PROBAST (Prediction model study Risk Of Bias Assessment Tool)

Criteria	Specify your systematic review question
Intended use of model:	The prediction models are intended to support clinical and operational decision-making in intensive care units (ICUs) during high-burden respiratory pandemics. Specifically: • The mortality model estimates the probability of in-hospital death among adults with confirmed COVID-19 who required ICU admission, to aid risk stratification and identification of patients at high risk of deterioration. • The LOS-ICU model predicts the expected length of stay in the ICU, supporting hospital capacity planning and bed turnover estimation.
Participants *including selection criteria and setting:*	Participants were drawn from the Brazilian national surveillance database for Severe Acute Respiratory Syndrome (SRAG/INFLUD) for the year 2021, which includes all hospitalized respiratory infection notifications nationwide. Inclusion criteria: Mortatlity prediction: • Adults aged ≥ 18 years. • Laboratory-confirmed COVID-19 (which were classified as CLASSI_FIN = 5 in the INFLUD 2021 dataset). • ICU admission recorded (which was classified as UTI = 1 in the INFLUD 2021 dataset). LOS-ICU prediction: • Adult patients (≥18 years) with confirmed COVID-19 (CLASSI_FIN=5) admitted to the ICU (UTI=1) and hospitalized (HOSPITAL=1) in the 2021 national SRAG (INFLUD) database, with both ICU admission and ICU discharge dates available (DT_ENTUTI and DT_SAIDUTI). ICU length of stay (LOS_ICU) was calculated as DT_SAIDUTI − DT_ENTUTI, restricted to 0–120 days and truncated at 40 days.
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):*	Demographic and Epidemiological • Age, sex, municipality of residence, epidemiological week, regional indicators. Clinical Symptoms and Support • Dyspnea, oxygen saturation, fever, cough, fatigue, vomiting, diarrhea, disorientation, abdominal pain, chest pain. • Need for ventilatory support (SUPORT_VEN). • Presence of any risk factors (FATOR_RISC). Comorbidities • Cardiopathy, pneumopathy, renal disease, immunodeficiency. • Additional comorbidities manually selected based on prior literature. Hospitalization and ICU Evolution • ICU admission and discharge dates (DT_ENTUTI, DT_SAIDUTI). • Calculated **LOS-ICU**. • Administrative variables (e.g., DT_EVOLUCA, DT_ENCERRA). Measurement Notes • All predictors originate from standardized national surveillance forms. • Collected during hospital admission or daily clinical updates, • No imaging or laboratory biomarkers are available in the INFLUD dataset,

	• No specialized equipment is required to obtain any predictor included in the model.
Outcome to be predicted:	• Mortality model • • Primary outcome: In-hospital mortality (EVOLUCAO = 2). • Defined by the official discharge status recorded in the national surveillance system. • Binary classification: death vs discharge (excluding outcome codes 3 and 9). • • LOS-ICU model • • Primary outcome: Length of stay in the ICU, calculated as: LOS_ICU = DT_SAIDUTI – DT_ENTUTI (in days). • Patients who died in the ICU are included with death treated as a valid endpoint for LOS.

Step 2: Classify the type of prediction model evaluation

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	Scope review: https://www.medrxiv.org/content/10.1101/2023.02.08.23285446v1 Thesis / Model development: https://repositorio.unifesp.br/items/7188041c-7d05-46fe-9637-d5d7880cd339
Models of interest	• Mortality prediction model (ICU COVID-19) • ICU Length-of-Stay (LOS-ICU) prediction model
Outcome of interest	• In-hospital mortality (death vs discharge) • Length of ICU stay (days)

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants

A. Risk of Bias

Describe the sources of data and criteria for participant selection:

The study used the Brazilian national surveillance database for Severe Acute Respiratory Syndrome (SRAG/INFLUD) for the year 2021, which includes all hospitalized respiratory infections across public and private hospitals in the 27 Brazilian states.

Participants were selected through the following predefined criteria:

1. Adults aged ≥ 18 years
2. Laboratory-confirmed COVID-19 (CLASSI_FIN = 5)
3. ICU admission recorded (UTI = 1)
4. Outcome status (EVOLUCAO) used as reported in the national surveillance system
5. No selection based on treatment received, hospital type, disease severity, socioeconomic status, or any post-baseline information

The resulting mortality cohort consisted of 392,572 adult ICU COVID-19 admissions, of which 96.9% (380,575) had complete outcome documentation.

No exclusions were made based on predictor availability aside from standard missing-data handling in the analytical phase.

		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?		Y	Y
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*	Low	Low

Rationale of bias rating:

Participant selection followed predefined, objective eligibility criteria based solely on:

- adult age (≥18 years),
- confirmed COVID-19 diagnosis (CLASSI_FIN = 5),
- recorded ICU admission (UTI = 1).

These criteria are independent of predictors and outcomes, and were applied uniformly to a nationwide consecutive surveillance cohort (SRAG/INFLUD 2021).

B. Applicability

Describe included participants, setting and dates:

Describe included participants, setting and dates:

Adult patients (≥18 years) with laboratory-confirmed COVID-19 who were admitted to an ICU in Brazil during the 2021 epidemiological year. Participants were identified from the national SRAG/INFLUD surveillance system, which captures mandatory hospitalization records from public and private hospitals across all 27 Brazilian states. The cohort reflects real-world ICU populations under routine clinical conditions.

Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*	Low	Low
Rationale of applicability rating: The study population precisely matches the intended target group for the prediction model—adult patients with laboratory-confirmed COVID-19 who required ICU admission. Participants were drawn from a nationwide, mandatory surveillance system (SRAG/INFLUD) that includes all public and private hospitals across Brazil's 27 states, ensuring broad representativeness and real-world applicability. The setting, time period (2021), and inclusion criteria are fully aligned with the review question and the intended operational use of the model. Therefore, there is no meaningful divergence between the included participants and the target population.			

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment:

The final set of predictors (34 variables) was derived using a hybrid selection strategy combining (1) automatic importance ranking from XGBoost and (2) manual inclusion of clinically essential comorbidities known to influence ICU outcomes. All predictors were measured at hospital admission or corresponded to baseline demographic or epidemiological characteristics, preventing temporal or information leakage.
The predictors included in the final model are grouped as follows:

1. Demographic and Epidemiological Predictors (n = 10)
Measured at first notification or hospital intake; no timing variability exists.

- NU_IDADE_N — patient age in years (continuous)
- SEXO — biological sex
- CO_MUN_RES, CO_MU_INTE, CO_MUN_NOT — municipality of residence, hospital, and notification
- SEM_PRI, SEM_NOT — epidemiological weeks of symptom onset and notification
- CO_REGIONA, CO_RG_INTE, CO_RG_RESI — regional health identifiers

These variables reflect structural epidemiologic context and patient demographic characteristics, showing excellent temporal consistency (assessed once at baseline).

2. Clinical Symptoms, Ventilatory Support and Risk Factors (n = 19)

All measured at hospital admission and prior to outcome occurrence.

- SUPORT_VEN — ventilatory support status at admission
- FATOR_RISC — presence of aggregated risk factors
- DISPNEIA, SATURACAO, FEBRE, TOSSE, GARGANTA, FADIGA, VOMITO, DIARREIA, DESORIENT, DOR_ABD, DOR_TORAC, OUTRO_SIN
- OUT_MORBI — “other comorbidities” (free-text coded into structured categories)
- CARDIOPATI, PNEUMOPATI, RENAL, IMUNODEPRE

These variables describe the acute clinical presentation and underlying risk conditions. All were recorded upon admission, avoiding forward-looking data or post-outcome information.

3. Manually Added Comorbidities (n = 2)
Not selected automatically but added based on strong clinical and epidemiological evidence of prognostic relevance.

- HEMATOLOGI — chronic hematologic disease
- HEPATICA — chronic liver disease

These were added to ensure clinically meaningful representation of comorbidity burden and to prevent omission of biologically important predictors in high-severity respiratory illness.

4. ICU Hospitalization Timing Variables (LOS-ICU model only; n = 3)
Used exclusively for the LOS-ICU regression model, not for mortality prediction.

- DT_ENTUTI — ICU admission date
- DT_SAIDUTI — ICU discharge date
- LOS_ICU — derived length of stay in ICU (DT_SAIDUTI – DT_ENTUTI, in days)

These variables were excluded from the mortality classification model to prevent outcome leakage. For LOS-ICU prediction, both timestamps are inherently baseline to the derived outcome and were handled consistently.

Removed Predictors to Avoid Bias (n = 2)
The following administrative timestamps were intentionally excluded from all models to prevent temporal leakage:

- DT_ENCERRA — system closure date

- DT_EVOLUCA — administrative outcome date

Their exclusion ensured that no predictor was influenced by, or temporally dependent on, the final outcome.

Risk of bias judgement for DOMAIN 2 — Predictors
RISK: Low

Rationale:
All predictors were defined prior to outcome occurrence, consistently assessed at admission, and free from post-treatment or post-outcome information. Manual additions (hematologic and hepatic comorbidities) strengthened clinical completeness without introducing bias.

Predictors prone to leakage (administrative timestamps) were manually purposefully excluded. Thus, the risk of bias introduced by predictor definition, measurement, or timing is low.

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

All predictors were captured through standardized national surveillance forms (SIVEP-Gripe), ensuring consistent definitions and identical measurement procedures for all participants across Brazil's public hospital network.

All predictors were recorded before outcome occurrence, at the time of admission or notification, and none required retrospective reconstruction. Administrative timestamps that could introduce temporal leakage (e.g., DT_EVOLUCA, DT_ENCERRA) were excluded from the modeling pipeline.

All included predictors would routinely be available to clinicians at the intended point of model use (hospital or ICU admission).

Therefore, the risk of bias arising from predictor definition, measurement, or timing is low.

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:

Predictors were defined using standardized national surveillance fields and measured consistently at or before ICU admission, matching the review focus on routinely available demographic and clinical variables. No specialized tests outside the intended clinical setting were required. Thus, applicability concerns are low.

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 mortality outcome was defined using the official discharge status (EVOLUCAO), a standardized field in the national surveillance system. Outcomes were determined after hospitalization or ICU stay, always occurring after predictor assessment at admission. The timing between predictor measurement and outcome determination is appropriate and consistently recorded, minimizing bias.

		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?		Y	Y
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:

Mortality is determined through standardized national discharge coding (EVOLUCAO), with high procedural consistency.

The EVOLUCAO codes follow a national surveillance standard agreed upon by the Ministry of Health. No predictor variable overlaps with or contributes to the construction of the EVOLUCAO field. Predictors are recorded at admission, and outcome is determined only after hospital or ICU completion, ensuring correct temporal direction.

Outcome assignment follows a uniform national protocol across all hospitals and reporting units. Outcome coding is performed at discharge and is not influenced by baseline predictors used in the model. Outcome definitions follow national standards, are uniformly applied, temporally appropriate, and fully independent of predictor variables, minimizing risk of systematic misclassification.

B. Applicability

At what time point was the outcome determined:

The outcome (in-hospital mortality, EVOLUCAO = 2) was determined only at the end of the hospitalization or ICU stay, following completion of treatment and discharge procedures.
Thus, outcome assessment occurs after all predictor measurements, preserving the correct temporal sequence for prognostic modelling.

If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:
Not applicable. The study uses a single, non-composite outcome (death vs. discharge). No subcomponents contribute to the definition.

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:

The outcome—in-hospital mortality (EVOLUCAO = 2)—is defined using a standardized national field, determined uniformly at discharge, and always assessed after predictor measurement. This aligns fully with the review question and the intended prognostic use of the model.

DOMAIN 4: Analysis
Risk of Bias
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor: ICU cohort included 392,572 adults with confirmed COVID-19. A total of 34 candidate predictors were used. Valid outcome data (EVOLUCAO = 1 or 2) were available for 380,575 participants, with 156,376 deaths and 217,723 discharges, yielding ~11,200 events per predictor, well above recommended thresholds for stable model estimation.
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): The model was developed using supervised machine learning, primarily logistic regression and gradient-boosting algorithms (XGBoost, LightGBM, CatBoost, Random Forest) for mortality, and regression variants of the same methods for LOS-ICU. Predictor selection followed a hybrid approach: initial automated ranking by XGBoost feature importance, complemented by manual inclusion of clinically relevant comorbidities. No risk groups were pre-defined; the models output individual probability estimates for mortality and continuous LOS predictions.
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 model underwent internal validation using a stratified random 70/30 train–test split, preserving the mortality ratio. No external validation was performed. Performance was evaluated on the independent test set using discrimination metrics (AUC), calibration checks, and standard classification measures.
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism: Model performance was assessed using discrimination metrics (AUC, F1-score, sensitivity, specificity, precision) and calibration evaluation through calibration curves and predicted-vs-observed probability plots. No net benefit analysis or decision-curve analysis was performed. Performance measures were not adjusted for optimism.
Describe any participants who were excluded from the analysis: Participants younger than 18 years, those without confirmed COVID-19 (CLASSI_FIN ≠ 5), and individuals without documented ICU admission (UTI ≠ 1) were excluded. Additionally, records with missing outcome status (EVOLUCAO) were removed from the mortality model, and cases lacking ICU admission or discharge dates were excluded from the LOS-ICU analysis.
Describe missing data on predictors and outcomes as well as methods used for missing data: Several predictors contained missing values, which were handled using simple zero-imputation consistent with prior DataSUS-based modelling studies. Most of the predictors have >70% completeness, mitigating the risk of bias, although a formal analysis was not performed. For the mortality model, participants without a recorded outcome (EVOLUCAO) were excluded from analysis.

For the LOS-ICU model, cases missing either ICU admission or discharge dates were removed because LOS could not be computed.

		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?		Y	Y
4.4 Were participants with missing data handled appropriately?		Y	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 multivariable analysis?		Y	
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*	Low	Low

Rationale of bias rating:

No univariable screening was used; instead, predictor selection combined data-driven XGBoost importance with supervised clinical curation.

No p-value filtering, univariable regressions or correlation-based exclusions were used.

Instead, a hybrid strategy was applied: initial variable ranking via XGBoost importance, followed by clinically supervised manual inclusion of two comorbidities (so the model works with the comprehensive list of available comorbidities) to ensure relevance and stability. This approach avoids the known biases of univariable selection.

The ICU mortality outcome had a very large number of events (>150,000 deaths), ensuring a strong events-per-predictor ratio.

Predictors were handled appropriately using standard preprocessing (zero-imputation for tree models and standardization for logistic regression). Only participants with a computable outcome were included, and missingness was handled transparently and consistently.

Data complexities were adequately addressed (no censoring for mortality; LOS handled as a continuous endpoint). Model performance was evaluated using discrimination metrics (AUC, F1, sensitivity, specificity) and checked for internal optimism via a 70/30 train–test split.

Final model weights and predictor contributions were consistent with multivariable learning algorithms, particularly XGBoost, LightGBM and CatBoost.

Step 4: Overall assessment

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: All four PROBAST domains (Participants, Predictors, Outcome, and Analysis) were rated as low risk of bias. The study used a very large national dataset (n ≈ 392,572 ICU patients), minimizing instability and overfitting. Internal validation was conducted through a stratified 70/30 train–test split, and no univariable predictor selection was used. Outcome definitions followed a national standardized surveillance system, predictors were measured before the outcome, and missing data were handled transparently. No domain presented methodological weaknesses substantial enough to elevate the risk of bias. Per PROBAST guidance, prediction models without external validation may be downgraded. However, the extremely large sample size, high event counts, and proper internal validation procedures justify maintaining an overall low risk of bias classification.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	Low
Summary of applicability concerns: The study population (adult ICU patients with confirmed COVID-19), the routinely collected predictors, and the standardized national outcome definitions closely match the intended context of use for prognostic ICU models. No discrepancies were identified between the model's design and the target clinical setting, resulting in no meaningful applicability concerns.		