## PROBAST

(Prediction model study Risk Of Bias Assessment Tool)

Published in Annals of Internal Medicine (freely available):

1. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies
2. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration

**What does PROBAST assess?**

PROBAST assesses both the *risk of bias* and *concerns regarding applicability* of a study that evaluates (develops, validates or updates) a multivariable diagnostic or prognostic prediction model. It is designed to assess primary studies included in a systematic review.

*Bias* occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results. For the purpose of prediction modelling studies, we have defined *risk of bias* to occur when shortcomings in the study design, conduct or analysis lead to systematically distorted estimates of a model's predictive performance or to an inadequate model to address the research question. Model predictive performance is typically evaluated using calibration, discrimination and sometimes classification measures, and these are likely inaccurately estimated in studies with high risk of bias. *Applicability* refers to the extent to which the prediction model from the primary study matches your systematic review question, for example in terms of the participants, predictors or outcome of interest.

A primary study may include the development and/or validation or update of more than one prediction model. A PROBAST assessment should be completed for each distinct model that is developed, validated or updated (extended) for making individualised predictions. Where a publication assesses multiple prediction models, only complete a PROBAST assessment for those models that meet the inclusion criteria for your systematic review. Please note that subsequent use of the term "model" includes derivatives of models, such as simplified risk scores, nomograms, or recalibrations of models.

PROBAST is not designed for all multivariable diagnostic or prognostic studies. For example, studies using multivariable models to identify predictors associated with an outcome but not attempting to develop a model for making individualised predictions are not covered by PROBAST.

PROBAST includes four steps.

| Step | Task | When to complete |
|---|---|---|
| **1** | Specify your systematic review question(s) | Once per systematic review |
| **2** | Classify the type of prediction model evaluation | Once for each model of interest in each publication being assessed, for each relevant outcome |
| **3** | Assess risk of bias and applicability | Once for each development and validation of each distinct prediction model in a publication |
| **4** | Overall judgment | Once for each development and validation of each distinct prediction model in a publication |

If this is your first time using PROBAST, we strongly recommend reading the detailed explanation and elaboration (E&E, see link above) paper and to check the examples on www.probast.org

**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 the risk of death in LBCL (large B-cell lymphoma) patients after disease progression to CAR (chimeric antigen receptor) T-cell therapy.* |
| ***Participants** including selection criteria and setting:* | *Adult patients with LBCL who experienced disease progression after CAR T-cell therapy administered in the third or later line 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):* | Each predictor provides up to 1 point to the score, all assessed at time of disease progression to CAR-T therapy. Highlighted are the cut-offs to assign the points.<br>*ECOG (0 vs. **≥1**)<br>*LDH (<2 vs. **≥2 xULN** [Upper Limit of Normal])<br>*Hemoglobin (**< 10** vs. ≥10 g/dL)<br>*Number of extranodal sites (<2 vs. **≥2**)<br>*Time from CAR-T therapy to disease progression (**<4** vs. ≥4 months) |
| *Outcome to be predicted:* | *Overall Survival after CAR-T progression.* |

**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 | YES | 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** | *Submitted, under consideration* |
| **Models of interest** | *PC-PI score* |
| **Outcome of interest** | *Overall Survival since CAR-T progression.* |

**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:***<br>-DEVELOPMENT COHORT: Patients with relapsed/refractory (R/R) LBCL who experienced disease progression after receiving CD19-targeted CAR T-cell therapy in the third or later line setting in Spain.<br>-VALIDATION COHORT: Patients with relapsed/refractory (R/R) LBCL who experienced disease progression after receiving CD19-targeted CAR T-cell therapy in the third or later line setting in three different European centers, namely CHU Lyon (Lyon, France), King's College Hospital (London, United Kingdom) and LMU University Hospital (Munich, Germany). | | | |
| | | Dev | Val |
| 1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data? | | YES | YES |
| 1.2 Were all inclusions and exclusions of participants appropriate? | | YES | YES |
| **Risk of bias introduced by selection of participants** | **RISK:** *(low/ high/ unclear)* | LOW | LOW |
| *Rationale of bias rating:*<br>A consecutive series of patients with disease progression after CAR T-cell therapy was included, with no concerns regarding eligibility. | | | |
| **B. Applicability** | | | |
| *Describe included participants, setting and dates:*<br>-DEVELOPMENT COHORT: 216 patients with relapsed/refractory (R/R) LBCL who experienced disease progression after receiving CAR T-cell therapy in the third or later line setting at 12 Spanish centers, from September 2018 until May 2022.<br>-VALIDATION COHORT: 204 patients treated at three different European centers, namely CHU Lyon (Lyon, France), King's College Hospital (London, United Kingdom) and LMU University Hospital (Munich, Germany) who also presented disease progression after CAR T-cell therapy administered as third or later line from June 2018 until May 2023. | | | |
| **Concern that the included participants and setting do not match the review question** | **CONCERN:** *(low/ high/ unclear)* | LOW | LOW |
| *Rationale of applicability rating:*<br>The included patients appear representative of the population specified in the review question. Since all centers are European with public healthcare systems, it would be interesting to see its applicability in other countries with different healthcare systems. | | | |

| DOMAIN 2: Predictors | | | |
|---|---|---|---|
| **A. Risk of Bias** | | | |
| *List and describe predictors included in the final model, e.g. definition and timing of assessment:*<br>Each predictor provides up to 1 point to the score, all assessed at time of disease progression to CAR-T therapy. Highlighted are the cut-offs to assign the points.<br>*ECOG performance status (0 vs. **≥1**)<br>*LDH (<2 vs. **≥2 xULN** [Upper Limit of Normal])<br>*Hemoglobin (**< 10** vs. ≥10 g/dL)<br>*Number of extranodal sites (<2 vs. **≥2**)<br>*Time from CAR-T therapy to disease progression (**<4** vs. ≥4 months) | | | |
| | | Dev | Val |
| 2.1 Were predictors defined and assessed in a similar way for all participants? | | YES | YES |
| 2.2 Were predictor assessments made without knowledge of outcome data? | | YES | YES |
| 2.3 Are all predictors available at the time the model is intended to be used? | | YES | YES |
| **Risk of bias introduced by predictors or their assessment** | **RISK:** *(low/ high/ unclear)* | LOW | LOW |
| *Rationale of bias rating:*<br>The predictors were included based on previous studies in the field and the clinical judgment of different collaborators in the study. | | | |
| **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:*<br>All predictors were measured at time of CAR-T progression. | | | |

| 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:*<br>*The outcome was Overall Survival since CAR-T progression, meaning the time from disease progression to CAR T-cells until death from any cause. OS was estimated from time of PD to CAR T-cells using the Kaplan-Meier method.* | | | |
| | | Dev | Val |
| 3.1 Was the outcome determined appropriately? | | YES | YES |
| 3.2 Was a pre-specified or standard outcome definition used? | | YES | YES |
| 3.3 Were predictors excluded from the outcome definition? | | YES | YES |
| 3.4 Was the outcome defined and determined in a similar way for all participants? | | YES | YES |
| 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? | | YES | YES |
| **Risk of bias introduced by the outcome or its determination** | **RISK:** *(low/ high/ unclear)* | LOW | LOW |
| *Rationale of bias rating:*<br>The predictors used in the score are widely available variables in this context, so their high predictive power on the outcome was expected. That's why "Probably Not" (PN) is the answer to point 3.5. | | | |
| **B. Applicability** | | | |
| *At what time point was the outcome determined:*<br>The overall survival (time to event) outcome was determined since CAR-T progression until death from any cause. In the development cohort, 72.69% of patients had the event of interest (death) and in the validation cohort, 73.04%.<br><br>*If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:*<br>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:*<br>It is the most important and commonly used outcome in this field, which is why it was chosen. | | | |

**DOMAIN 4: Analysis**

**Risk of Bias**

*Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:*

*DEVELOPMENT: 216 participants, 15 candidate predictors, 157 events, 10.5 events per candidate predictor*
*VALIDATION: 204 participants, 149 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):*

We used LASSO-Cox with lambda1se to select the model variables. For model creation, we conducted a multivariable stratified Cox regression with post-CART treatment as the stratification factor. Each variable coefficient was rounded to 1 to facilitate clinical application. The risk group definition was formulated to ensure similar group sizes in the development cohort.

*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):*

*The external validation used the same variables and cutoffs as the development model. Other European countries provided the data for this external validation (Germany, UK, France), different from the development cohort setting (Spain).*

*Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:*

*c-statistic (discrimination) and calibration plot (calibration).*

*Describe any participants who were excluded from the analysis:*

*The amount of missing data was very low (<4%) but, in order to avoid losing patients, mice imputation was employed.*

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

*DEVELOPMENT: No missing data for "Number of extranodal sites at time of CAR-T progression" and "Time in months from CAR-T to progression". MICE was used to impute missing values (<4%) of ECOG, LDH and Hemoglobin at time of CAR-T progression.*
*VALIDATION: No imputation was done.*

| | | Dev | Val |
|---|---|---|---|
| 4.1 Were there a reasonable number of participants with the outcome? | | YES | YES |
| 4.2 Were continuous and categorical predictors handled appropriately? | | PN | PN |
| 4.3 Were all enrolled participants included in the analysis? | | YES | YES |
| 4.4 Were participants with missing data handled appropriately? | | YES | YES |
| 4.5 Was selection of predictors based on univariable analysis avoided? | | YES | |
| 4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately? | | YES | YES |
| 4.7 Were relevant model performance measures evaluated appropriately? | | YES | YES |
| 4.8 Were model overfitting and optimism in model performance accounted for? | | YES | |
| 4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? | | YES | |
| **Risk of bias introduced by the analysis** | **RISK:** *(low/ high/ unclear)* | LOW | LOW |

*Rationale of bias rating:*

The model's primary objective was to enhance usability for individuals with limited statistical proficiency. Consequently, we rounded the coefficients of the multivariate model to streamline its application. However,

this simplification comes at the cost of precision, mandating the categorization of all variables for ease of use, avoiding the need of an application.

**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.<br>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:**<br>*(low/ high/ unclear)* | LOW (DEVELOPMENT)<br>LOW (VALIDATION) |
| *Summary of sources of potential bias:*<br>The risk of bias is low due to the involvement of different countries and the homogeneous patient population in this specific context. | | |
| **Overall judgement of applicability** | **CONCERN:**<br>*(low/ high/ unclear)* | LOW (DEVELOPMENT)<br>LOW (VALIDATION) |
| *Summary of applicability concerns:*<br>No concerns, but it would be interesting to see the application of the model in other healthcare systems and continents. | | |