Supplemental File 1. PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Supplemental File 2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Abstract and p. 7
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7-8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplemental file 3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8-9
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8-9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9; Supplemental Files 4 and 5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9; Supplemental File 5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8-9
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	11; Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11-13; Table 3, 4, 5
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	13-14; Figures 3 and 4

Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	11-14; Table 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	15-19
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	18
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	19
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	20

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Supplemental File 2. PICOTS System

Population	Pulmonary tuberculosis cases
Intervention	Any prognostic model developed to predict tuberculosis treatment outcome. This includes model development studies with and without external validation
Comparator	Models will be compared to each other, as there is no other relevant comparator for this systematic review
Outcome	TB treatment outcome. The primary outcome of interest is the probability of unsuccessful TB treatment outcome, defined by the WHO as the combination of death, treatment failure, default, and/or not evaluated, as compared to successful TB treatment outcome, defined as the combination of cure and treatment completion. Included studies should evaluate at least one of the following outcomes: cure, treatment completion, death, treatment failure, default, and not evaluated. Default and not evaluated are sometimes referred to collectively as lost to follow-up. Some prediction models will look at only single endpoints, whereas other look at composite outcomes.
Timing	The timespan of prediction may vary between studies, depending on the duration of treatment and follow-up, but we expect most studies will evaluate endpoints around 6-9 months.
Setting	Model designed for use in clinical or hospital setting at the time of TB treatment initiation to aid in targeted treatment or programmatic support for individuals at greatest risk for unsuccessful TB treatment outcomes.

Supplemental File 3. Search Strategy

Database	*Search terms*
PubMed	1. ((validat*[tiab] OR predict*[ti] OR rule*[tiab]) OR (predict*[tiab] AND (outcome*[tiab] OR risk*[tiab] OR model*[tiab])) OR ((history[tiab] OR variable*[tiab] OR criteria[tiab] OR scor*[tiab] OR characteristic*[tiab] OR finding*[tiab] OR factor*[tiab]) AND (predict*[tiab] OR model*[tiab] OR decision*[tiab] OR identif*[tiab] OR prognos*[tiab])) OR (decision*[tiab] AND (model*[tiab] OR clinical*[tiab] OR "Logistic Models"[Mesh])) OR (prognostic[tiab] AND (history[tiab] OR variable*[tiab] OR criteria[tiab] OR scor*[tiab] OR characteristic*[tiab] OR finding*[tiab] OR factor*[tiab] OR model*[tiab])) 2. (stratification[tiab] OR "ROC Curve"[Mesh] OR discrimination[tiab] OR discriminate[tiab] OR "c-statistic"[tiab] OR "c statistic"[tiab] OR "area under the curve"[tiab] OR AUC[tiab] OR calibration[tiab] OR indices[tiab] OR algorithm[tiab] OR multivariable[tiab]) 3. (tuberculosis[Mesh] OR tuberculosis[tiab]) 4. (outcome*[tiab] OR mortality*[tiab] OR death*[tiab] OR fail*[tiab] OR recur*[tiab] OR relapse*[tiab] OR default*[tiab] OR abandon*[tiab] OR loss*[tiab] OR cure*[tiab] OR success*[tiab] OR unsuccess*[tiab] OR die[tiab] OR died[tiab] OR dies[tiab])) 5. 1 OR 2 6. 3 AND 4 7. 5 AND 6 AND (humans[Filter]) AND ("1995"[Date - Publication] : "3000"[Date - Publication])
Embase	1. (validat$ or predict$ or rule$).ti. OR (predict$ and (outcome$ or risk$ or model$)).ti,ab. OR ((history or variable$ or criteria or scor$ or characteristic$ or finding$ or factor$) and (predict$ or model$ or decision$ or identif$ or prognos$)).ti,ab. OR (decision$.ti,ab. and ((model$ or clinical$).ti,ab. or "statistical model"/)) OR (prognostic and (history or variable$ or criteria or scor$ or characteristic$ or finding$ or factor$ or model$)).ti,ab. 2. (stratification or discrimination or discriminate or c-statistic or "c statistic" or "area under the curve" or AUC or calibration or indices or algorithm or multivarriable).ti,ab. or "receiver operating characteristic"/ 3. tuberculosis/ or tuberculosis.ti,ab 4. (outcome$ or mortality$ or death$ or fail$ or recur$ or relapse$ or default$ or abandon$ or loss$ or cure$ or success$ or unsuccess$ or die or died or dies).ti,ab. 5. 1 or 2 6. 3 and 4 7. 5 and 6 8. limit 7 to (human and yr="1995 -Current")
Web of Science	1. TI=(validat* or predict*. or rule*) OR TS=(predict* and (outcome* or risk* or model*)) OR TS=((history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)) OR TS=(decision* and ((model* or clinical*). or "statistical model")) OR TS=(prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)) 2. TS=(stratification or discrimination or discriminate or c-statistic or "c statistic" or "area under the curve" or AUC or calibration or indices or algorithm or multivariable or "receiver operating characteristic") 3. TS=(tuberculosis) 4. TS=(outcome* or mortality* or death* or fail* or recur* or relapse* or default* or abandon* or loss* or cure* or success* or unsuccess* or die or died or dies) 5. 1 or 2 6. 3 and 4 7. 5 and 6; IC Timespan=1995-2019
Google scholar	tuberculosis treatment outcome prediction prognostic model development validation

Supplemental File 4. CHARMS Checklist

Domain	Key items	Reported on page #
SOURCE OF DATA	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	
	Participant description	
	Details of treatments received, if relevant	
	Study dates	
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	
	Was the same outcome definition (and method for measurement) used in all patients?	
	Type of outcome (e.g., single or combined endpoints)	
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	
	Time of outcome occurrence or summary of duration of follow-up	
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	
	Definition and method for measurement of candidate predictors	
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation)	
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	
SAMPLE SIZE	Number of participants and number of outcomes/events	
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)	
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	
	Number of participants with missing data for each predictor	
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques)	
	Modelling assumptions satisfied	
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	
	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)	
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals	
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	
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)	
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	
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)	
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	

	Comparison of the distribution of predictors (including missing data) for development and validation datasets	
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	
	Comparison with other studies, discussion of generalizability, strengths and limitations.	

Supplemental File 5. Prediction model Risk Of Bias Assessment Tool (PROBAST)

Link to full explanation and elaboration document
Citation: Moons KG, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019;170:W1–W33. doi: https://doi.org/10.7326/M18-1377

Domain 1: Participants
The overall aim for prediction models is to generate absolute risk predictions that are correct in new individuals. Certain data sources or designs are not suited to generate absolute probabilities. Problems may also arise if a study inappropriately includes or excludes participant groups from entering the study

	Signaling question	Yes/probably yes	No/probably no	No information
1	What study design was used and was it appropriate?	Yes: If a cohort design (including RCT or proper registry data) was used and you have confidence in data quality and participant enrollment is clearly described Probably yes: a nested case–control or case–cohort design (with proper adjustment of the baseline risk/hazard in the analysis) has been used or a cohort design was used but participant enrollment was data quality is unclear	No: If a non-nested case–control design has been used Probably no: a nested case-control study was used without proper adjustment of baseline risk/hazard	If the method of participant sampling is unclear.
2	Were all inclusion and exclusion criteria appropriate?	Yes: Inclusion and exclusion are clear and selection participants was appropriate, so participants correspond to unselected participants of interest (i.e. the target population). Probably yes: Inclusion and exclusion criteria are not entirely clear, but it seems like the population is representative of the target population	No: If participants are included who would already have been identified as having the outcome and so are no longer at risk of developing outcome, or if specific subgroups are excluded that may have altered the performance of the prediction model for the intended target population. Probably no: inclusion and exclusion criteria are unclear and it seems possible that there was bias in selection of participants that could lead to the model being applied to a population that is unrepresentative of the target population.	When there is no information on whether inappropriate inclusions or exclusions took place.

Low risk of bias	**High risk of bias**	**Unclear risk of bias**
If the answer to all signaling questions is "Yes" or "Probably yes," then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no," the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low.	If the answer to any of the signaling questions is "No" or "Probably no," there is a potential for bias, except if defined at low risk of bias above.	If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.

Domain 2: Predictors
Bias in model performance can occur when the definition and measurement of predictors is flawed. Predictors are the variables evaluated for their association with the outcome of interest. Bias can occur, for example, when predictors are not defined in a similar way for all participants or knowledge of the outcome influences

	Signaling question	**Yes/probably yes**	**No/probably no**	**No information**
1	Were predictors defined and assessed in a similar way for all participants?	Yes: It is clear that definitions of predictors and their assessment were similar for all participants. Probably yes: Some predictors were based off subjective judgement, but carried out by persons with the necessary skills to evaluate the predictor, or if data from multiple sources was used but predictor definitions were standardized between sources.	No: If different definitions were used for the same predictor or if predictors requiring subjective interpretation were assessed by differently experienced assessors Probably no: Data from multiple sources was used and its unclear whether definitions were standardized between sources or if subjective measurements were likely not carried out by persons with appropriate training.	If there is no information on how predictors were defined or assessed.
2	Were predictor assessments made without knowledge of data outcome?	Yes: If outcome information was stated as not used during predictor assessment or was clearly not (yet) available to those assessing predictors (i.e. prospective data collection).	If it is clear that outcome information was used when assessing predictors.	No information on whether predictors were assessed without knowledge of outcome information.

		Probably yes: If it is likely that outcome information was not used during predictor assessment, but not entirely clear (retrospective data collection/surveillance data)		
3	Are all predictors available at the time the model was intended to be used?	All included predictors would be available at the time the model is intended to be used for prediction	Predictors would not be available at the time the model is intended to be used for prediction.	No information on whether predictors would be available at the time the model is intended to be used for prediction.

Low risk of bias	**High risk of bias**	**Unclear risk of bias**
If the answer to all signaling questions is "Yes" or "Probably yes," then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no," the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low.	If the answer to any of the signaling questions is "No" or "Probably no," there is a potential for bias, except if defined at low risk of bias above.	If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.

Domain 3: Outcome

Bias in model performance can occur when methods used to determine outcomes incorrectly classify participants with or without the outcome. Bias in methods of outcome determination can result from use of suboptimal methods, tests, or criteria that lead to unacceptably high levels of errors in outcome determination, when methods are inconsistently applied across participants, or when knowledge of predictors influence outcome determination. Incorrect timing of outcome determination can also result in bias.

	Signaling question	**Yes/probably yes**	**No/probably no**	**No information**
1	Was the outcome determined appropriately?	If a method of outcome determination has been used which is considered optimal or acceptable by guidelines or previous publications on the topic Note: This is about level of measurement error within the method of determining the outcome (see concerns for applicability about whether the definition of the outcome method is appropriate).	If a clearly suboptimal method has been used that causes unacceptable error in determining outcome status in participants	No information on how outcome was determined
2	Was the outcome pre-specified or standard?	Yes: If the method of outcome determination is objective, or if a standard outcome definition is used, or if prespecified categories are used to group outcomes. (i.e. outcome assessment is based on previously published studies, published study protocol, or clinical guidelines) Probably yes: The outcome determination is not clearly based on guidelines or previous research, but outcome assessment is objective and would not inadvertently alter study results	No: If the outcome definition was not standard and not prespecified Probably no: a non-standard or non-prespecified outcome was used, and it is unclear whether the outcome definition could introduce bias. *Caution with composite outcomes that favor a better model by excluding typical outcome components or including atypical events	No information on whether the outcome definition was prespecified or standard
3	Were predictors excluded from outcome definition?	Yes: None of the predictors are included in the outcome definition (clearly stated) Probably yes: None of the predictors are included in the outcome definition (assumed)	If ≥1 of the predictors forms part of the outcome definition	No information on whether predictors are excluded from the outcome definition
4	Was the outcome defined and determined in a similar way for all participants?	Yes: If outcomes were defined and determined in a similar way for all participants (clearly stated) Probably yes: If outcomes were defined and determined in a similar way for all participants (assumed)	If outcomes were clearly defined and determined in a different way for some participants	No information on whether outcomes were defined or determined in a similar way for all participants
5	Was the outcome determined without predictor information	Yes: If predictor information was not known when determining the outcome status, or outcome status determination is clearly reported as determined without knowledge of predictor information. Probably yes: predictor information might have been available at time of outcome assessment, but outcome definition is objective and knowing information about predictors would not influence outcome	No: If it is clear that predictor information was used when determining the outcome status Probably no: it is likely predictor information was available at the time of outcome assessment, and outcome definition is subjective and knowledge of predictors could influence outcome determination.	No information on whether outcome was determined without knowledge of predictor information

		assessment (i.e death, treatment failure based on culture results, etc)		
6	Was the time interval between predictor assessment and outcome determination appropriate	If the time interval between predictor assessment and outcome determination was appropriate to enable the correct type and representative number of relevant outcomes to be recorded, or if no information on the time interval is required to allow a representative number of the relevant outcome occur or if predictor assessment and outcome determination were from information taken within an appropriate time interval.	If the time interval between predictor assessment and outcome determination is too short or too long to enable the correct type and representative number of relevant outcomes to be recorded.	If no information was provided on the time interval between predictor assessment and outcome determination.

Low risk of bias	**High risk of bias**	**Unclear risk of bias**
If the answer to all signaling questions is "Yes" or "Probably yes," then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no," the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low.	If the answer to any of the signaling questions is "No" or "Probably no," there is a potential for bias, except if defined at low risk of bias above.	If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.

Domain 4: Analysis

Statistical analysis is a critical part of prediction model development and validation. The use of inappropriate statistical analysis methods increases the potential for bias in reported model performance measures. Model development studies include many steps where flawed methods can distort results. We recommend reviewers seek statistical advice when completing

	Signaling question	**Yes/probably yes**	**No/probably no**	**No information**
1	Were there a reasonable number of participants with the outcome?	For model development studies, if the number of participants with the outcome relative to the number of candidate predictor parameters is ≥20 (EPV ≥20).* For model validation studies, if the number of participants with the outcome is ≥100.	For model development studies, if the number of participants with the outcome relative to the number of candidate predictor parameters is <10 (EPV <10).* For model validation studies, if the number of participants with the outcome is <100.	For model development studies, no information on the number of candidate predictor parameters or number of participants with the outcome, such that the EPV cannot be calculated. For model validation studies, no information on the number of participants with the outcome.
		* For EPVs between 10 and 20, the item should be rated as either probably yes or probably no, depending on the outcome frequency, overall model performance, and distribution of the predictors in the model. For more guidance, see references 145 to 147.		
2	Were continuous and categorical predictors handled appropriately?	Yes: If continuous predictors are kept as continuous or if continuous predictors are examined as linear or non-linear using restricted cubic splines or fractional polynomials. Probably yes: If continuous predictors are not converted into >2 categories when included in the model (i.e., dichotomized or categorized) using a prespecified method or in a way that avoids sparse data/would not intentionally improve statistical significance. For model validation studies, if continuous predictors are included using the same definitions or transformations, and categorical variables are categorized using the same cut points, ascompared with the development study.	No: For model development studies, if continuous predictors are converted into 2 categories when included in the model. Probably no: If categorical predictor group definitions do not use a prespecified method or continuous variables were split into >2 groups, but the decision of how to split variables is unclear. For model validation studies, if continuous predictors are included using different definitions or transformations, or categorical variables are categorized using different cut points, as compared with the development study.	No information on whether continuous predictors are examined for nonlinearity and no information on how categorical predictor groups are defined. For model validation studies, no information on whether the same definitions or transformations and the same cut points are used, as compared with the development study.
3	Were all enrolled participants included in the analysis?	If all participants enrolled in the study are included in the data analysis.	If some or a subgroup of participants are inappropriately excluded from the analysis (because they were missing data, unknown outcome, outliers)	No information on whether all enrolled participants are included in the analysis.
4	Were participants with missing data handled appropriately?	Yes: If there are no missing values of predictors or outcomes and the study explicitly reports that participants are not excluded on the basis of missing data, or if missing values are handled using multiple imputation.	No: If participants with missing data are omitted from the analysis, or if the method of handling missing data is clearly flawed, e.g., missing indicator method or inappropriate use of last value carried forward, or	If there is insufficient information to determine if the method of handling missing data is appropriate

		Probably yes: If a small percentage of persons with missing data were excluded and authors provide comparison of included vs. excluded participants or if sensitivity analysis with imputation methods are convincing that bias is low	if the study had no explicit mention of methods to handle missing data. Probably no: If authors provide comparison of included vs. excluded participants or if sensitivity analysis with imputation methods are reported, but the results are not convincing to rule out bias from excluding missing data	
5	Was selection of predictors based on univariable analysis avoided?	If the predictors are not selected on the basis of univariable analysis prior to multivariable modeling.	If the predictors are selected on the basis of univariable analysis prior to multivariable modeling.	If there is no information to indicate that univariable selection is avoided.
6	Were complexities in the data (censoring, competing risks, sampling of control participants) accounted for appropriately?	If any complexities in the data are accounted for appropriately, or if it is clear that any potential data complexities have been identified appropriately as unimportant.	If complexities in the data that could affect model performance are ignore. For example, case-control studies that do not estimate baseline risk or studies with censoring or competing risks that do not use survival analysis or other appropriate methods.	No information is provided on whether complexities in the data are present or accounted for appropriately if present.
7	Were relevant model performance measures evaluated appropriately?	Yes: If both calibration (via calibration plot) and discrimination (c-index) are evaluated appropriately (including relevant measures tailored for models predicting survival outcomes). Probably yes: if authors present a table of predicted probabilities with confidence intervals and corresponding outcome frequencies across subgroups	If both calibration and discrimination are not evaluated, or if only goodness-of-fit tests (Hosmer-Lemeshow test), are used to evaluate calibration or if for models predicting survival outcomes performance measures accounting for censoring are not used, or if classification measures (like sensitivity, specificity, or predictive values) were presented using predicted probability thresholds derived from the data set at hand, but calibration is not otherwise evaluated.	Either calibration or discrimination are not reported, or no information is provided as to whether appropriate performance measures for survival outcomes are used (e.g., references to relevant literature or specific mention of methods, such as using Kaplan–Meier estimates), or no information on thresholds for estimating classification measures is given.
8	Were model overfitting, underfitting, and optimism in model performance accounted for?	Yes: If internal validation techniques (bootstrapping and cross-validation) including all model development procedures, were used to account for any optimism in model fitting, and subsequent adjustment of the model performance estimates were applied. Probably yes: If internal validation was used and optimism was estimated as very low, and then optimism-corrected performance measures were not appropriately calculated (accounting for all model development procedures)	No: If no internal validation has been performed, or if internal validation consists only of a single random split-sample of participant data, Probably no: Internal validation with bootstrapping or cross-validation was conducted but did not include all model development procedures including any variable selection or were not used to correct model performance measures.	No information: No information is provided on whether internal validation techniques, including all model development procedures, have been applied.
9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	If the predictors and regression coefficients in the final model correspond to reported results from multivariable analysis.	If the predictors and regression coefficients in the final model do not correspond to reported results from multivariable analysis. (i.e. rounding of model coefficients to create a "risk score" are inappropriately determined).	If it is unclear whether the regression coefficients in the final model correspond to reported results from multivariable analysis.

Low risk of bias	**High risk of bias**	**Unclear risk of bias**
If the answer to all signaling questions is "Yes" or "Probably yes," then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no," the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low.	If the answer to any of the signaling questions is "No" or "Probably no," there is a potential for bias, except if defined at low risk of bias above.	If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.

Applicability				
	Domain	**Low concern**	**High concern**	**Unclear concern**

	Participants: do you have concern that the included participants or setting do not match the review question?	Included participants and clinical setting match the review question.	Included participants and clinical setting were different from the review question.	If relevant information about the participants and clinical setting are not reported.
	Predictors: does the definition, assessment, or timing of predictors match the review questions?	Definition, assessment, and timing of predictors match the review question.	Definition, assessment, or timing of predictors were different from the review question	If relevant information about the predictors is not reported.
	Outcome: does the definition, timing, or determination of outcome match the review question?	Outcome definition, timing, and method of determination defines the outcome as intended by the review question.	Choice of outcome definition, timing, and method of outcome determination defines another outcome as intended by the review question	If relevant information about the outcome, timing, and method of determination is not reported.

Supplemental File 6. Model outcome definitions

Study ID	Outcome category	Full outcome definition from the source paper
Hussain / 2019	Treatment completion	The target variable TreatmentComplete consists of 64.37% positive (treatment complete) and 35.62% negative (treatment incomplete)
Abdelbary / 2017 - Death	Death	All causes of death (TB or non-TB related) during the course of TB treatment
Abdelbary / 2017 - TB-DM / Death	Death	Death included all causes of death (TB and non-TB related) during the course of TB treatment
Aljohaney / 2018	Death	Not defined, but seems to be death during hospitalization.
Bastos / 2016	Death	Deaths that occurred during the first 6 months after diagnosis were classified as TB death
Gupta-Wright / 2019	Death	The outcome was mortality risk at 2 months after admission.
Horita / 2013	Death	'Discharged alive' was defined as being discharged alive and satisfying the discharge criteria, i.e., when the patient was receiving effective treatment, showed clinical improvement and negative conversion was confirrmed. Negative conversion was defined as three or more consecutive sputum samples obtained on different days being smear-negative for acid-fast bacilli or when appropriate sputum sample(s) were culture-negative. 'Died in hospital' was defined as death from any cause.
Koegelenberg / 2015	Death	Patients were categorised as either ICU/hospital survivors or non-survivors.
Nguyen (general pop) / 2018	Death	Documented treatment outcome of 'completed' or 'died'
Nguyen (TB-DM) / 2019	Death	TB treatment outcome of either 'completed' or 'died'
Nguyen (TB-HIV) / 2018	Death	Given the main purpose of our study is to predict the mortality during TB treatment in HIV-infected patients against the treatment completion, patients who had an outcome coding other than completed or died.
Pefura-Yone / 2017	Death	At treatment completion, patients are ranked into the following mutually exclusive categories 1) cured-patient with negative smear at the last month of treatment and at least one of the preceding months; 2) treatment completed-patient who has completed the treatment and for whom the smear results at the end of the last month are not available; 3) failure-patient with positive smear at the 5th month or later during treatment; 4) death-death from any cause during treatment; 5) defaulter-patient who's treatment has been interrupted for at least two consecutive months; 6) transfer-patient transferred to complete his treatment in another center and who's treatment outcome is unknown Cured and treatment completed are considered successful treatment
Podlekareva / 2013	Death	Death within 12 months of TB diagnosis
Valade / 2012	Death	Final outcomes of survival or death were recorded
Wang / 2019	Death	The outcome was estimated with all-cause mortality, with the mortality in 12 months as the primary outcome and the mortality in 3, 6, 9 months as other outcome
Wejse / 2008	Death	Mortality: ability to predict death
Zhang / 2019	Death	Primary treatment outcome was documented either survival or death when HIV/TB co-infected patients left hospital. Patients who survived when discharged received 12-month follow-up, and the date of last known alive was documented in electronical medical records base on records of last follow-up
Abdelbary / 2017 - Failure	Treatment failure	Treatment failure indicated smear-positive persistence at or after 5 months of treatment with first-line anti-TB medications.
Kalhori (logistic) / 2010	Treatment failure	The dependent variable was failing in treatment course completion.
Keane / 1997	Treatment failure	Failing to clear the sputum of acid-fast bacilli with standard treatment and having to start second line therapy
Luies / 2017	Treatment failure	From the original samples, all treatment failure cases were included.
Mburu / 2018 - Failure	Treatment failure	The secondary analyses only compared 'cures' versus 'failures' at similar time points as is the standard practice when examining chemotherapy efficacy
Thompson / 2017	Treatment failure	Patients' clinical outcomes were classified as 'cured' if they proved and maintained sputum culture negativity by month 6 after treatment initiation (M6), 'failed' if the M6 culture was still positive, and 'un-evaluable' if contamination caused uncertainty in outcome. We note that none of the treatment failures achieved culture negativity at any time point during treatment.
Abdelbary / 2017 - TB-DM / Default	Default, Abandon, or LTF (interruption >2 months)	Never defined
Belilovsky / 2010	Default, Abandon, or LTF (interruption >2 months)	We evaluated TI initiated by the patient (significant noncompliance with the doctor's prescribed course of treatment and serious violations of public order in hospitals) resulting in inpatient treatment cancellation.
Chang / 2004	Default, Abandon, or LTF	Default was defined as failure to collect drugs for 2 months or more after registration

	(interruption >2 months)	
Chee / 2000	Default, Abandon, or LTF (interruption >2 months)	Defaulter or cases were defined as patients on anti-tuberculosis treatment at the TBCU who failed to turn up for their scheduled appointments despite usual attempts to recall them by phone or mail, as described below, and from whom at least one home visit during the study was recorded
Cherkaoui / 2014	Default, Abandon, or LTF (interruption >2 months)	Treatment default was defined as an interruption in TB treatment for >=2 consecutive months.
Rodrigo / 2012	Default, Abandon, or LTF (interruption >2 months)	Interruption of treatment for any reason for more than 2 months, non-completion of treatment within 9 months when the patient is placed on a 6 month regimen. or drug intake of <80% the prescribed dose.
Kalhori (predicting) / 2009	Treatment success (cure + completion)	For each patient dependent variable was recorded whether or not the patient finished the treatment course and get cured.
Sauer / 2018	Unfavorable outcome (death + failure)	The primary outcome was treatment failure, which we defined as failure of therapy or death.
Baussano / 2008	Unfavorable outcome (death, failure, LTF, NE)	Treatment interruption or default, treatment failure, transferred out cases and those lost to follow-up were grouped as 'unsuccessful outcomes
Costa-Veiga / 2017	Unfavorable outcome (death, failure, LTF, NE)	In line with WHO criteria, SVIG-TB categorized a six possible and mutually exclusive categories for treatment outcomes, grouped in this study into a binary outcome: (i) Successful outcome-if PTB patients were treated before and declared cured, including both negative smear microscopy at the end of treatment at least one previous follow-up test and in case of not providing sputum samples, cure is declared if treatment completed and absent of disease clinical evidences (categories 1 and 2). (ii) Unsuccessful outcome-if treatment of PTB patients resulted in failure (i.e. remaining smear-positive after 5 months of treatment, cat. 3), default (i.e. patients who interrupted their treatment for two consecutive months or more after registration, cat. 4), death (cat. 5) or were transferred-out (cat. 6)
Killian / 2019	Unfavorable outcome (death, failure, LTF, NE)	We label 'Cured' and 'Treatment Complete' to be favorable outcomes and 'Died', 'Treatment failed', and 'Lost to follow-up' to be unfavorable outcomes
Madan / 2018	Unfavorable outcome (death, failure, LTF, NE)	Favourable treatment outcomes included cure and treatment completed. Unfavourable treatment outcomes included death, loss to follow-up, treatment failure, transfer out, or a switch to MDR TB treatment.
Mburu / 2018 - Unfavorable	Unfavorable outcome (death, failure, LTF, NE)	The primary analyses compared favorable versus unfavorable outcomes at end of treatment
Kalhori (fuzzy) / 2009	Other composite outcome	The values of outcomes might be any values from 1 to 5 which means different outcomes. Value 1 means patient completed the treatment course in frame of DOTS, 2 means the patient has been cured, 3 means patients has quitted the course, 4 means patients has failed and finally 5 is a sign of dead as outcome of TB treatment course

Supplemental File 7. Model presentation

Study ID	Final model
Abdelbary / 2017 - Death	2 + 2*(Age 41-65) + 5*(Age>=65) + 2*(Male gender) + 4*(MDR TB) + 3*(HIV) + 3*(Malnutrition) + 2*(Alcoholism) + 2*(Male*diabetes) + 3*(HIV*pulmonary TB) - 1*(diabetes) - 1*(pulmonary TB)
Abdelbary / 2017 - Failure	8*(No or low education) + 40*(MDR) + 10*(AFB smear +2) + 15*(AFB smear +3)
Abdelbary / 2017 - TB-DM / Death	2 + 3*(Male gender) + 3*(Malnutrition) - 1*(BCG vaccinated) - 1*(AFB smear positive)
Abdelbary / 2017 - TB-DM / Default	2 + 2*(Age<40) + 2*(Male gender) + 4*(HIV)
Aljohaney / 2018	Don't report final model, but show the beta coefficients. The coefficients are written as predictor (beta-coefficient): age ³ 65 (2.497), congestive heart failure (1.231), bilateral disease on chest x-ray (1.192)
Bastos / 2016	3*(Hypoxemic respiratory failure) + 2*(Age>=50) + 1*(Bilateral involvement) + 1*(At least one of: HIV, diabetes, liver failure/cirrhosis, congestive heart failure, chronic respiratory disease) + 1*(Hemoglobin<12)
Baussano / 2008	Nomogram with: residency status (residential vs. homeless), sex, geographic origin (non-EU vs. EU), case definition (other than definite vs. definite), treatment setting (inpatient and unknown vs. outpatient), age (continuous)
Belilovsky / 2010	-3.2 + 0.8*(male gender) + 0.7*(unemployment) + 0.4*(retreatment case) + 1.1*(alcohol abuse) + 0.6*(no data about alcohol) + 0.8*(severe TB form) - 0.3*(urban residence) + 0.4*(age 25-50) + 0.8*(pulmonary TB) + 0.5*(prison history)
Chang / 2004	Dont report final model. Just show odds ratios of predictors but don't report intercept term, which are written as predictor (OR) as follows: Current smokers (3.44), ex-smokers (2.48), history of default (10.74), no history of default (0.80),
Chee / 2000	The OR for each predictor is as follow in the format predictor (OR): Non-Chinese race (8.08), Living with family vs. living alone/with friends (0.08), Treatment duration (1.85). Treatment duration is categorical as 6 months, 9 months, and >9 months, but only one OR is presented.
Cherkaoui / 2014	2 points for yes to the following questions: Are you younger than 50 years of age? Do you feel work is interfering with your ability to take TB treatment? Are you taking a retreatment regimen for TB? Do you or doctor think you are having moderate or severe side effects from TB treatment Are you required to get your TB treatment daily? Have you told your friends that you have TB? (1 point for no) Are you a current smoker (1 point for yes) Did you TB symptoms go away within 2 months of starting TB treatment (1 point for yes) Do you know how long your TB treatment is supposed to last (1 point for no) Have you ever smoked cigarettes (-1 point for no)
Costa-Veiga / 2017	Nomogram with: HIV, previous treatment, age class (25-44, 15-24, 45-64, >64), IV drug use, pathologies (other disease comorbidity: yes/no)
Gupta-Wright / 2019	9*(Male sex) + 7*(patient aged 55+) + 6*(currently taking ART) + 7*(unable to walk unaided) + 7*(hemoglobin <80, severe anemia) + 6*(positive on urine TB-LAM)
Horita / 2013	1*Age (years) + 10*(oxygen requirement) - 20*(albumin) + 5*(semi-dependent, ADL) + 10*(total dependent, ADL)
Hussain / 2019	None
Kalhori (fuzzy) / 2009	Learned parameters by training set for each predictor written as predictor (learned parameter): Case type (0.467), treatment category (-0.079), risky sex (-0.945), prison (0.992), sex (0.400), recent TB infection (0.793), diabetes (2.445), low body weight (1.313), TB type (0.950), length (-0.235), previous imprisonment (2.398), age (0.237), area (0.8895), HIV (0.731)
Kalhori (logistic) / 2010	exp(-0.93 - 0.71*(gender) + 0.02*(age) - 0.02*(weight) + 0.5*(nationality) + 0.99*(prison) + 0.16*(case type))
Kalhori (predicting) / 2009	exp(-(1.58 - 0.12*(age) + 0.807*(gender) - 0.039*(nationality) - 0.263*(prison) + 0.15*(area) + 0.021*(weight))
Keane / 1997	Unclear. No constant term provided. Here are the predictor (OR): Mediastinal shift (2.1), average smear score (1.5), extensive lesions (3.6), any previous treatment (2.3), cavities (1.7), weight (0.98)
Killian / 2019	LEAP = Lstm rEal-time Adherence Predictor with 2 input layers, 1) LSTM with 64 hidden units and a dense layer with 48 units for the dense layer and 4 units for the penultimate layer
Koegelenberg / 2015	One point for each parameter: septic shock, HIV with CD4 < 200, creatinine > 140 (male) or >120 (female), P:F O2 ratio < 200, chest radiograph showing miliary pattern/parenchymal infiltrates, absence of TB treatment at admission
Luies / 2017	Written as predictor (OR): 3,5,-Dihydroxybenzoic acid (25.6), 3-(4-Hydroxy-3-methoxyphenyl) propionic acid (1.3)
Madan / 2018	Written as predictor (OR): New TB with 1+ smear grade (5.78), New TB with 2+ smear grade (2.69), New TB with 3+ smear grade (1.69), New TB without smear (1.67), New TB with smear positive, unknown grade (1.00), Previously treated, smear negative TB (1.35), previously treated with scanty smear (4.74), previously treated with 1+ smear grade (1.61), previously treated with 2+ smear grade (1.05), previously treated with 3+ smear grade (7.54), previously treated with no sputum smear (2.46), previously treated with unknown grade (30.37), pulmonary TB (1.83), pulmonary and extrapulmonary TB (5.86), HIV+ on ART with CD4 350-500 (8.09), HIV+ on ART with CD4 200-350 (6.14), HIV+ on ART with CD4 50-200 (16.35), HIV+ on ART with CD4 <50 (38.76), HIV+ not on ART with CD4 350-500 (53.44), HIV+ not on ART with CD4 200-350 (65.98), HIV+ not on ART with CD4 50-200 (6.94), HIV+ not on ART with CD4 <50 (49.20), HIV+ diagnosed after TB with CD4>500 (1.05), HIV+ diagnosed after TB with CD4 350-500 (2.49), HIV+ diagnosed after TB with CD4 200-350 (8.88), HIV+ diagnosed after TB with CD4 50-200 (6.79), HIV+ diagnosed after TB with CD4 <50 (13.99), Female 25-34 (9.41), Female 35-44 (1.75), Female >= 45 (4.49), Male 15-24 (10.63), Male 25-34 (2.74), Male 35-44 (2.9), Male >= 45 (3.96)
Mburu / 2018 - Failure	Present relative scores for each covariate included with scores of 100, 72.61, 69.19, 55.39, 49.87, 48.74, 48.18, 46.51, 39.69, and 37.69 for hba1c, regimen, age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine, respectively
Mburu / 2018 - Unfavorable	Present relative scores for each covariate included, not sure if this was how it should be used. Relative scores are 100, 79.38, 70.09, 63.93, 62.47, 62.63, 61.63, 55.62, 39.21, 34.48 for hba1c, regimen, creatinine, BMI, BUN, weight, age, random blood glucose, HIV positive result, male gender, respectively
Nguyen (general pop) / 2018	6*[Age 45-64] + 12*[Age>65] + 2*[US born] + 2*[Homeless] + 4*[Resident of LTCF] + 8*[Chronic kidney failure] + 10*[Meningeal TB] + 4*[Miliary TB] + 6*[TB-CXR] + 6*[HIV positive] + 6*[HIV unknown]
Nguyen (TB-DM) / 2019	16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[IDU] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliary TB] + 6*[AFB positive smear] + 24*[Positive HIV]
Nguyen (TB-HIV) / 2018	Prognostic score: 5*[Age >= 65] + 12*[Resident of LTCF] + 9*[Meningeal TB] + 6*[abnormal CXR] + 9*[diagnosis confirmed with positive culture or NAA] + 10*[culture not converted or unknown]

	Model: -6.994499 + 1.069024*[Age >= 65] + 2.541147*[Resident of LTCF] + 1.998852*[Meningeal TB] + 1.37995*[abnormal CXR] + 1.899108*[diagnosis confirmed with positive culture or NAA] + 2.186305*[culture not converted or unknown]
Pefura-Yone / 2017	1/(1 + exp(-1.3120 + 0.0474*[age] - 0.1866*[adjusted BMI] + 1.1637*[PTB-] + 0.5418*[ETB] + 1.3820*[HIV]
Podlekareva / 2013	1*[DST performed] + 2*[Initial treatment with RHZ] + 2*[cART started before or up to 1 month after TB diagnosis]
Rodrigo / 2012	1*[Immigrant] + 1*[Living alone] + 1*[Living in an institution] + 2*[Previous TB treatment] + 2*[Linguistic barriers] + 4*[IV drug use] + 1*[Unknown IV drug use]
Sauer / 2018	Negatively correlated: drug sensitivity (sensitive), employment status (employed), microscopy: 1 to 99 acid-resistant bacteria in 100 fields of view when stained by Ziehl-Nielsen, dissemination (diffuse pulmonary nodules detected)
Thompson / 2017	Heatmap of differentially expressed genes
Valade / 2012	Sum of three parameters: military tuberculosis (yes: +1, no: 0), required mechanical ventilation on ICU admission (yes: +1, no: 0), and required vasopressor infusion (yes: +1, no: 0).
Wang / 2019	Unknown
Wejse / 2008	1 point for each variable: cough, hemoptysis, dyspnea, chest pain, night sweating, anemia conjunctivae, tachycardia, positive funding at lung auscultation, temperature >37, BMI <18, BMI<16, MUAC<220, MUAC<200
Zhang / 2019	2*[Anemia (HGB < 90g/L)]+ 2*[Tuberculous meningitis] + 5*[Severe pneumonia] + 2*[Hypoalbuminemia] + 7* [Unexplained infections or space-occupying lesions] + 5* [Malignancies]

Supplemental File 8. Comparison of model performance and quality by population characteristics.

For each analysis below, results were stratified on the basis of whether the study population included, excluded, or did not report on two population characteristics of interest: MDR and younger age group (minimum age <18 vs. minimum age ≥18).

Note: The unit of measure for these analyses is the model (N=37) not the study (N=33), which explains differences in numbers between this and Table 4 of the main manuscript.

A) MDR

	Included (N=11)	**Excluded (N= 7)**	**Unknown (N=19)**
Prevalence of MDR, Median [IQR]	1% [1%-1%]	0% [0%-0%]	
C-statistic, Median [IQR]	0.77 [0.69-0.81]	0.77 [0.73-0.81]	0.75 [0.69-0.85]
Unknown	1	3	4
Outcome			
Death	7 (64%)	1 (14%)	8 (42%)
Treatment failure	2 (18%)	1 (14%)	3 (16%)
Default, LTF, or treatment interruption	1 (9.1%)	2 (29%)	3 (16%)
Composite outcome*	1 (9.1%)	3 (43%)	5 (26%)
Risk of Bias (Population)			
Low	6 (55%)	4 (57%)	11 (58%)
High	0 (0%)	2 (29%)	4 (21%)
Unclear	5 (45%)	1 (14%)	4 (21%)
Risk of Bias (Predictors)			
Low	1 (9.1%)	3 (43%)	9 (47%)
High	5 (45%)	0 (0%)	5 (26%)
Unclear	5 (45%)	4 (57%)	5 (26%)
Risk of Bias (Outcomes)			
Low	5 (45%)	4 (57%)	12 (63%)
High	0 (0%)	1 (14%)	3 (16%)
Unclear	6 (55%)	2 (29%)	4 (21%)
Risk of Bias (Analysis)			
Low	0 (0%)	0 (0%)	0 (0%)
High	11 (100%)	7 (100%)	19 (100%)
Unclear	0 (0%)	0 (0%)	0 (0%)
Top 5 predictors included^	Age (7), x-ray findings (5), extrapulmonary TB (4), HIV (4), other comorbidities (4), smear result (4)	Nationality (3), Age (2), HIV (2), living situation (2), previous TB (2), sex (2), treatment regimen (2)	Age (12), previous TB (9), BMI (8), extrapulmonary TB (6), sex (6)

Abbreviations: BMI=body mass index, LTF=losses to follow-up, MDR=multi-drug resistance, TB=tuberculosis
*Composite outcome includes unfavorable outcome (combination of death, failure, and default/LTF/treatment interruption) or treatment success (combination of cure and treatment completion)
^Witten as predictor (number of models included in). Top 5 unless there was a tie, in which case more predictors were listed.

Summary: Overall, the study population for 11 models included individuals with MDR, whereas 7 excluded patients with MDR, and the inclusion of MDR was unknown in 19 models. In models that included patients with MDR, the overall prevalence of MDR was low, with a median 1% prevalence. Model performance, as measured by the c-statistic, of studies that included and excluded patients with MDR was comparable and both were slightly higher than in studies where the prevalence of MDR was unknown. There were notable differences in outcome definition for the studies that included vs. excluded MDR patients, such as most studies that included patients with MDR examined death as the primary endpoint, whereas studies that excluded patients with MDR were more likely to use a composite outcome or evaluate default/LTF/treatment interruptions. Risk of bias assessment for the population and analysis domains were similar between all groups, but studies that included patients with MDR seemed to have higher amounts of bias in the predictors domain and more unclear risk of bias in the outcomes domain. For all groups, age was an important predictor of treatment outcome, but the other frequently included predictors varied between groups.

B) Age <18

	Included (N=10)	**Excluded (N= 11)**	**Unknown (N=16)**
Minimum age			
15	8 (80%)	0 (0%)	-
16	1 (10%)	0 (0%)	-
17	1 (10%)	0 (0%)	-
18	0 (0%)	10 (91%)	-
20	0 (0%)	1 (9.1%)	-
Age[#], Median [IQR]	34 [32-38]	43 [43-50]	44 [40-49]
Unknown	4	3	8
C-statistic, Median [IQR]	0.78 (0.65, 0.80)	0.70 (0.68, 0.84)	0.75 (0.74, 0.85)
Unknown	1	0	7
Outcome			
Death	5 (50%)	7 (64%)	4 (25%)
Treatment failure	2 (20%)	1 (9.1%)	3 (19%)
Default, LTF, or treatment interruption	0 (0%)	3 (27%)	3 (19%)
Composite outcome*	3 (30%)	0 (0%)	6 (38%)
Risk of Bias (Population)			
Low	10 (100%)	9 (82%)	2 (12%)
High	0 (0%)	0 (0%)	6 (38%)
Unclear	0 (0%)	2 (18%)	8 (50%)
Risk of Bias (Predictors)			
Low	6 (60%)	5 (45%)	2 (12%)
High	2 (20%)	5 (45%)	3 (19%)
Unclear	2 (20%)	1 (9.1%)	11 (69%)
Risk of Bias (Outcomes)			
Low	8 (80%)	9 (82%)	4 (25%)
High	0 (0%)	1 (9.1%)	3 (19%)
Unclear	2 (20%)	1 (9.1%)	9 (56%)
Risk of Bias (Analysis)			
Low	0 (0%)	0 (0%)	0 (0%)
High	10 (100%)	11 (100%)	16 (100%)
Unclear	0 (0%)	0 (0%)	0 (0%)
Top 5 predictors included^	Age (7), HIV (5), BMI (4), extrapulmonary TB (4), previous TB (4)	Age (7), sex (5), extrapulmonary TB (4), hemoglobin (3), HIV (3), MDR (3), other lab values (3), x-ray findings (3)	Age (7), nationality (5), previous TB (5), BMI (4), sex (4), treatment regimen (4), x-ray findings (4)

Abbreviations: BMI=body mass index, LTF=losses to follow-up
[#]Based on measure of central tendency reported in the study
*****Composite outcome includes unfavorable outcome (combination of death, failure, and default/LTF/treatment interruption) or treatment success (combination of cure and treatment completion)
^Witten as predictor (number of models included in). Top 5 unless there was a tie, in which case more predictors were listed.

Summary: In total, the study population of 10 models included individuals younger than 18, 11 had a minimum age of 18, and the minimum age of participants was not reported for 16 models. The age distribution of studies that included patients less than 18 was lower than that of studies with a minimum age of 18 or unreported minimum age. The c-statistic of studies that included younger patients (minimum age <18) was seemingly higher than studies with a minimum age of 18. Treatment outcome definitions varied between groups, such that none of the studies including younger patients examined default/LTF/treatment interruption as an outcome and none of the studies with age 18 as the minimum age used a composite outcome. Risk of bias for the population and predictors domain was somewhat lower for studies with a younger age population, and studies with unknown minimum age were more likely to be regarded as having unclear risk of bias. Across all groups, age was the most important predictor of outcome, but other important predictors varied between groups.