Supplementary Table S1. Search strategy

PubMed	
#1	"stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("gastric"[Title/Abstract] AND "cancer"[Title/Abstract]) OR "gastric cancer"[Title/Abstract] OR ("stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("stomach"[Title/Abstract] AND "cancer"[Title/Abstract]) OR "stomach cancer"[Title/Abstract]) OR (("gastrics"[Title/Abstract] OR "stomach"[MeSH Terms] OR "stomach"[Title/Abstract] OR "gastric"[Title/Abstract]) AND ("carcinoma"[MeSH Terms] OR "carcinoma"[Title/Abstract] OR "carcinomas"[Title/Abstract] OR "carcinoma s"[Title/Abstract])) OR (("stomach"[MeSH Terms] OR "stomach"[Title/Abstract] OR "stomachs"[Title/Abstract] OR "stomach s"[Title/Abstract] OR "stomachal"[Title/Abstract] OR "stomaches"[Title/Abstract]) AND ("carcinoma"[MeSH Terms] OR "carcinoma"[Title/Abstract] OR "carcinomas"[Title/Abstract] OR "carcinoma s"[Title/Abstract])) OR ("stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("gastric"[Title/Abstract] AND "neoplasm"[Title/Abstract]) OR "gastric neoplasm"[Title/Abstract]) OR ("stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("stomach"[Title/Abstract] AND "neoplasm"[Title/Abstract]) OR "stomach neoplasm"[Title/Abstract]) OR ("stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("stomach"[Title/Abstract] AND "tumor"[Title/Abstract]) OR "stomach tumor"[Title/Abstract]) OR ("stomach neoplasms"[MeSH Terms] OR ("stomach"[Title/Abstract] AND "neoplasms"[Title/Abstract]) OR "stomach neoplasms"[Title/Abstract] OR ("gastric"[Title/Abstract] AND "tumor"[Title/Abstract]) OR "gastric tumor"[Title/Abstract])
#2	(("artificial intelligence"[MeSH Terms] OR ("artificial"[Title/Abstract] AND "intelligence"[Title/Abstract]) OR "artificial intelligence"[Title/Abstract]) OR ("deep learning"[MeSH Terms] OR ("deep"[Title/Abstract] AND "learning"[Title/Abstract]) OR "deep learning"[Title/Abstract])) OR ("machine learning"[MeSH Terms] OR ("machine"[Title/Abstract] AND "learning"[Title/Abstract]) OR "machine learning"[Title/Abstract])
#3	"mortality"[MeSH Subheading] OR "mortality"[Title/Abstract] OR "survival"[Title/Abstract] OR "survival"[MeSH Terms] OR "survivability"[Title/Abstract] OR "survivable"[Title/Abstract] OR "survivals"[Title/Abstract] OR "survive"[Title/Abstract] OR "survived"[Title/Abstract] OR "survives"[Title/Abstract] OR "surviving"[Title/Abstract]
	#1 AND #2 AND #3 Results: N=442
Web of Science	
#1	(((((((((TS=(survival)) OR TS=(mortality)) AND TS=(gastric cancer)) OR TS=(stomach cancer)) OR TS=(gastric neoplasm)) OR TS=(stomach neoplasm)) OR TS=(gastric tumor)) OR TS=(stomach tumor)) OR TS=(gastric carcinoma)) OR TS=(stomach carcinoma)

#2	((TS=(machine learning)) OR TS=(deep learning)) OR TS=(artificial intelligence)
#3	(TS=(survival)) OR TS=(mortality)
	#1 AND #2 AND #3 Results: N=935

Supplementary Table S2: CHARMS Checklist.

Domain	S A Rahman (2021)	Reported on page #
SOURCE OF DATA	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,4
	Details of treatments received, if relevant	2,4
	Study dates	2,4
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation)	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	3
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)	3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques)	3-4
	Modelling assumptions satisfied	3-4
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3-4
	Method for selection of predictors **during multivariable modelling** (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion)	3-4
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3-4
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals	3-4
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	3-4
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	4
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	4

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-6
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-6
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-6
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	6-8
	Comparison with other studies, discussion of generalizability, strengths and limitations.	6-8

Domain	Tao Chen (2019)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,4
	Details of treatments received, if relevant	2,4
	Study dates	2,4
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	3
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	3
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	5-6
	Comparison with other studies, discussion of generalizability, strengths and limitations.	5-6

Domain	Mengxin Tian (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2-3
	Participant description	2-3,7
	Details of treatments received, if relevant	2-3,7
	Study dates	2-3
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	3
	Was the same outcome definition (and method for measurement) used in all patients?	3
	Type of outcome (e.g., single or combined endpoints)	3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	3
	Time of outcome occurrence or summary of duration of follow-up	3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	3
	Definition and method for measurement of candidate predictors	3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	3
SAMPLE SIZE	Number of participants and number of outcomes/events	2-3
	Number of outcomes/events in relation to the number of candidate predictors (Events	3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	3
	Number of participants with missing data for each predictor	3
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	3
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3-4
	Modelling assumptions satisfied	3-4
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3-4
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3-4
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3-4
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	4
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	4
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	4
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	4

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	5-11
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	5-11
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	5-11
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	11-13
	Comparison with other studies, discussion of generalizability, strengths and limitations.	11-13

Domain	Warid Islam (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2
	Details of treatments received, if relevant	2
	Study dates	2
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	3
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	3
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	5-6
	Comparison with other studies, discussion of generalizability, strengths and limitations.	5-6

Domain	Hongcai Chen (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2
	Details of treatments received, if relevant	2
	Study dates	2
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2
	Definition and method for measurement of candidate predictors	2
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2
	Modelling assumptions satisfied	2
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	2,4-5
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	2,4-5
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	2
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	2

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	3-5
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	3-5
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	3-5
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	6-9
	Comparison with other studies, discussion of generalizability, strengths and limitations.	6-9

Domain	Naoki Kuwayama (2023)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2
	Details of treatments received, if relevant	2
	Study dates	2
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2
	Definition and method for measurement of candidate predictors	2
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	4-5
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	4-5
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	2-3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	2-3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	3-5
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	3-5
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	3-5
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	5-9
	Comparison with other studies, discussion of generalizability, strengths and limitations.	5-9

Domain	Junjie Zeng (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,4
	Details of treatments received, if relevant	2,4
	Study dates	2,4
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2
	Definition and method for measurement of candidate predictors	2
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	5-7
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	5-7
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-7
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-7
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-7
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	7-10
	Comparison with other studies, discussion of generalizability, strengths and limitations.	7-10

Domain	Mengjie Wu (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,5
	Details of treatments received, if relevant	2,5
	Study dates	2,5
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2-3
	Was the same outcome definition (and method for measurement) used in all patients?	2-3
	Type of outcome (e.g., single or combined endpoints)	2-3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2-3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2-3
	Time of outcome occurrence or summary of duration of follow-up	2-3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2-3
	Number of outcomes/events in relation to the number of candidate predictors (Events	2-3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2-3
	Number of participants with missing data for each predictor	2-3
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2-3
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3-4
	Modelling assumptions satisfied	3-4
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3-4
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3-4
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3-4
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	10-12
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	10-12
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3-4
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3-4

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-5
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-5
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-5
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	5-8
	Comparison with other studies, discussion of generalizability, strengths and limitations.	5-8

Domain	Xunjun Li (2022)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2-3
	Details of treatments received, if relevant	2-3
	Study dates	2-3
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2
	Definition and method for measurement of candidate predictors	2
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	4-5
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	4-5
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	2
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	2

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	2-5
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	2-5
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	2-5
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	5-7
	Comparison with other studies, discussion of generalizability, strengths and limitations.	5-7

Domain	Rocío Aznar-Gimeno (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	3
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	3
	Participant description	3,7-8
	Details of treatments received, if relevant	3,7-8
	Study dates	3,7-8
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	3
	Was the same outcome definition (and method for measurement) used in all patients?	3
	Type of outcome (e.g., single or combined endpoints)	3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	3
	Time of outcome occurrence or summary of duration of follow-up	3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	3-4
	Definition and method for measurement of candidate predictors	3-4
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	3-4
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	3-4
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	3-4
SAMPLE SIZE	Number of participants and number of outcomes/events	3
	Number of outcomes/events in relation to the number of candidate predictors (Events	3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	3
	Number of participants with missing data for each predictor	3
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	3
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	4-7
	Modelling assumptions satisfied	4-7
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	4-7
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	4-7
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	4-7
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	11-13
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	11-13
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	4-7
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	4-7

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	7-17
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	7-17
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	7-17
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	17-21
	Comparison with other studies, discussion of generalizability, strengths and limitations.	17-21

Domain	Yuming Jiang (2022)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,5
	Details of treatments received, if relevant	2,5
	Study dates	2,5
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	3
	Number of outcomes/events in relation to the number of candidate predictors (Events	3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3-4
	Modelling assumptions satisfied	3-4
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3-4
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3-4
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3-4
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	3-4
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	3-4
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	4
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	4

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-6
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-6
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-6
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	6-8
	Comparison with other studies, discussion of generalizability, strengths and limitations.	6-8

Domain	Fan Li (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2
	Details of treatments received, if relevant	2
	Study dates	2
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2
	Was the same outcome definition (and method for measurement) used in all patients?	2
	Type of outcome (e.g., single or combined endpoints)	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2
	Time of outcome occurrence or summary of duration of follow-up	2
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	2-3
	Modelling assumptions satisfied	2-3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	2-3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	2-3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	2-3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	3-4
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	3-4
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	2-3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	2-3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	5-8
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	5-8
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	5-8
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	8-9
	Comparison with other studies, discussion of generalizability, strengths and limitations.	8-9

Domain	Tianbao Liao (2024)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2-3
	Participant description	2-3,5
	Details of treatments received, if relevant	2-3,5
	Study dates	2-3,5
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	3
	Was the same outcome definition (and method for measurement) used in all patients?	3
	Type of outcome (e.g., single or combined endpoints)	3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	3
	Time of outcome occurrence or summary of duration of follow-up	3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3
	Modelling assumptions satisfied	3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	6-7
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	6-7
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-6
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-6
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-6
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	6-10
	Comparison with other studies, discussion of generalizability, strengths and limitations.	6-10

Domain	Ting Wei (2022)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2-3
	Participant description	2-3,6
	Details of treatments received, if relevant	2-3,6
	Study dates	2-3,6
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2-3
	Was the same outcome definition (and method for measurement) used in all patients?	2-3
	Type of outcome (e.g., single or combined endpoints)	2-3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2-3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2-3
	Time of outcome occurrence or summary of duration of follow-up	2-3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2
	Number of outcomes/events in relation to the number of candidate predictors (Events	2
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2
	Number of participants with missing data for each predictor	2
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3-4
	Modelling assumptions satisfied	3-4
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3-4
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3-4
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3-4
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	7
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	7
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	6-9
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	6-9
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	6-9
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	9-11
	Comparison with other studies, discussion of generalizability, strengths and limitations.	9-11

Domain	Mohammad Reza Afrash (2023)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	3
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	3-4
	Participant description	3-4
	Details of treatments received, if relevant	3-4
	Study dates	3-4
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	4
	Was the same outcome definition (and method for measurement) used in all patients?	4
	Type of outcome (e.g., single or combined endpoints)	4
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	4
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	4
	Time of outcome occurrence or summary of duration of follow-up	4
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	4
	Definition and method for measurement of candidate predictors	4
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	4
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	4
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	4
SAMPLE SIZE	Number of participants and number of outcomes/events	3
	Number of outcomes/events in relation to the number of candidate predictors (Events	3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	3
	Number of participants with missing data for each predictor	3
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	3
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	5-6
	Modelling assumptions satisfied	5-6
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	5-6
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	5-6
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	5-6
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	9-12
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	9-12
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	6
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	6

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	6-12
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	6-12
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	6-12
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	12-14
	Comparison with other studies, discussion of generalizability, strengths and limitations.	12-14

Domain	Junjie Zeng (2023)	Reported on page #
SOURCE OF	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	2
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	2
	Participant description	2,5
	Details of treatments received, if relevant	2,5
	Study dates	2,5
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	2-3
	Was the same outcome definition (and method for measurement) used in all patients?	2-3
	Type of outcome (e.g., single or combined endpoints)	2-3
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	2-3
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	2-3
	Time of outcome occurrence or summary of duration of follow-up	2-3
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	2-3
	Definition and method for measurement of candidate predictors	2-3
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment	2-3
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	2-3
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	2-3
SAMPLE SIZE	Number of participants and number of outcomes/events	2-3
	Number of outcomes/events in relation to the number of candidate predictors (Events	2-3
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	2-3
	Number of participants with missing data for each predictor	2-3
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	2-3
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning	3
	Modelling assumptions satisfied	3
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	3
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike	3
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	3
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination	9-11
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	9-11
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	3
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	3

RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	4-7
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	4-7
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	4-7
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	7-11
	Comparison with other studies, discussion of generalizability, strengths and limitations.	7-11