TRIPOD+AI Expanded Checklist (Explanation & Elaboration Light)

Section/Topic	Item		Checklist item
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** ● *Informative titles aid the identification of prediction model studies by potential readers and also systematic reviewers* ● *Report an informative title that provides key information about the target population and the outcome being predicted by the model*
ABSTRACT			
Abstract	2	D;E	**See TRIPOD+AI for Abstracts checklist** ● *Report an abstract addressing each item in the TRIPOD+AI for Abstracts checklist*
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** ● *Describe the healthcare setting where the model is intended to be used or needed* ● *Where an existing prediction model is available, provide a clear justification for developing a new model* ● *For studies evaluating an existing model, provide the rationale for the evaluation, and provide references to all models being evaluated*
	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)** ● *Describe who is the target population for the developed or evaluated model, e.g., people of a certain age, in a specific country, or with a specific disease* ● *Describe the intended purpose of the model, including the clinical decision or guidance the model is intended used to support (e.g., referral for further testing or hospital admission, triage, starting a treatment, or changing a lifestyle) and the point in the care pathway the model is to be intended used* ● *Describe who the intended users of the model are, and if the model is for healthcare professionals, patients, public or other*
	3c	D;E	**Describe any known health inequalities between sociodemographic groups**

Section/Topic	Item		Checklist item
			• *In the context of the healthcare setting where the model is intended to be used, describe any known health inequalities between sociodemographic groups in the target population (along with citations to support the health inequalities)*
Objectives	4	D;E	**Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)** • *Provide an explicit statement of all objectives of the study, describing whether the study is developing a prediction model, evaluating the performance of a prediction model, or both*
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 representativeness of the data** • *Provide a description of the source of the data used for model development and evaluation of model performance, including whether the data are (for example) from a randomised trial, a cohort, a registry or from electronic routine healthcare records* • *Specify whether the study is using existing data or is prospectively collecting new data for the purpose of the prediction model study* • *Where existing data are being used (i.e., they were originally collected for a different purpose), provide the rationale for using these data, and comment on the suitability (particularly if data are being used from a different setting or country to the intended target population) and representativeness of these data with respect to the intended target population and context* • *A description of the data sources should be provided for all data sets, and separately for development and evaluation* • *If any synthetic data have been used, then provide reasons as to why, and provide all details on how the synthetic data have been created (and code, see item 18f) and used in the study*
	5b	D;E	**Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up** • *Specify the start and end dates of the period for which the participants or the used data were selected* • *For models predicting prognosis, the duration of follow-up is important so report the date of end of follow-up*
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 centres** • *Describe the healthcare setting, and where the participants in the study were recruited from* • *Report the geographical location (at a minimum, the country) and centres (including the number of centres) of the study*
	6b	D;E	**Describe the eligibility criteria for study participants** • *The eligibility criteria for participants should be reported to understand the potential applicability and generalisability of the prediction model*

Section/Topic	Item		Checklist item
			• *This includes reporting any restrictions of continuous variables, e.g., age range used to define the eligibility of the included participants*
	6c	D;E	**Give details of any treatments received, and how they were handled during model development or evaluation, if relevant** • *Any treatments received before or at the start of follow-up should be reported, and whether and how this was handled during the development or evaluation of the prediction model (if relevant)* • *Any treatments received between the moment the prediction model is used and the measurement of the outcome, that could modify the probability of the outcome, should be reported (if relevant)*
Data preparation	7	D;E	**Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups** • *Describe any data cleaning steps, this includes any feature engineering, transformation of raw data, feature reduction and data quality checks. All code used for data cleaning should be made available (see item 18f)* • *For analyses using data from multiple sources (e.g., data from different studies, cohorts, or registries), describe any harmonisation (e.g., of outcome and predictors)* • *Confirm whether all data pre-processing/data cleaning steps were similar across key socio demographic groups, if relevant* • *If the data pre-processing/data cleaning steps are extensive, consider reporting this information in the supplementary material*
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** • *For diagnostic prediction models, the outcome should be clearly defined, including whether a (widely accepted) reference standard (ground truth) was used to determine the presence or absence of the outcome* • *For prognostic models, i.e., models predicting an outcome in the future, authors should report the time-horizon of the outcome prediction. For example, predicting the 28-day risk of mortality following cardiothoracic surgery, or the 10-year risk of fractures in patients with osteoporosis. Also, the frequency of outcome assessment during follow-up should be reported* • *If standard definitions are used, e.g., using ICD[1] codes, this should be stated and referenced* • *Any discrepancies in the outcome assessment across socio-demographic groups should be reported* • *In some instances, it may be necessary to confirm that no predictors were used to define the outcome or are a proxy for the outcome*
	8b	D;E	**If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors** • *For outcomes that require a subjective interpretation (e.g., interpreting the results from an imaging test, describe the number, qualification, and demographic characteristics of the outcome assessors)*

[1] ICD stands for International Classification of Diseases

Section/Topic	Item		Checklist item
			• *If the measurement and interpretation of the outcome require (additional) training or specific instructions, these should be reported.* • *If extensive, consider reporting this information in the supplementary material*
	8c	D;E	**Report any actions to blind assessment of the outcome to be predicted** • *The outcome being predicted should be assessed blind to information about the predictors – particularly relevant for outcomes requiring a subjective interpretation thereby avoiding data (label) leakage* • *If appropriate, authors should describe which information was available to the outcome assessors and report any specific actions to blinding the outcome assessment*
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** • *Provide details on how the initial list of predictors were considered for inclusion in the model building, and whether they were chosen based on a (systematic) review of the literature, clinical input (domain experts), or simply whether using all predictors in the available data* • *If any pre-selection of predictors, before model building, was carried out, then provide details how this was done. For example, were predictors omitted for model building due to high amounts of missing data, or predictors not considered plausibly (clinically) related to the outcome being predicted* • *The list of initial predictors may be extensive, in these instances reporting these in the supplementary material is advisable*
	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)** • *All predictors included in the modelling should be clearly defined, along with units of measurement, and all categories for categorical predictors, so that readers and others can replicate, implement, or evaluate the performance of the model* • *Details on how and when the predictor values were measured. Note that predictors should be measured before or at the time the model is intended to be used* • *For predictors requiring subjective interpretation, it may be important to interpret this blind to the values of other predictors considered in the modelling (e.g., avoiding data leakage). Authors should report any actions to blind the assessment of the predictor measurement to other predictors* • *Specifically for diagnostic models, the measurement of the predictors should be done without knowledge of the outcome of the individual as this could artificially inflate the association between the predictors and the outcome. Authors should report any actions to blind the assessment of the predictor measurements to the outcome value* • *In some instances, the number of predictors can be very large and thus reporting them all in the main manuscript is unhelpful, in these instances, it is still important to clearly define all the predictors, and reporting this in the supplementary material should be considered*

Section/Topic	Item		Checklist item
	9c	D;E	**If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors** ● *For predictors that require a subjective interpretation (e.g., interpreting the results from an imaging test), the qualifications and demographic characteristics of the predictor assessors should be reported* ● *If the measurement and interpretation require (additional) training or specific instructions, then these should be reported. This could be reported in the supplementary material*
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** ● *Describe how the sample size was determined – this should be done separately for determining the sample size needed for model development and the sample size needed to evaluate the performance of the model irrespective of whether data are being prospectively collected or using existing data* ● *Provide details and all estimates used in any sample size calculation* ● *If no formal sample size calculation was done, e.g., all available data were used, provide a justification whether the size of the data was sufficient to answer the research question*
Missing data	11	D;E	**Describe how missing data were handled. Provide reasons for omitting any data** ● *Missing data is an omnipresent problem. Authors should report for each predictor being considered for inclusion in the model the number of missing values* ● *The handling of missing values should be reported, including any assumptions for the reason of the missingness* ● *If individuals (or predictors) have been omitted due to the missing values, this should be reported, and reasons given* ● *If missing values have been imputed, then full details of the method for imputing any missing values should be reported* ● *If missing values have been imputed confirm it was done separately for the training and any test data (i.e., avoiding leakage)*
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** ● *Describe how the available data were used to develop the model and to evaluate model performance, including whether and how the data were partitioned, and the reasons for partitioning the data (e.g., model development, hyperparameter tuning, evaluating model performance, internal-external cross-validation)* ● *If the data has been partitioned, report whether sample size requirements (see item 10) were considered during the partitioning, and whether the size of the partitioned data are sufficient to carry out the analyses and answer the research question* ● *If the data has been partitioned into training (including any hyperparameter tuning data) and test data, confirm that there has been no data leakage*

Section/Topic	Item		Checklist item
			• *If the data contain multiple records or samples from the same individual, and if the data has been partitioned into training (including any hyperparameter tuning data) and test data, confirm there has been no leakage of individuals across any of the partitioned data or if not, how describe how this was handled in the analysis (see item 12c)*
	12b	D	**Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation)** • *For any predictors that have been transformed during the analysis, i.e., rescaled, or standardised, describe how this was done* • *For any categorical predictors, where collapsing of categories has been carried out, e.g., due to small sample size/too few outcome events, provide the details and reasons*
	12c	D	**Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation** • *Clearly specify the type of model (or models) being developed (e.g., logistic regression, Cox regression, random forest, neural network) and provide a rationale for using each model building method – consider the type of outcome being predicted and how the prediction model will be implemented in practice* • *For each model, clearly describe all the steps in the model building, including any hyperparameter tuning, what hyperparameters have been tuned and how this was done. If many model building approaches are being applied and word limits prohibit a full description, then use supplementary material to provide the details* • *For studies that are developing more than one model (e.g., using different model building methods), clearly describe the criteria to choose which is the model being put forward (if any), see item 12e and item 23 on model performance)* • *The internal validation approach (to evaluate model performance) during model development should be clearly described, e.g., was k-fold cross validation or bootstrapping used. Clarify whether all model building steps (including hyperparameter tuning) was replayed during the method of internal evaluation* • *Clearly describe any methods (e.g., bootstrapping) used to examine model stability (e.g., in terms of predictor selection, predictive performance and individual predictions) (Riley & Collins, Biom J 2023; 65: 2200302 [DOI: 10.1002/bimj.202200302])* • *If the data contain multiple records or samples from the same individual, describe how this was handled in the model building and internal validation (e.g., if k-fold cross-validation was used, confirm if all records/samples for an individual were included in the same fold (e.g., avoiding data leakage)*
	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 for additional considerations[3]**

[2] Separately for all model building approaches.
[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (e.g., at different hospitals/centres).

Section/Topic	Item		Checklist item
			• *If the analysis has accounted for any clustering in the data (e.g., from combining individual participant data from multiple studies, or data clustered by medical centre/hospital, or country) during the model development or evaluation of model performance, the rationale and methods used to account for clustering should be clearly described* • *For specific reporting recommendations for prediction model studies that have accounted for clustering and heterogeneity in model parameter values and performance, authors should consult the TRIPOD-Cluster checklist (Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018])*
	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** • *Report all the measures used to evaluate model performance. It is generally expected that as a minimum, model discrimination and calibration (including calibration plots) are presented* • *If the prediction model is predicting a time-to-event outcome, then clearly describe the measures and methods that have been used to account for the time-to-event nature (i.e., censoring). Similarly, the handling of any competing risks should also be stated (if applicable)* • *For prognostic models, report all time-points at which the model's predictive performance was evaluated* • *Report the methods used for graphical displays of model performance, such as calibration plots (with smooth calibration curves) and decision curves* • *If multiple models are being compared, i.e., comparing against an existing model or comparing multiple modelling approaches, then the methods used for comparing these models, and the criteria for making any judgements on superior performance should be clearly explained*
	12f	E	**Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings** • *If the model is updated following the validation, such as recalibration or refitting – whether in the entire cohort or in a specific socio demographic group, then provide details on the methods used to update the model*
	12g	E	**For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)** • *For studies evaluating an existing model in a separate data set (i.e., an external validation study), provide details on how the individual predictions from the model were calculated. If a model is not freely/publicly available, explain how the predictions were obtained* • *If a regression model equation was being evaluated, provide details of this equation (e.g., consider presenting this equation, provide a citation to the original study that developed the equation)*

Section/Topic	Item		Checklist item
			• *For studies evaluating a prediction model where there is no equation (e.g., a neural network, random forest), provide details on how the predictions were made, e.g., code, software object, API, and where can this be found (i.e., URL[4], DOI[5])* • *If individual predictions from the model were used to create risk groups or classifications (that were not specified in the model development) then details on how and why this was done should be reported (see item 15)*
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** • *If class imbalance methods (e.g., under/over sampling, SMOTE[6]) have been used, then provide a rationale for doing so, and how this was done – considering any impact on sample size (e.g., for undersampling methods)* • *Imbalance corrections have an impact on model calibration (van den Goorbergh et al, J Am Med Inform Assoc 2022; 29: 1525--1534 [DOI: 10.1093/jamia/ocac093]), yielding probability estimates that are too high (which also has an impact on defining any risk groups), describe the methods used to recalibrate the model or the model predictions*
Fairness	14	D;E	**Describe any approaches that were used to address model fairness and their rationale** • *Fairness refers to ensuring that a prediction model does not discriminate against individuals or groups, for example based on personal attributes such as race, gender, age and all approaches used to address fairness should be clearly explained along with their rationale* • *It is important to ensure the data contains representative groups (of the target population) when developing the model and evaluating its performance and researchers should attempt to demonstrate this* • *If the prediction model is developed using data with underrepresented groups or particular groups not included, then evaluation in these groups in representative data is needed to evaluate the model in these groups, as to increase generalisability to more groups of individuals beyond those in the development and evaluation data*
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** • *Most models output a probability estimate for an individual, whilst some models turn the output into a classification (e.g., into low- or high-risk groups), this should be clearly stated. If classification or risk groups have been created, then the rationale for doing so in the context of the care pathway and how these risk groups inform any clinical decisions should be made* • *For models producing a classification or risk groups, this should be clearly reported, and any thresholds (e.g., range of estimated probabilities defining the groups) should be specified (whether these are based on the literature, clinical guidelines, statistical considerations or ad-hoc)*

[4] URL stands for uniform resource locator
[5] DOI stands for digital object identifier
[6] SMOTE stands for synthetic minority oversampling technique

Section/Topic	Item		Checklist item
			• *If uncertainty intervals for individual prediction model outputs have been presented then provide details on how this was done (e.g., using the variance-covariance matrix of parameter estimates or conformal prediction)*
Development versus evaluation	16	D;E	**Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors** • *Prediction models developed in one setting, centre or country are not necessarily useful in a different setting, centre, or country. Eligibility criteria, outcome and predictors definitions might (intentionally) differ between data from different sources. Describing any differences between the development data and data used to evaluate model performance is useful to understand and interpret the performance and generalisability of the model in the context of the original model development data*
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** • *If the study has no institutional research board or ethics approval, then clearly state so, with reasons*
OPEN SCIENCE			
Funding	18a	D;E	**Give the source of funding and the role of the funders for the present study** • *Provide details on whether the study was funded and provide any details on the role the funder had in the study.* • *Provide any additional funding sources all authors*
Conflicts of interest	18b	D;E	**Declare any conflicts of interest and financial disclosures for all authors** • *Disclose any of the authors' relationships or activities that readers could consider pertinent or that may have influenced the study design, conduct, interpretation, or reporting*
Protocol	18c	D;E	**Indicate where the study protocol can be accessed or state that a protocol was not prepared** • *Provide all details on the availability of the study protocol, including where the study protocol can be found (e.g., publication details, in supplementary material, publicly available in a repository such as on the Open Science Framework), including a URL or DOI* • *Clearly state if no study protocol was developed or publicly available (and reasons)* • *If there are any notable deviations from what was specified in the study protocol, provide a summary and reasons for the deviation*
Registration	18d	D;E	**Provide registration information for the study, including register name and registration number, or state that the study was not registered** • *If the study has been registered (e.g., on clinicaltrials.gov, Open Science Framework), then provide details on the registration number, the name of the register and a link to the registration (including any DOI)* • *Clearly state if the study has not been registered*

Section/Topic	Item		Checklist item
Data sharing	18e	D;E	**Provide details of the availability of the study data** • *Provide details on the availability of the study data, including where the data can be found (e.g., public repository, URL, DOI), how it can be retrieved, any conditions or restrictions on obtaining and using the data. A data dictionary should accompany any shared data.* • *If data cannot be shared, provide reasons as to why* • *Avoid platitudes such as 'Data available upon reasonable request' without specifying conditions for what constitutes a reasonable request*
Code sharing	18f	D;E	**Provide details of the availability of the analytical code**[7] • *Provide all details on the availability of the analytical code (and documentation on how to run the code), including where the code can be found (e.g., code repository, DOI, link), how it can be retrieved, any conditions or licences to obtain and use the code should be reported (and version)* • *The analytical code is all code needed to replicate (in principle) all the reported results and findings of the study (including any code for data cleaning). The software and any packages needed to reproduce (in principle) the study findings should be reported (including any version numbers). In some instances, more details on the computing environment may need to be reported (e.g., hardware, operating system, CPU, RAM)*
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** • *Describe how patients or public were involved in the planning, design, conduct, reporting or dissemination of the study and its findings.* • *Were the findings of the study presented to patients or the public?* • *Consider using the GRIPP2 statement to report patient and public involvement in the research (Staniszewska et al, BMJ 2017; j3453 [DOI: 10.1136/bmj.j3453])* • *If no patients or public were involved in any aspect of the study, then clearly state so*
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**

[7] This relates to the analysis code, e.g., any data cleaning, feature engineering, model building, evaluation.

Section/Topic	Item		Checklist item
			• *A flow diagram can be useful to describe the flow of participants through a study, where the entry point to the flow diagram is the source of participants, and then successive steps can relate to eligibility criteria, follow-up (if applicable). and data availability* • *Other useful information to present in the flow diagram include the number of participants with missing values, and the number of outcome events* • *For studies of prognosis or diagnosis with delayed reference testing, a summary of the follow-up time should be reported (e.g., median follow-up, and range)*
	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** • *Report, possibly using a table, a summary of all data sets used, including the distribution of outcomes, predictors (e.g., mean/median, standard deviation/interquartile range, frequency), any treatments received, the sample size (and number of outcome events, summary of the follow-up time, and for each predictor, the number and proportion of missing values* • *If relevant, it may be useful to report any differences across key demographic groups of interest*
	20c	E	**For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).** • *For studies evaluating the performance of an existing model (including those within a model development study) provide a comparison of the distribution of important variables (e.g., mean/median, standard deviation/interquartile range, frequency), such as demographics, predictors in the model, and outcome, including proportion of missing values. This is probably best presented in a table and consider reporting this by outcome status*
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)** • *The sample size (including the number of outcome events) should be reported for each analysis (i.e., each model development, each model evaluation), as they can often vary across different analyses in a prediction model study (e.g., after data partitioning, model hyperparameter tuning), and particularly in the presence of missing data* • *If the data contain multiple samples or records for an individual report also report the number of individuals*
Model specification	22	D	**Provide details of the full prediction model (e.g., formula, code, object, API) 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)**[8] • *The 'product' of a prediction model development study is the prediction model. It is therefore important to provide details on the model, and how it can be used to allow predictions for new individuals to be made. For example, provide the equation for a*

[8] This relates to the code to implement the model to get estimates of risk for a new individual.

Section/Topic	Item		Checklist item
			regression model, for models developed using methods where the model cannot be 'written down' as an equation, provide details on the availability of code, software objects or API so that others can evaluate this model in their own data, or implement it in daily practice ● *If multiple models have been developed, then provide details on the availability of all models* ● *Explain how to use the model to allow others to make predictions in new individuals.* ● *Provide details of any hardware requirements, and software (and packages) to enable third-party testing, implementation and monitoring* ● *If a model cannot be made publicly available (e.g., for commercial reasons), this should be clearly reported, and any conditions on gaining access to the model to enable predictions to be calculated for new individuals and third-party evaluation should be reported*
Model performance	23a	D;E	**Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation** ● *Estimates of all model performance measures described in item 12e should be presented along with confidence intervals.* ● *Report model performance estimates for the overall population and for any key groups (e.g., sex, ethnicity) of interest (e.g., as part of fairness checks) with confidence intervals* ● *Use plots to present and aid evaluation, such as calibration plots (with smooth calibration curves and distributions of predicted values) and decision curves* ● *Report performance estimates for all evaluations undertaken (e.g., in development data; in evaluation data; from internal validation process, etc), including at each time-point examined (for prognostic models)* ● *Report any examinations of model stability, e.g., in terms of performance estimates and variability of individual predictions across models developed in bootstrap samples (Riley & Collins, Biom J 2023; 65: 2200302 [DOI: 10.1002/bimj.202200302])* ● *Clearly indicate which data have been used to present each performance estimate*
	23b	D;E	**If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details** ● *If the evaluation of model performance has accounted for any clustering in the data (e.g., from combining individual participant data from multiple studies, or data clustered by centre/hospital, or country), the results should be reported, along with confidence intervals (see item 23a)* ● *For specific reporting recommendations for prediction model studies that have accounted for clustering and heterogeneity in model performance, authors should consult the TRIPOD-Cluster checklist (Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018])*
Model updating	24	E	**Report the results from any model updating, including the updated model and subsequent performance** ● *If the prediction model has been updated (e.g., recalibrated, re-fit) following the validation, details of the updated prediction model to enable third-party evaluation and implementation, including any restrictions to access or re-use should be reported (see item 22)*

Section/Topic	Item		Checklist item
			• *The performance of the updated model should be reported (see items 23a, potentially 23b)*
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** • *Interpretation of the study results places the findings in context of other evidence. If there are existing models, then discuss the findings in the context of these existing studies* • *For studies evaluating the performance of an existing prediction model, if existing studies have evaluated the performance of the model, then it's important to discuss and summarise these findings and place them in context* • *Ensure the interpretation of the findings do not go beyond the findings reported from the development and evaluation of the model to prevent overinterpretation or 'spin'* • *It is useful for the reader to understand how performance of the model in the evaluation data compares to the performance of the model in any other evaluation studies of that model. When the results diverge, possible reasons for the difference in model performance should be discussed*
Limitations	26	D;E	**Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability** • *Acknowledgement of limitations is an important aspect of any scientific paper – and can refer to any aspect of the study design, conduct or analysis. Provide a meaningful discussion of the study limitations factoring in any concerns related to representativeness of the data used in the analysis, sample size, overfitting and missing data/data quality*
Usability of the model in the context of current care	27a	D	**Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model** • *Authors should comment on how to handle unavailable predictor values at the moment the model is intended to be used as part of the care pathway in daily practice. Any strategies to impute missing values at the moment the model is intended to be used should also be evaluated (and thus mentioned in the Methods and Results)* • *Similarly, at the point of implementation, authors should discuss (if relevant) the handling of poor quality input data (e.g., image resolution, data format)*
	27b	D	**Discuss 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** • *Provide details on how users are expected or required to interact with the prediction model for the model to be used as intended, for example any considerations for handling the input data* • *Is any expertise or training needed or required to use the model, handle or collect the input data, and if so, provide details*

Section/Topic	Item		Checklist item
	27c	D;E	**Discuss any next steps for future research, with a specific view to applicability and generalizability of the model** • *Are further evaluations of the model needed, e.g., in different populations or subgroups, or is the model ready for evaluation in clinical trials, or implementation as part of the care pathway*

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

TRIPOD+AI

Version: 11-January-2024

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	
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	
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	
	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)	
	3c	D;E	Describe any known health inequalities between sociodemographic groups	
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	
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 representativeness of the data	
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	
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 centres	
	6b	D;E	Describe the eligibility criteria for study participants	
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	
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	
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	
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	
	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)	
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	
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	
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	
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	
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	
	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 for additional considerations[3]	
	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	
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	
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	
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	
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	

[1] 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

[2] Separately for all model building approaches.

[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	
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	
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was 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	
Data sharing	18e	D;E	Provide details of the availability of the study data	
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	
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.	
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.	
	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.	
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	
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)	
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]	
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	
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	
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	
	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	
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

[4] This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.
[5] This relates to the code to implement the model to get estimates of risk for a new individual.

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
1. Title		Identify the study as developing and/or validating (i.e., testing) a multivariable prediction model, the target population, and the outcome to be predicted.	1%	5%	11%	83%	94%
2. Abstract		Provide a summary of objectives, study design or data sources, setting, participants, sample size, predictors/features, outcome, analytical methods, intended use of the prediction model, results, and conclusions.	1%	1%	11%	86%	97%
3. Introduction	a. Background	i. Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the prediction model, including references to existing (AI or non-AI) models, and the main advantages of the used design and analyses.	1%	4%	23%	72%	95%
		ii. Explain the intended purpose (e.g., for prognosis or diagnostic predictions) and use for the AI model in the context of the clinical pathway, including its intended users (e.g., healthcare professionals, patients, public).	2%	4%	25%	69%	94%
	b. Objectives	Specify the study objectives, including whether the study describes the development or validation (e.g., testing) of the model or both.	0%	1%	8%	91%	99%
4. Methods	a. Sources of data	i. Describe the study design or source of data (e.g., randomized trial, cohort, routine care or registry data), separately for the development and validation (test) datasets, if applicable.	0%	2%	8%	90%	98%
		ii. Describe the origin of the data, and how data were identified, requested and collected.	0%	8%	20%	72%	92%
		iii. Specify the key dates of the collected participant data, including start and of participant/data accrual; and, if applicable, end of follow-up.	2%	9%	31%	57%	88%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	b. Participants	i. Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres or data sources.	0%	6%	28%	66%	94%
		ii. Describe the eligibility criteria for participants or data sources: how, where, and when potentially eligible participants were identified (e.g., symptoms, results from previous tests, inclusion in the registry, patient-care setting, location).	1%	4%	16%	79%	95%
		iii. Give details of treatments received, and how they were handled, if relevant.	5%	10%	41%	45%	86%
	c. Data preparation (predictors/feature and outcome/label)	Describe any data pre-processing steps, including any cleaning, harmonisation, sampling, linkage, de-identification methods, and quality checks.	2%	6%	21%	72%	93%
	d. Outcome/labelling	i. Clearly define the outcome (e.g., ground truth or reference standard) that is predicted by the prediction model (including the time horizon), including how and when assessed and the rationale for choosing this outcome measurement (if alternatives exist).	0%	0%	4%	96%	100%
		ii. Describe the qualifications of the assessors of the outcomes which require subjective interpretation and whether additional materials were used.	2%	13%	41%	44%	85%
		iii. Report any actions to blind assessment of the outcome to be predicted (e.g., details on what, if any, information was used to assess/label the outcome).	3%	14%	35%	48%	83%
		iv. Describe any measurement of inter- and intra- rater variability for outcomes requiring subjective interpretation (e.g., imaging), including methods to mitigate variability or resolve discrepancies.	2%	15%	42%	41%	83%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	e. Predictors	i. Clearly define all predictors/features used in developing the multivariable prediction model, including how and when they were measured. Consider using supplementary material for large numbers of predictors.	0%	5%	12%	83%	95%
		ii. Report the source of predictors and any known biases that may be associated with them.	4%	13%	34%	49%	83%
		iii. Describe the qualifications of the assessors of the predictors/features which require subjective interpretation and whether additional materials were used.	7%	25%	36%	32%	68%
		iv. Report any actions to blind assessment of predictors for the outcome and other predictors.	7%	21%	34%	39%	73%
		v. Describe any measurement of inter- and intra- rater variability for predictors requiring subjective interpretation (e.g., imaging), including methods to mitigate variability and/or resolve discrepancies.	5%	24%	36%	34%	70%
	f. Sample size	Explain how the study size was arrived at (and provide a justification that the study size was sufficient). Report assumptions and estimates to support any sample size calculation.	3%	11%	28%	58%	86%
	g. Missing data	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation or other data augmentation method.	0%	2%	16%	81%	97%
	h. Analytical methods	i. Consider a diagram to illustrate the analytical processes.	10%	19%	38%	33%	71%
		ii. Describe how predictors/features were handled in the analyses (functional form and any standardisation).	1%	7%	25%	67%	92%
		iii. Describe any pre-selection of predictors/features prior to model building.	1%	6%	23%	70%	93%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		iv. Describe any predictor/feature rescaling or transformations prior to model building.	3%	10%	29%	58%	87%
		v. Specify the type of model, all model-building procedures (including any predictor selection), and method for internal validation (e.g., bootstrapping, cross-validation).	0%	4%	9%	88%	97%
		vi. Describe any ensemble techniques (e.g., to combine model predictions), if applicable.	4%	11%	24%	62%	86%
		vii. Provide a detailed description of the model, including inputs, outputs, all intermediate layers and connections.	4%	12%	25%	59%	84%
		viii. Describe the initialization of model parameters (e.g., randomization, transfer learning).	5%	10%	30%	54%	84%
		ix. Details of model training approaches, including hyperparameters, number of models trained, used data sets.	2%	6%	19%	73%	92%
		x. Describe how any heterogeneity in the model parameter values was handled across clusters (e.g., hospitals, countries).	2%	13%	38%	47%	85%
		xi. Describe how any heterogeneity in model performance was handled and quantified across clusters (e.g., hospitals, countries).	3%	12%	38%	47%	85%
		xii. Specify all measures used to assess model performance (e.g., discrimination, calibration) and, if relevant, to compare multiple models.	1%	2%	9%	88%	97%
		xiii. Describe any model updating (e.g., recalibration) arising from the validation (testing), either overall or for particular populations or settings.	2%	4%	25%	68%	93%
		xiv. Describe the method of selecting the final model.	1%	3%	13%	84%	97%
		xv. Describe methods for explainability or interpretability and how they were validated.	4%	13%	33%	49%	82%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		xvi. For external validation/testing (outside the model development study), describe how the predictions by the model were calculated (e.g., using code made publicly available on GitHub).	4%	8%	24%	64%	88%
	i. Risk groups	Provide details on how risk groups were created, if done.	5%	15%	29%	51%	80%
	j. Model development (e.g., training) versus validation (e.g., testing)	For external validation (testing), identify any differences from the development (training) data in setting, eligibility, criteria, outcome, and predictors.	1%	4%	15%	81%	96%
	k. Software	Provide details of any software libraries, frameworks, or packages used.	2%	6%	19%	73%	92%
5. Results	a. Participants	i. 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.	1%	5%	19%	75%	94%
		ii. Report the characteristics overall and where applicable for each data source or setting, including the key dates, key predictors/features (including demographics, ethnicity), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data.	1%	4%	20%	76%	96%
		iii. For external validation (testing), show a comparison with the development data of the distribution of important predictors/features (demographics, predictors, and outcome).	2%	9%	22%	67%	89%
	b. Model development	Specify the number of participants and outcome events in each analysis (e.g., model building, parameter tuning, testing).	1%	2%	23%	74%	97%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	c. Model specification	i. Provide details on the full prediction model to allow predictions for individuals to allow third-party evaluation and implementation (e.g., regression coefficients, input parameters, sharing of code/any dependencies). Provide reasons for not sharing code.	2%	7%	23%	67%	90%
		ii. Specify the output of the AI model (e.g., probabilities, classification, risk grouping).	3%	3%	9%	85%	94%
	d. Model performance	i. Report performance measures (with confidence intervals, CIs) for the prediction model.	0%	2%	8%	90%	98%
		ii. Report results of any heterogeneity across clusters in model performance.	2%	13%	38%	46%	84%
		iii. Describe the results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, explain why not.	4%	21%	37%	38%	75%
	e. Model updating	Report the results from any model updating (including the updated model and subsequent performance), overall and for each cluster (e.g., hospital, if done).	6%	11%	28%	55%	83%
	f. Usability of the model	i. Explain how (and when in the clinical pathway) to use the prediction AI model to permit third-party testing and implementation.	7%	14%	24%	55%	79%
		ii. Describe how the AI model will be integrated into the target setting and clinical pathway, including any onsite or offsite requirements (e.g., hardware/software requirements).	12%	20%	34%	34%	68%
		iii. Describe how poor quality or unavailable input data will be assessed and handled when implementing the AI model.	8%	18%	38%	36%	74%
		iv. Specify whether human-AI interaction will be required in the handling of the input data (e.g., steps required to go from the raw clinical data into a form usable by the model), and what level of expertise is required of users.	11%	18%	30%	41%	71%

Supplementary Table 3: Aggregated TRIPOD+AI responses from Delphi round 1

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	g. Sensitivity analysis	Report results from any subgroup or sensitivity analysis.	2%	16%	35%	48%	83%
6. Discussion	a. Limitations	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on biases, statistical uncertainty, and generalizability.	0%	2%	9%	90%	99%
	b. Interpretation	i. For external validation (testing), discuss the results with reference to performance in the development data, and any other validation data.	1%	7%	23%	69%	92%
		ii. Give an overall interpretation of the main results, including heterogeneity across clusters in model performance, in the context of the objectives and previous studies (e.g., any comparisons to existing prediction models).	1%	7%	19%	73%	92%
	c. Implications	i. Discuss the potential use of the model and implications for future research, with a specific view to generalizability and applicability of the model across different settings or (sub)populations.	2%	8%	27%	62%	89%
		ii. Make clear how the AI/ML model and its outputs may be used and change clinical practice and work-up.	5%	12%	32%	51%	83%
7. Other	a. Supplementary information	i. Provide information about the availability of supplementary resources (e.g., study protocol, data sets).	2%	9%	25%	63%	88%
		ii. State whether and how the AI model and/or its code can be accessed, including any restrictions to access or re-use.	4%	6%	18%	71%	89%
	b. Funding	Give the source of funding and the role of the funders for the present study.	1%	1%	10%	87%	97%

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
1. Title		Identify the study as developing or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1%	4%	15%	80%	95%
2. Abstract		Provide a summary of objectives, study design, data sources, setting, participants, sample size, predictors, outcome, analytical methods, intended use of the prediction model, availability of the model, results, and conclusions.	0%	2%	14%	84%	98%
3. Introduction	a. Background	i. Explain the clinical context (including whether diagnostic or prognostic) and rationale for developing or validating the prediction model, including references to existing models.	1%	1%	15%	84%	99%
		ii. Explain the intended purpose (e.g., for prognosis or diagnostic predictions) and use for the prediction model in the context of the clinical pathway, including its intended users (e.g., healthcare professionals, patients, public).	1%	4%	25%	70%	95%
		iii. Describe the key population groups relevant to the disease outcome, including any known health inequalities between demographic groups.	4%	9%	42%	46%	88%
	b. Objectives	Specify the study objectives, including whether the study describes the development or validation of the model or both.	0%	3%	7%	90%	97%
4. Methods	a. Source of data	i. Describe the study design, source of data (e.g., randomized trial, cohort, routine care or registry data), separately for the development and validation datasets, and the rationale (and representativeness of the target clinical population) for using these data.	0%	1%	8%	91%	99%
		ii. Specify the key dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up.	1%	5%	29%	65%	94%
	b. Participants	i. Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres.	1%	6%	23%	70%	93%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		ii. Describe the eligibility criteria for participants: how, where, and when potentially eligible participants were identified (e.g., symptoms, results from previous tests, inclusion in the registry, patient-care setting, location).	1%	3%	13%	83%	96%
		iii. Give details of any treatments received, and how they were handled during model development or validation, if relevant.	3%	12%	26%	59%	85%
	c. Data preparation	Describe any data pre-processing steps, including any cleaning, feature engineering, harmonisation, sampling, linkage, de-identification methods, augmentation and exclusions or changes made based on data quality data quality. Include details on whether data quality was similar across relevant demographic groups.	1%	4%	16%	80%	96%
	d. Outcome definition	i. Clearly define the outcome that is being predicted (and the time horizon), including how and when assessed, the rationale for choosing this outcome definition (if alternatives exist), and whether the method of outcome assessment is consistent across demographic groups.	0%	0%	5%	95%	100%
		ii. In case of outcome assessment requiring subjective interpretation, describe the qualifications of the outcome assessors and any measurement of inter- or intra- rater variability.	2%	9%	33%	56%	89%
		iii. Report any actions to blind assessment of the outcome to be predicted.	5%	9%	37%	49%	86%
	e. Predictors (features)	i. Clearly define all predictors, including how and when they were measured, any actions to blind assessment of predictors for the outcome and other predictors. Consider using supplementary material for large numbers of predictors.	1%	4%	14%	81%	95%
		ii. Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors prior to model building.	3%	6%	28%	63%	91%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		iii. In case of predictor measurement requiring subjective interpretation, describe the qualifications of the predictor assessors and any measurement of inter- or intra-rater variability. Including any methods to mitigate or resolve discrepancies and materials needed for the predictor measurement. Consider whether the method of predictor measurement is consistent across demographic groups.	3%	9%	38%	50%	88%
	f. Class imbalance	If class imbalance was addressed, state why and how this was done, and any methods to recalibrate the model or the model predictions	2%	6%	23%	69%	92%
	g. Sample size	Explain how the study size was arrived at (separately for development and validation), including details of any sample size calculation. Provide a justification that the study size was sufficient to answer the research question.	1%	7%	25%	67%	92%
	h. Missing data	Describe in detail how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation, surrogate splits, pattern submodels). Provide reasons for omitting any data.	0%	2%	14%	84%	98%
	i. Analytical methods	i. Describe how the data were used in the analysis (e.g., any partitioning, and how this was done considering any sample size requirements, see item 4g). Provide details on how the data were used for any internal validation.	1%	4%	11%	85%	96%
		ii. Consider adding a diagram to illustrate the analytical processes.	8%	19%	49%	24%	73%
		iii. Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation or any standardisation).	2%	3%	28%	67%	95%
		iv. Specify the type of model (and its rationale), how multinomial outcomes, survival (e.g., censored observations), or competing risks were handled (if appropriate), and all model-building procedures (including any data-driven predictor selection), and considering the following issues (where applicable).	1%	4%	17%	78%	95%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		a) Describe the initialization of any model parameters (e.g., random, transfer learning).	1%	6%	31%	62%	93%
		b) Give a detailed description of the model, including inputs, outputs, and any intermediate layers and connections.	4%	11%	24%	61%	85%
		c) Describe the model training, including the procedure for any hyperparameter tuning, and number of models trained.	1%	4%	21%	74%	95%
		d) Describe the method of selecting the final model.	0%	2%	15%	83%	98%
		e) Describe any ensemble techniques (e.g., to combine model predictions).	2%	7%	28%	62%	90%
		f) Give details, including rationale, for any continual learning.	5%	12%	38%	45%	83%
		v. Describe how any heterogeneity in the model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations.	4%	5%	42%	49%	91%
		vi. Specify all measures used (and their rationale) to assess model performance (e.g., discrimination, calibration, net benefit) and, if relevant, to compare multiple models.	0%	5%	11%	84%	95%
		vii. Describe the methods for any analysis of prediction errors carried out. If no such analysis was carried out, explain why not.	3%	10%	31%	57%	88%
		viii. Describe any model updating (e.g., recalibration) arising from the validation, either overall or for particular populations (see item 4j on fairness) or settings.	2%	7%	30%	60%	90%
		ix. For external validation, describe how the model predictions were calculated (e.g., using code made publicly available).	3%	3%	21%	73%	94%
	j. Fairness	Provide details on how steps to recognise and avoid bias during the study, and whether decisions were taken during model development and validation to prioritise fairness and reduce health inequality.	5%	12%	28%	56%	84%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	k. Model output and risk groups	Specify the output of the prediction model (e.g., probabilities, classification, risk grouping. Provide details (and rationale) for any risk groups/classifications, and how the thresholds were identified.	1%	2%	17%	81%	98%
	l. Development versus validation	For validation, identify any differences from the development (training) data in healthcare setting, eligibility, criteria, outcome, and predictors.	2%	2%	21%	74%	95%
	m. Ethical approval	Name the institutional research board or ethics committee that approved the study and provide a description of participant informed consent or an ethics committee waiver of informed consent.	3%	7%	16%	74%	90%
	n. PPI	Provide details of any patient and public involvement during the design, conduct, reporting (and interpretation) or dissemination of the study.	11%	20%	39%	30%	69%
5. Open Science	o. Protocol and registration	Provide details where the study protocol can be publicly accessed or state that a protocol was not prepared. Give details of registration and name of the registry.	6%	10%	33%	51%	84%
	p. Data	Provide details on availability of any data, or reasons for not sharing.	4%	10%	25%	61%	86%
	q. Code	Provide details of any software libraries, frameworks, or packages used in the work, as well as versions and licences under which they are available. Report the availability of any analytical code or previously unpublished code, provide the DOI or link if available. Give reasons for not sharing.	3%	11%	25%	62%	87%
6. Results	a. Participants	i. 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.	3%	5%	15%	78%	93%
		ii. 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.	1%	4%	18%	77%	95%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
		iii. For validation, show a comparison with the development data of the distribution of important predictors/features (demographics, predictors, and outcome).	2%	8%	26%	63%	89%
	b. Model development	Specify the number of participants and outcome events in each analysis (e.g., model building, parameter tuning, validation).	2%	5%	14%	80%	94%
	c. Model specification	i. Provide details on the full prediction model (e.g., regression coefficients, input parameters, sharing of code/any dependencies), to allow predictions in new individuals, and to enable third-party evaluation and implementation.	2%	9%	18%	71%	89%
		ii. Give details on where the code (for the model) can be accessed, including any restrictions to access or re-use (e.g., freely available, proprietary). Report reasons for not sharing code.	4%	8%	24%	64%	88%
	d. Model performance	i. Report performance measures (with confidence intervals) for the prediction model.	0%	1%	5%	95%	100%
		ii. Report the results (e.g., model performance) in key subgroups (e.g., race, ethnicity, age, gender, sex) and any subgroups relevant to the outcome.	2%	9%	26%	63%	89%
		iii. Report results of any heterogeneity in model performance across clusters.	3%	8%	34%	55%	89%
		iv. Describe the results of any analysis of prediction errors and how errors were identified.	3%	8%	35%	54%	89%
	e. Model updating	Report the results from any model updating (including the updated model and subsequent performance), overall and for each cluster (e.g., hospital, if done).	2%	13%	28%	58%	86%
7. Discussion	a. Limitations	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on biases, statistical uncertainty, and generalizability.	0%	1%	8%	91%	99%

Supplementary Table 4: Aggregated TRIPOD+AI responses from Delphi round 2

			Can be omitted	Possibly include	Desirable for inclusion	Essential for inclusion	Desirable + Essential
	b. Interpretation	i. Give an overall interpretation of the main results, including issues of fairness, heterogeneity across clusters in model performance, in the context of the objectives and previous studies (e.g., any comparisons to existing prediction models).	1%	2%	17%	80%	97%
		ii. For external validation, discuss the results with reference to performance in the development data, and any other validation data.	1%	4%	20%	75%	95%
	c. Usability of the model in the context of current care	i. Explain how (and when in the clinical pathway) to use the prediction model including any onsite or offsite requirements (e.g., hardware/software requirements to permit third-party testing and implementation.	2%	12%	26%	60%	86%
		ii. Describe how poor quality or unavailable input data (e.g., predictors) should be assessed and handled when implementing the prediction model.	5%	15%	34%	46%	80%
		iii. Specify whether human-AI interaction will be required in the handling of the input data (e.g., steps required to go from the raw clinical data into a form usable by the model), and what level of expertise is required of users.	5%	14%	38%	44%	82%
		iv. Discuss the potential use of the model and implications and need for future research, with a specific view to generalizability and applicability of the model (e.g., across different settings or (sub)populations).	2%	8%	30%	61%	91%
8. Other	d. Non-technical summary	Consider providing a non-technical summary/box including the aim of the model (including the outcome being predicted), the population on whom it was trained and validated, and the results of model performance. Describe any future work required before the model can be implemented.	10%	17%	35%	38%	73%
	a. Conflicts of interest	State any conflicts of interest and financial disclosures for all authors.	2%	4%	5%	88%	93%
	b. Funding	Give the source of funding and the role of the funders for the present study.	2%	4%	14%	81%	95%

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378