METHODS

Study Population and Outcomes

Based on data consistency, comparability, and sufficient follow-up time to obtain survival outcomes, we extracted data on PLC cases with complete survival information from the Surveillance, Epidemiology, and End Results (SEER) database (2000-2017) using SEER*Stat software version 8.4.4 (National Cancer Institute, Bethesda, MD, USA)[1]. The follow-up cutoff date was December 31, 2022, allowing all patients diagnosed up to December 31, 2017, to have complete 60-month follow-up. A 60-month endpoint was chosen because it is consistent with international clinical consensus on cancer prognosis evaluation. Patients with PLC were identified using the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), based on primary site and morphology codes for the liver (C22.0–C22.9). As the SEER database is publicly available and de-identified, institutional review board approval and informed consent were not required for this study[2]. Patients diagnosed with PLC between 2000 and 2017 were identified using the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), based on the primary site and morphology codes corresponding to the liver (C22.0–C22.9). This study used de-identified, publicly available data from the SEER database. As the data contain no personal identifiers, institutional review board approval and informed consent were not required.

Inclusion and Exclusion Criteria

Inclusion criteria were as follows: (1) diagnosed with primary liver cancer; (2) known race; between 2000 and 2017; (3) known age; (4) known survival time. The exclusion criteria were as follows: (1) cases with missing covariates; (2) non-primary liver cancer; (3) cases reported based on autopsy or death certificate only; and (4) cases without a pathological diagnosis confirmed by positive histology. The flowchart is shown in Supplementary Figure S1, and a total of 57,526 PLC patients were ultimately included in the study.

Variables Included in the Analysis

The variable**s** in the study included the following variables: age (numeric); survival months (numeric); race (white and hispanic, black, Asian or Pacific Islander, American Indian/Alaska Native); year of diagnosis (2000–2005, 2006–2011, 2012–2017); median household income (inflation-adjusted to 2022): ≤ 90,000/year (household), > 90,000/year (household); residential status (counties in metropolitan areas, counties not in metropolitan areas); marital status (married categories, single categories, unknown); histologic type [hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), other]; stage (localized, regional, distant, unknown/unstaged); surgery (no/unknown, yes); radiotherapy (no/unknown, yes); chemotherapy (no/unknown, yes); time from diagnosis to treatment (≤ 2 months, > 2 months, no/unknown); outcome [alive, dead (attributable to PLC), dead (attributable to other causes)]. The specific coding for each group is shown in Supplementary Table S1, where the group with the smaller code is considered the reference group.

Statistical Analyses

Considering that “dead (attributable to other causes)” and "dead (attributable to PLC)" are competing risks, we used univariate and multivariable Fine-Gray competing risks models[3] to

analyze the effect of each variable on the prognosis of PLC patients. Then, based on the SEER cause-specific death classification, the outcomes were combined into "alive or dead of other causes" and "dead (attributable to PLC)", and univariate and multivariable Cox proportional hazards regression models were used to analyze the impact of each variable on the prognosis of PLC patients.

We then divided the entire cohort into a training set and a test set in a 7:3 ratio. The baseline characteristics of both the training and test sets were compared using the χ^2 test and Fisher's exact test, while the non-parametric Kruskal-Wallis test was applied for numeric variables. The machine learning methods we used included Light Gradient Boosting Machine (LightGBM)[4], eXtreme Gradient Boosting machine (XGBoost)[5], random forest[6], logistic regression, support vector machine (SVM)[7], and artificial neural networks (ANN)[8]. The DeLong test was used to compare the differences in AUC between different models.

Based on the aforementioned prognosis factors, 6 machine learning methods were applied to the training set. Combined with 5-fold cross-validation, hyper-parameter tuning was conducted, and 6 prediction models (6 machine learning classifiers) were constructed for the 60-month PLC-specific survival (considering only whether the patient died within 60 months). The models' performances were then evaluated on the test set using area under the receiver operating characteristic curve (AUC) and calibration curves. AUC measures the models' discrimination ability, indicating how well the model can distinguish between patients who died from PLC within 60 months and those who did not, while calibration curves assess the agreement between predicted probabilities and observed outcomes. The best performing model was selected to construct the final 60-month PLC-specific mortality prediction model, and prognostic stratification was performed on the entire cohort by using the 60-month PLC-specific mortality prediction model. The flowchart is shown in Supplementary Figure S1B. In addition, the SHapley Additive exPlanations (SHAP)[9] method was applied to assess the impact of each variable on the performance of the final model. Finally, the best performing model was deployed as a prognostic stratification calculator tool.

Data analysis was performed using R software version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria). The packages used in this study include: "survival", "pROC", "caret", "lightgbm", "e1071", "RSNNS", and "randomForest". Statistical significance was set at a *P* value < 0.05.

This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis-Artificial Intelligence (TRIPOD-AI) reporting guidelines[10] (Supplementary TRIPOD checklist).

Supplementary Table S1. Characteristics of groups and code for prognosis factors

Characteristics	Characteristics type	Description	Code
Diagnosis year	Categorical	2000-2005	1
		2006-2011	2
		2012-2017	3
Age	Numeric		
Sex	Categorical	Female	1
		Male	2
Race	Categorical	White and Hispanic	1
		Black	2
		Asian or Pacific Islander	3
		American Indian/Alaska Native	4
Income	Categorical	≤ 90,000/year (Household)	1
		> 90,000/year (Household)	2
Residential Status	Categorical	Counties in metropolitan areas	1
		Counties not in metropolitan areas	2
Marital status	Categorical	Married categories	1
		Single categories	
		Unknown	
Histologic type	Categorical	HCC	1
		ICC	2
		Other	3
Survival months	Numeric		
Stage	Categorical	Localized	1
		Regional	2
		Distant	3
		Unknown/unstaged	4
Surgery	Categorical	No/Unknown	1
		Yes	2
Radiotherapy	Categorical	No/Unknown	1
		Yes	2
Chemotherapy	Categorical	No/Unknown	1
		Yes	2
Time from diagnosis to treatment	Categorical	≤ 2 months	1
		> 2 months	2
		No/Unknown	3
Status	Categorical	Alive	0
		Dead (attributable to PLC)	1
		Dead (attributable to other causes)	2

Note. PLC, primary liver cancer; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma.

Supplementary Table S2. Comparison of baseline characteristics by 60-month PLC-specific mortality in the training and test sets

Characteristics	Total (*n* = 57,526)	Test (*n* = 17,257)	Training (*n* = 40,269)	*P*
Age, Median [Q1, Q3]	64.00 [56.00, 73.00]	64.00 [56.00, 73.00]	64.00 [56.00, 73.00]	0.562
Sex, *n* (%)				
Female	15,153 (26.34)	4,493 (26.04)	10,660 (26.47)	0.281
Male	42,373 (73.66)	12,764 (73.96)	29,609 (73.53)	
Race, *n* (%)				
White and Hispanic	40,681 (70.72)	12,150 (70.41)	28,531 (70.85)	0.634
Black	6,771 (11.77)	2,037 (11.80)	4,734 (11.76)	
Asian or Pacific Islander	9,524 (16.56)	2,897 (16.79)	6,627 (16.46)	
American Indian/Alaska Native	550 (0.96)	173 (1.00)	377 (0.94)	
Income, *n* (%)				
≤ 90,000/year (Household)	41,865 (72.78)	12,562 (72.79)	29,303 (72.77)	0.958
> 90,000/year (Household)	15,661 (27.22)	4,695 (27.21)	10,966 (27.23)	
Residential Status, *n* (%)				
Counties in metropolitan areas	51,768 (89.99)	15,514 (89.90)	36,254 (90.03)	0.645
Counties not in metropolitan areas	5,758 (10.01)	1,743 (10.10)	4,015 (9.97)	
Marital status, *n* (%)				
Married categories	30,851 (53.63)	9,221 (53.43)	21,630 (53.71)	0.525
Single categories	24,015 (41.75)	7,256 (42.05)	16,759 (41.62)	
Unknown	2,660 (4.62)	780 (4.52)	1,880 (4.67)	
Histologic Type, *n* (%)				
HCC	48,590 (84.47)	14,627 (84.76)	33,963 (84.34)	0.363
ICC	1,723 (3.00)	518 (3.00)	1,205 (2.99)	
Other	7,213 (12.54)	2,112 (12.24)	5,101 (12.67)	
Stage, *n* (%)				
Localized	27,050 (47.02)	8,154 (47.25)	18,896 (46.92)	0.610
Regional	14,487 (25.18)	4,369 (25.32)	10,118 (25.13)	
Distant	10,621 (18.46)	3,132 (18.15)	7,489 (18.60)	
Unknown/unstaged	5,368 (9.33)	1,602 (9.28)	3,766 (9.35)	
Surgery, *n* (%)				
No/Unknown	38,694 (67.26)	11,597 (67.20)	27,097 (67.29)	0.844
Yes	18,832 (32.74)	5660 (32.80)	13,172 (32.71)	
Radiotherapy, *n* (%)				
No/Unknown	53,064 (92.24)	15,913 (92.21)	37,151 (92.26)	0.866
Yes	4,462 (7.76)	1,344 (7.79)	3,118 (7.74)	
Chemotherapy, n(%)				
No/Unknown	37,794 (65.70)	11,443 (66.31)	26,351 (65.44)	0.045
Yes	19,732 (34.30)	5,814 (33.69)	13,918 (34.56)	
Time from diagnosis to reatment, *n*				

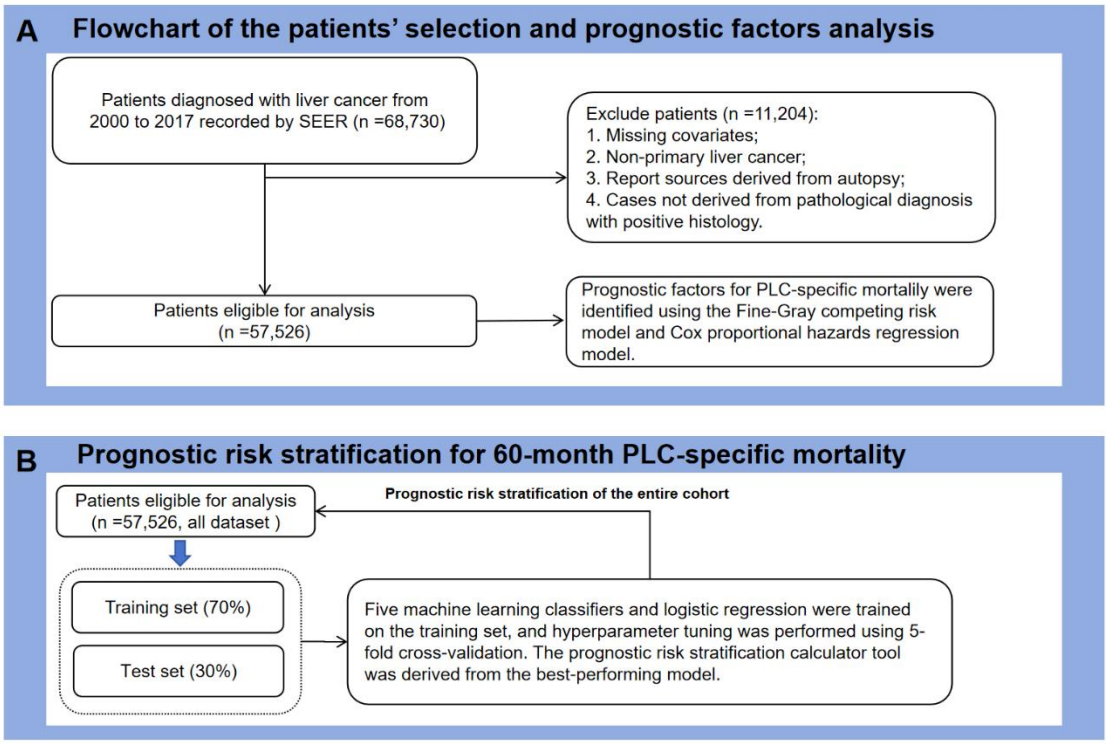
(%)				
≤ 2 months	21,255 (36.95)	6,369 (36.91)	14,886 (36.97)	0.853
> 2 months	10,601 (18.43)	3,204 (18.57)	7,397 (18.37)	
No/Unknown	25,670 (44.62)	7,684 (44.53)	17,986 (44.66)	
Status (60 months)				
Alive or dead of other causes	19,595 (34.06)	5,926 (34.34)	13,669 (33.94)	0.364
Dead (attributable to PLC)	37,931 (65.94)	11,331 (65.66)	26,600 (66.06)	

Note. PLC, primary liver cancer; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma.

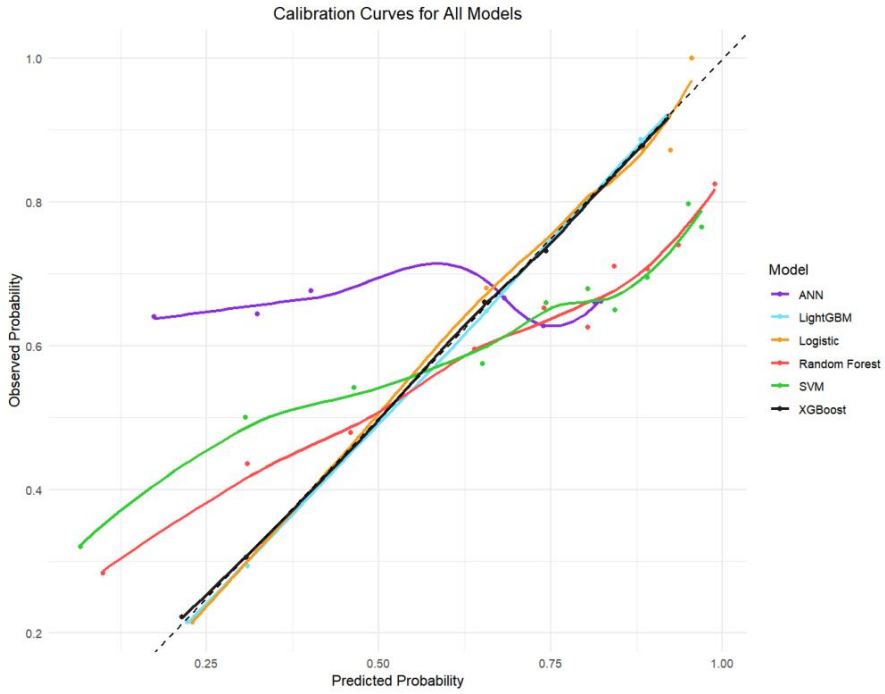
Supplementary Table S3. Hyperparameter space explored for different machine learning classifiers

Machine learning classifiers	Hyperparameters	Range	Step	Final choice
LightGBM	num_leaves	{5, 50}	5	20
	nrounds	{100, 1000}	100	500
	learning_rate	{0.01 ,0.1}	0.01	0.01
	subsample	{0.5, 1}	0.1	0.9
	max_depth	{3, 30}	3	6
	colsample_bytree	{0.5 1}	0.1	0.7
Random Forest	mtry	{2, 8}	2	2
	ntree	{100, 1000}	100	500
	nodesize	{1, 19}	3	1
ANN	number of layers	{1, 5}	1	3
	layer size	{5, 20}	5	10
	maxit	{100, 300}	100	100
	learnFuncParams	{0.01, 0.1}	0.01	0.01
XGBoost	nrounds	{100, 1000}	100	100
	max_depth	{3, 15}	3	9
	eta	{0.01, 0.1}	0.01	0.05
	subsample	{0.5, 1}	0.1	1
	colsample_bytree	{0.5, 1}	0.1	0.6
	min_child_weight	{1, 9}	2	5
	gamma	{0, 5}	1	5
SVM	cost	{0.01, 100}	*10	10
	sigma	{0.05, 0.1, 0.5, 1, 2}	\	2
	kernel	{sigmoid, kernel}	\	kernel

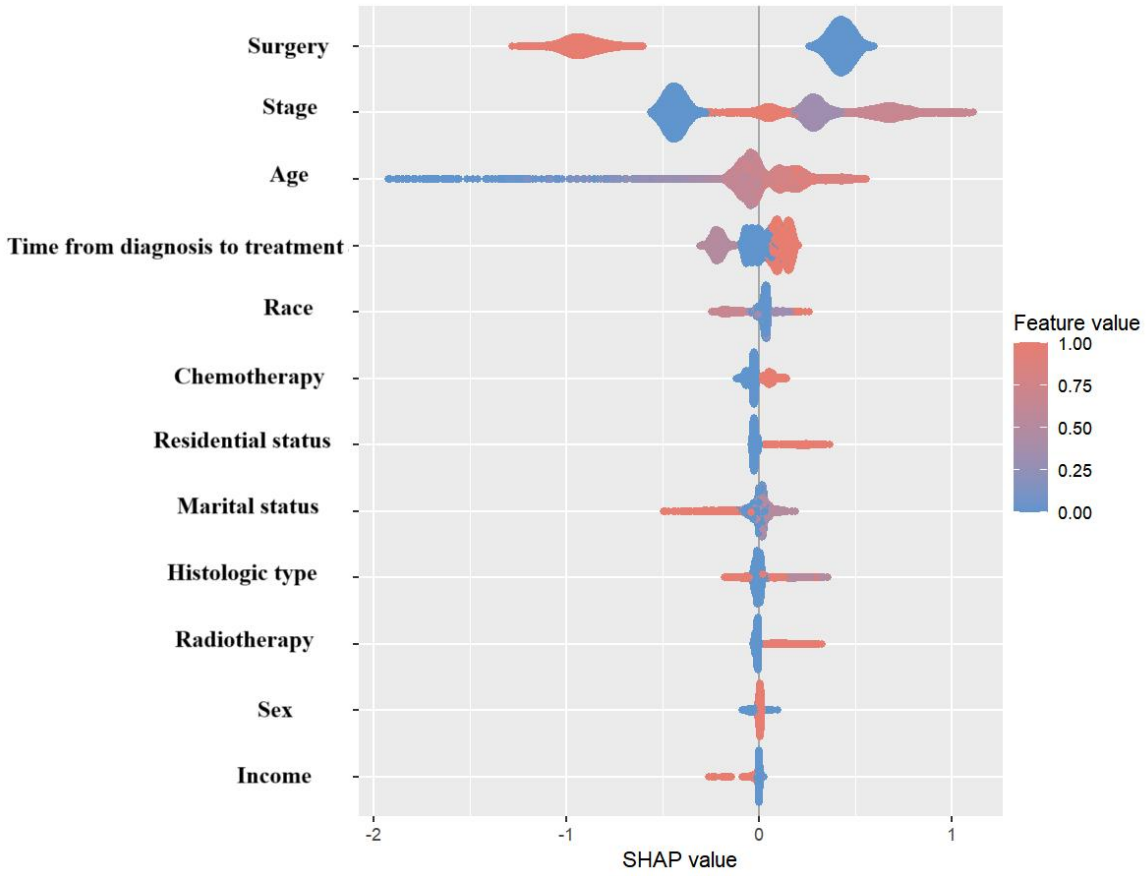
Note. ANN, artificial neural network; SVM, support vector machine; LightGBM, light gradient boosting machine; XGBoost, extreme gradient boosting machine.



Supplementary Figure S1. Flowchart of this study. (A) Flowchart; (B) prognostic risk stratification.



Supplementary Figure S2. Calibration curves of machine learning classifiers for predicting 60-month PLC-specific mortality on test set. ANN, artificial neural network; SVM, support vector machine; LightGBM, light gradient boosting machine; XGBoost, extreme gradient boosting machine.



Supplementary Figure S3. SHAP visualisation of selected predictors in LightGBM model.

Each participant was represented as a data point and was coded with gradient colors to indicate the value of the predictor. The horizontal spread of the points reflects the predictive power of each variable, with a wider range indicating stronger influence on the model output. The position on the x-axis shows whether a higher value of the predictor increases or decreases the likelihood of PLC-specific death within 60 months. Specifically, points on the right indicate that the predictor contributes to a higher risk of death, while points on the left indicate a protective effect. The color gradient further helps interpret how low or high values of the predictor are associated with changes in risk. In this way, the SHAP plot not only quantifies the importance of each predictor but also illustrates the direction and magnitude of its effect on prognosis.

Receiving radiotherapy, not undergoing surgery, having a lower income, living in counties not in metropolitan areas, and being diagnosed at a more advanced stage were all associated with an increased likelihood of PLC-specific death within 60 months. In contrast, a time from diagnosis to treatment of more than 2 months and unknown marital status were associated with a decreased likelihood of PLC-specific death within 60 months.

REFERENCE

1. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database. 2024.
2. Duggan MA, Anderson WF, Altekruse S, et al. The Surveillance, Epidemiology, and End Results (SEER) program and pathology: toward strengthening the critical relationship. Am J Surg Pathol, 2016; 40, e94-e102.
3. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 1999; 94, 496-509.
4. Ke G, Meng Q, Finley T, et al. editors. LightGBM: A highly efficient gradient boosting decision tree. Neural Information Processing Systems; 2017.
5. Chen T, Guestrin C. XGBoost: A scalable tree boosting system 2016. 785-794.
6. Breiman L. Random Forests. Machine learning, 2001; 45, 5-32.
7. Cortes C, Vapnik V. Support-vector networks. Machine learning, 1995; 20, 273-97.
8. Werbos PJ. Generalization of backpropagation with application to a recurrent gas market model. Neural Networks, 1988; 1, 339-56.
9. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc. 2017; 4768–77.
10. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 2024; 385, e078378.

Supplementary TRIPOD checklist

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D; E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Page 1
ABSTRACT				
Abstract	2	D; E	See TRIPOD+AI for Abstracts checklist	Not applicable
INTRODUCTION				
Background	3a	D; E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Pages 1-2
	3b	D; E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	Pages 1-2
	3c	D; E	Describe any known health inequalities between sociodemographic groups	Pages 1-2
Objectives	4	D; E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Pages 2
METHODS				
Data	5a	D; E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care, or registry data), the rationale for using these data, and the representativeness of the data	Appendix Methods
	5b	D; E	Specify the dates of the collected participant data, including the start and end of participant accrual and, if applicable, the end of follow-up	Appendix Methods
Participants	6a	D; E	Specify key elements of the study setting (e.g., primary care, secondary care, general population), including the number and location of centers	Appendix Methods
	6b	D; E	Describe the eligibility criteria for study participants	Appendix Methods
	6c	D; E	Give details of any treatments received and how they were handled during model development or evaluation, if relevant	Not applicable
Data preparation	7	D; E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Not applicable
Outcome	8a	D; E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Pages 3
	8b	D; E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	Appendix Methods
	8c	D; E	Report any actions to blind assessment of the outcome to be predicted	Not applicable
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	Pages 3

	9b	D; E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	Pages 3
	9c	D; E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	Not applicable
Sample size	10	D; E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	Not applicable
Missing data	11	D; E	Describe how missing data were handled. Provide reasons for omitting any data	Page 2 and eFIG 1
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	Appendix Methods
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardization).	Appendix Methods
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	AAppendix Methods
	12d	D; E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster foradditional considerations[3]	Not applicable
	12e	D; E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	FIG 1
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	Not applicable
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	Appendix Methods
Class imbalance	13	D; E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	Not applicable
Fairness	14	D; E	Describe any approaches that were used to address model fairness and their rationale	Not applicable
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Page 4-5
Training versus evaluation	16	D; E	Identify any differences between the development and evaluation data in a healthcare setting, eligibility criteria, outcome, and predictors	Not applicable
Ethical approval	17	D; E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	Appendix Methods
OPEN SCIENCE				
Funding	18a	D; E	Give the source of funding and the role of the funders for the present study	Not applicable

Conflicts of interest	18b	D; E	Declare any conflicts of interest and financial disclosures for all authors	Page 6
Protocol	18c	D; E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Not prepared
Registration	18d	D; E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	Not registered
Data sharing	18e	D; E	Provide details of the availability of the study data	Not applicable
Code sharing	18f	D; E	Provide details of the availability of the analytical code[4]	Not applicable
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D; E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	No involvement
RESULTS				
Participants	20a	D; E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	eFIG 1
	20b	D; E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	Table 1
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	Table 1
Model development	21	D; E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Table 1
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	Appendix eTable 1
Model performance	23a	D; E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid the presentation.	FIG 1 and Appendix eFIG 2 and eFIG 3
	23b	D; E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	Not applicable
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	Not applicable
DISCUSSION				
Interpretation	25	D; E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Pages 4-5
Limitations	26	D; E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, or missing data) and their effects on any biases, statistical uncertainty, and generalizability	Page 5
Usability of the model in the	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	Pages 4-5

context of current care	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model and what level of expertise is required of users	Pages 4-5
	27c	D; E	Discuss any next steps for future research, with a specific view to the applicability and generalizability of the model	Pages 5

Note. D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D; E=items applicable to both the development and evaluation of a prediction model.

From: Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 2024; 385, e078378.