PROBAST Example: diagnostic model development study

Rietveld RP, ter Riet G, Bindels PJ, Sloos JH, van Weert HC**.** Predicting bacterial cause in infectious conjunctivitis: cohort study on informativeness of combinations of signs and symptoms. BMJ. 2004;329(7459):206-10.

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:	Diagnosis of a bacterial cause in patients presenting with an acute infectious conjunctivitis
***Participants** including selection criteria and setting:*	Adult patients presenting with recent onset (<7 days) red eye and discharge or sticking eyelids to their general practitioner.
***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):*	Signs and symptoms measured at time of presentation.
***Outcome** to be predicted:*	Bacterial infection (confirmed by culture)

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	Rietveld RP et al. Predicting bacterial cause in infectious conjunctivitis: cohort study on informativeness of combinations of signs and symptoms. BMJ 2004; 329: 206-210, doi:10.1136/bmj.38128.631319.AE
Models of interest	"Final model" (combination of early morning glued eye(s), itch, and a history of conjunctivitis)
Outcome of interest	Presence of positive bacterial culture

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: "We asked nine designated general practitioners, working in 25 care centres with a total of 41 general practitioners, in the Amsterdam and Alkmaar region to include patients with a red eye and either (muco)purulent discharge or sticking of the eyelids. The exclusion criteria were age younger than 18 years, pre-existing symptoms for longer than seven days, acute loss of vision, wearing of contact lenses, use of systemic or local antibiotics within the previous two weeks, ciliary redness, eye trauma, and a history of eye surgery. All eligible patients were referred to one of the nine designated general practitioners for enrolment in the study."			
		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; exclusions appear appropriate.			
B. Applicability			
Describe included participants, setting and dates: Consecutive patients presenting to their general practitioner. Mean age around 44, 59% female. Inclusion between Sept 1999 and December 2002			
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.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: "At inclusion of each participant, general practitioners completed a standardised questionnaire and physical examination (index tests). The questionnaire contained questions about medical history (self-reported), duration of symptoms (days), self-medication and self-treatment, itching, burning sensation, foreign body sensation, and the number of glued eyes in the morning (0, 1, or 2). The physical examination included investigation of the degree of redness (peripheral, whole conjunctiva, or whole conjunctiva and pericorneal), the presence of periorbital oedema, the kind of discharge (watery, mucous, or purulent), and bilateral involvement…. The general practitioners did not receive the culture results" "For each patient one eye was designated as the 'study eye'." In the case of two diseased eyes, the diseased eye with worse signs or symptoms was the study eye. In the case of two equally affected eyes, the first affected eye was the study eye."			
		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?		Y	
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: Predictors were assessed by 41 including GPs. This may have led to some differences in assessment, but the predictors are routinely assessed by GPs and so differences are considered unlikely. GPs were blinded for the outcome (culture result).			
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: Predictors were assessed at presentation and defined in a standard way. They appear to match 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:

Diagnosis of a bacterial origin of acute infectious conjunctivitis:

"The general practitioner then took one conjunctival sample of each eye for a bacterial culture (reference standard). The general practitioners did not receive the culture results, and the microbiologist who analysed the cultures had no knowledge of the results of the index tests."

"General practitioners took one sample of the conjunctiva of each eye by rolling a cotton swab (Laboratory Service Provider, Velzen-Noord, Netherlands) over the conjunctiva of the lower fornix. They put the swabs into transport medium and sent them to the investigating laboratory in Alkmaar. Directly after arrival, we inoculated the swabs on to blood agar enriched with 5% sheep blood, MacConkey agar, and chocolate agar. All media were made at the laboratory with standard ingredients (Becton Dickinson, Cockeysville, MD, USA). After standard inoculation, we incubated the blood agar and MacConkey agar plates for 48 hours at 35°C; we incubated the chocolate agar plates for the same period and at the same temperature, but in a 7% CO_2 atmosphere. We analysed cultures daily according to the guidelines of the American Society for Microbiology. We identified all pathogens by using routine standard biochemical procedures. Colonies suspected to be pathogens were selected and investigated by Gram stain. Depending on the results of the Gram stain, we did additional tests. In the case of Gram positive cocci, we did a catalase test followed by, for example, a coagulase test (staphylococci) or an optochine (pneumococci) test. In the case of Gram negative rods or cocci, we did sugar tests."

		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?		Y	
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:

Culture is not perfect in establishing whether there is a bacterial cause. However, it is the best available method currently available. The authors use appropriate methods to obtain culture results and explicitly state that the microbiologist was unaware of the index test results.

B. Applicability

At what time point was the outcome determined:

Outcome (bacterial infection diagnosed by culture) was determined at the same moment as the measurement of the predictors which is in line with the diagnostic nature of the study

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**	

Rationale of applicability rating:

The outcome is the same as that specified in 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): 182 participants enrolled in the study, 177 participants included in analysis, 12 candidate predictors, 57 bacterial conjunctivitis			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): "We assessed the associations between findings from the index tests and the presence of a positive bacterial culture in the study eye by using a stepwise forward logistic regression analysis. The dependent variable was the presence or absence of a bacterium. We entered variables with a univariate P value of ≤ 0.10 into the model. We considered variables with a multivariate P value of < 0.15 to be independent indicators of the presence of bacteria and retained them in the final model. We modelled determinants with more than two categories as dummy variables. We assessed all second order interactions between the variables entered into the final model. We deemed interaction to be present if the P value associated with an interaction term was < 0.05."			
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): "Validation of this model with the bootstrap technique showed hardly any indication of undue influence by particular patients (corrected 95% confidence interval of area under curve 0.63 to 0.80)." However, it seems likely that the authors have not included all statistical modelling steps in the bootstrapping, and merely bootstrapped the final model. This would explain the negligible difference observed in the area under the curve, which is surprising given the small sample size.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: "We quantified the ability of the final model to discriminate between patients with and without a positive bacterial culture by using the area under the receiver operating characteristics curve with 95% confidence intervals. We quantified the reliability or calibration of the model by using the Hosmer-Lemeshow goodness of fit test. Finally, we bootstrapped the receiver operating characteristics curve a thousand times to counteract potential undue influence of atypical patients on the predictions of the final model." Mean predicted probabilities with corresponding confidence intervals across subgroups and the corresponding observed outcome frequencies are reported in table 4.			
Describe any participants who were excluded from the analysis: Five patients excluded (3 index test incomplete, 2 results of culture unknown)			
Describe missing data on predictors and outcomes as well as methods used for missing data: Not described			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		N	
4.2 Were continuous and categorical predictors handled appropriately?		NI	
4.3 Were all enrolled participants included in the analysis?		PN	
4.4 Were participants with missing data handled appropriately?		N	
4.5 Was selection of predictors based on univariable analysis avoided?		N	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	
4.7 Were relevant model performance measures evaluated appropriately?		PY	
4.8 Were model overfitting and optimism in model performance accounted for?		N	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		N	
Risk of bias introduced by the analysis	**RISK:**	**high**	

	(low/ high/ unclear)		
Rationale of bias rating: At least 12 different candidate predictors considered but only 57 participants with an event. Five patients (out of 182) had missing data and were not included in the analysis. This is unlikely to have substantially altered the results. Selection of predictors was based on univariable analysis. Simplified clinical score used whole number scores and assigned weights of the predictors do not correspond to the results in the final multivariable analysis due to rounding.			

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: The assessment identified potential for risk of bias in domain 4: At least 12 different candidate predictors considered but only 57 participants with an event. Selection of predictors based on univariable analysis. Assigned weights of the predictors do not correspond to the results in the final multivariable analysis.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**low**
Summary of applicability concerns: No concerns		