handled appropriately?			NI
4.3 Were all enrolled participants included in the analysis?			Y
4.4 Were participants with missing data handled appropriately?			Y
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?			PY
4.7 Were relevant model performance measures evaluated appropriately?			Y
4.8 Were 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 the reported multivariable analysis?			
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*		**low**

Rationale of bias rating: Logistic regression model which does not account for any censored data. However, no loss to follow up, so no censored data. States all patients in database followed for 14 day and 6 month outcomes. Cohort study without sampling of controls.

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**
Summary of sources of potential bias: Outcome might have been available when assessing predictors retrospectively.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**Low**
Summary of applicability concerns: No concerns.		