PROBAST Example: Diagnostic model validation study

Oudega R, Hoes AW, Moons KG. The Wells rule does not adequately rule out deep venous thrombosis in primary care patients. Ann Intern Med. 2005;143:100-7. [PMID: 16027451].

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:	Validation of the original (categorised) diagnostic Wells rule to exclude deep vein thrombosis (DVT) in the primary care setting
***Participants** including selection criteria and setting:*	People suspected of having DVT in the primary care 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):*	The nine clinical features included in the Wells rule (Lancet. 1995 May 27;345(8961):1326-30), measured when diagnosis of DVT was suspected primary care patients.
Outcome to be predicted:	Absence or presence of DVT

Step 2: Classify the type of prediction model evaluation

Use the following table to classify the evaluation as model development, model validation or model update, or combination. 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	Oudega R et al. The Wells rule does not adequately rule out deep venous thrombosis in primary care patients. Ann Intern Med 2005;143:100-7
Models of interest	The so-called Wells DVT diagnostic rule
Outcome of interest	Absence or presence of DVT

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 regarding 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:

"From 1 January 2002 to 1 March 2003, we prospectively identified consecutive adult (age 18 years) patients who visited their primary care physician and were clinically suspected to have DVT. The suspicion of DVT was based on swelling, redness, or pain in the legs that had been present for 30 days or less. Patients in whom pulmonary embolism was also suspected were excluded. We conducted the study in a large, circumscribed geographic region of The Netherlands that included 3 non-teaching hospitals with a catchment area of 110 primary care physicians and about 250 000 inhabitants. The diagnostic work-up, including study forms for documentation of patient information and referral of patients for further work-up, was integrated into the regular work-up of primary care patients. All 110 primary care physicians in the catchment area participated in the study. On average, 9 to 10 patients with suspected DVT per physician per year were included. The study area, participating general practitioners, and study patients can be considered representative of the country."

		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:

A cohort approach was used in which study subjects were included based on their presence of at least one of three predefined signs of DVT presence. Exclusion of suspicion of PE, which clinically may demand for another diagnostic work-up, were appropriately excluded.

B. Applicability

Describe included participants, setting and dates:

See above

Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*		**low**

Rationale of applicability rating:

The study population of the individual paper matches the targeted population of 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:

"After informed consent was obtained, the primary care physician systematically documented information on the patient's history and physical examination by using a standard form on which the items and possible answers were specified.

Patient history included sex, presence of previous DVT, family history of DVT, history of cancer (active cancer in the past 6 months), immobilization for more than 3 days, recent surgery (within the past 4 weeks), and duration of the 3 main symptoms (a painful, red, or swollen leg). Physical examination included the presence of tenderness along the deep venous system, distention of collateral superficial (nonvaricose) veins, pitting edema, swelling of the affected limb, and a difference between the circumference of the 2 calves. The 9 items of the Wells rule (**Table 1**) were also included in the standard form for patient history and physical examination.

All of the items were measured according to the original Wells rule (5, 21, 22). The last item of the rule—presence of an alternative diagnosis— has never been unambiguously defined and often causes controversy among users of the rule (23). In our study, physicians were asked to give their own assessment of the patient's probability of having DVT by using a score of 1 to indicate high probability of DVT, no alternative diagnosis likely; 2 to indicate moderate probability of DVT, alternative diagnosis possible; or 3 to indicate low probability of DVT, alternative diagnosis certain. To tailor the judgment of the physician on this item, 7 common alternative diagnoses for patients with suspected DVT were provided on the study form. If a low or moderate probability was assigned to a patient, we subtracted 2 points from the Wells score in the analysis.

The items of the Wells score were documented only for the purposes of our study and were not used to determine a patient's further diagnostic work-up."

		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y
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:

The items (predictors) of the Wells DVT rule were assessed similarly for all study subjects, as it was implemented in the routine care procedures. They were clearly measured before the reference standard was executed, and at the right moment of the diagnostic work-up (i.e. at the intended moment of the application of the Wells DVT rule).

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:

The definition, measurement and timing of the predictors matches the review question and targeted context (primary care).

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:

"Finally, as a reference test to determine the true presence or absence of DVT, each patient underwent real-time, B-mode compression ultrasonography of the lower extremities by using a 5- to 7.5-MHz linear-array sonographic scanner (system V, GE Healthcare, Milwaukee, Wisconsin). Color duplex imaging was used to identify venous vessels and venous flow patterns and was repeated 7 days later in patients with a normal result on ultrasonography. Proximal DVT was considered present if results of 1 of the 2 ultrasonographic tests were abnormal. The ultrasonogram was considered abnormal when the common femoral vein, the superficial femoral vein, or the popliteal vein up to the trifurcation was not completely compressible. Obstruction of the iliac veins was tested with color duplex imaging.

The person who performed the imaging was blinded to patient history, findings on physical examination, and results of D-dimer testing."

Time interval between predictor assessment and outcome determination not explicitly stated but it was clearly stated: "after history taking and physical examination".

		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?			PY
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*		**low**

Rationale of bias rating:

All items of this domain were explicitly or implicitly addressed

B. Applicability

At what time point was the outcome determined:

Time point not explicitly stated but it was stated: "after history taking and physical examination"

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 definition, measurement and timing of the outcome matches the review question and targeted context (primary care).

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor: Total: 1295 participants, 289 events Wells score ≥3 (high risk): 467 participants, 175 events			
Describe how the model was developed (for example in regards to modelling technique (e.g. survival or logistic modelling), predictor selection, and risk group definition): N/A – validation only			
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): "We validated the categorized Wells rule by quantifying per score the overall proportion of patients and the actual prevalence of DVT. The sensitivity, specificity, negative predictive value, and likelihood ratio for a negative test result (negative likelihood ratio = [1–sensitivity]/specificity), with corresponding 95% CIs, were manually calculated. Because ruling out DVT is the main purpose of both the Wells rule and the D-dimer assay, the positive predictive value and the likelihood ratio for a positive test result are not presented. For similar reasons, only the diagnostic accuracy variables for the threshold between the low Wells score and the medium and high scores are presented."			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism: Diagnostic performance reported in Table 3 of the paper (Probability of DVT)			
Describe any participants who were excluded from the analysis: "Of the 1326 patients considered for inclusion, 31 were excluded because the physician did not supply sufficient information about fulfilment of the inclusion and exclusion criteria"			
Describe missing data on predictors and outcomes as well as methods used for missing data: "If data on a patient were missing, the research physician contacted the patient's primary care physician immediately after the study forms were received. Nonetheless, 1 or more study variables was missing for 127 patients (45 patients had a missing value on 1 or more of the 9 items of the Wells rule). The missing data per study variable ranged from only 0.4% to 4% (the lowest was 5 missing responses for "paralysis or recent immobilization of the lower extremity," and the highest was 52 missing responses for "thrombosis in family"). We imputed the missing values of all study variables by using a linear regression method with the addition of a random error term (SPSS software, version 12.0 for Windows, SPSS, Inc., Chicago, Illinois). We also performed a complete case analysis. Because this yielded similar results to those obtained after imputation, only the results after imputation are presented."			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			Y
4.2 Were continuous and categorical predictors handled appropriately?			Y
4.3 Were all enrolled participants included in the analysis?			N
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?			N
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 multivariable analysis?			
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*		**Low**

Rationale of bias rating:

More than 100 patients with the outcome present were included in this validation study. Thirty-one patients were “excluded because the physician did not supply sufficient information about fulfilment of the inclusion and exclusion criteria”. No indication that these excluded participants were a ‘completely random’ subset of the original study population.

Missing data were imputed.

Calibration, discrimination, (re)classification and net benefit of the Wells rule (uncategorised) were not reported. Only diagnostic classification measures of the categorised Wells rule (as originally was published) is reported, but this was conform the review question: validating the original Wells DVT rule (which was indeed dichotomised to exclude DVT in suspected patients). Hence, the overall rating of this domain is LOW. Otherwise, this validation study did not validate a continuous prediction model (and thus did not report the necessary calibration and discrimination measures), and in that strict sense the overall rating of this domain could be HIGH.

Step 4: Overall assessment

Use the following tables to reach overall judgements about risk of bias and concerns regarding 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 a 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 regarding applicability	If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have **low concerns regarding applicability**.
High concerns regarding applicability	If high concerns regarding applicability for at least one domain, the prediction model evaluation is judged to have **high concerns regarding applicability**.
Unclear concerns regarding applicability	If unclear concerns (but no "high concern") regarding applicability for at least one domain, the prediction model evaluation is judged to have **unclear concerns regarding 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**
Summary of sources of potential bias: See above explanations and thus an overall score of low		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**Low**
Summary of applicability concerns: See above explanations and thus an overall score of low		