

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<u>http://bmjopen.bmj.com</u>).

If you have any questions on BMJ Open's open peer review process please email <u>info.bmjopen@bmj.com</u>

BMJ Open

BMJ Open

A systematic review of prediction models for tuberculosis treatment outcomes

Journal:	BMJ Open
Manuscript ID	bmjopen-2020-044687
Article Type:	Original research
Date Submitted by the Author:	10-Sep-2020
Complete List of Authors:	Peetluk, Lauren; Vanderbilt University School of Medicine, Epidemiology Ridolfi, Felipe; Instituto Nacional de Infectologia Evandro Chagas Rebeiro, Peter; Vanderbilt University School of Medicine, Epidemiology; Vanderbilt University School of Medicine, Division of Infectious Diseases Liu, Dandar; Vanderbilt University School of Medicine, Biostatistics Rolla, Valeria; Instituto Nacional de Infectologia Evandro Chagas Sterling, Timothy; Vanderbilt University School of Medicine, Division of Infectious Diseases
Keywords:	Tuberculosis < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, STATISTICS & RESEARCH METHODS





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

reliez oni

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A systematic review of prediction models for tuberculosis treatment outcomes

Lauren S. Peetluk, MPH,¹ Felipe M. Ridolfi, MD, MSc,² Peter F. Rebeiro, PhD, MHS,^{1,3} Dandan

Liu, PhD,⁴ Valeria C. Rolla, MD, PhD,² Timothy R. Sterling, MD³

¹Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine,

Nashville, Tennessee, USA

²Instituto Nacional de Infectologia Evandro Chagas (INI) – Fiocruz, Rio de Janeiro, Brazil

³Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of

Medicine, Nashville, TN, USA

⁴Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA

.edu Corresponding author: Lauren S. Peetluk, MPH A2209 Medical Center North 1161 21st Avenue South Nashville, TN 37203 E-mail: lauren.s.peetluk@vanderbilt.edu

Word count main text: 3547

ABSTRACT (word count: 261)

Objective: To systematically review and critically evaluate prediction models developed to predict tuberculosis (TB) treatment outcomes among persons with pulmonary tuberculosis. **Design:** Systematic review

Data sources: PubMed, Embase, Web of Science, and Google Scholar were searched for studies published January 1, 1995 - January 9, 2020.

Study selection and data extraction: Studies that developed a model to predict pulmonary TB treatment outcomes were included. Study screening, data extraction, and quality assessment were conducted independently by two reviewers. Study quality was evaluated using the Prediction model Risk Of Bias Assessment Tool (PROBAST).

Results: 14,739 articles were identified, 536 underwent full-text review, and 33 studies presenting 37 prediction models were included. Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=9, 25%). Most models (n=29, 78%) measured discrimination (median c-statistic=0.75; IQR: 0.68-0.84), and 17 (46%) reported calibration, often the Hosmer-Lemeshow test (n=13). Nineteen (51%) models were internally validated, and six (16%) were externally validated. Eighteen studies (54%) mentioned missing data, and of those half (n=9) used complete case analysis. The most common predictors included age, sex, extrapulmonary TB, body mass index (BMI), chest x-ray results, previous TB, and HIV. Risk of bias varied across studies, but all studies had high risk of bias in their analysis.

Conclusions: TB outcome prediction models are heterogeneous with disparate outcome definitions, predictors, and methodology. We do not recommend applying any in clinical settings without external validation, and encourage future researchers adhere to guidelines for developing and reporting of prediction models.

Registration: The study was pre-registered on OSF (<u>https://osf.io/rz3wp</u>).

totoeet etter ont

ARTICLE SUMMARY:

Strengths and limitations

- Prediction models for tuberculosis treatment outcomes have the potential to inform interventions or treatment management protocols to promote cure among tuberculosis patients at the greatest risk of unsuccessful treatment outcomes, but the methods and clinical utility of existing models had not been formally evaluated.
- This was the first systematic review of prediction models for tuberculosis treatment outcomes.
- The review used a comprehensive search strategy, conducted thorough bias assessment with the PROBAST tool, and offers recommendations for future model development and validation studies for predicting tuberculosis treatment outcomes.
- Evidence synthesis and quality assessment were limited by incomplete reporting in primary studies
- External validation studies or studies written in languages other than English, Spanish,
 Portuguese, or French were excluded.

BACKGROUND

Tuberculosis (TB) is one of the top ten causes of death worldwide and a leading cause of death from an infectious disease. In 2018, 10 million people developed TB and 1.45 million people died from it globally, despite widespread availability of curative treatment.(1) Global treatment success was 85% for all new and relapse TB patients in 2018. For HIV-associated TB, it was 75%. These proportions are lower than the End TB Strategy target of \geq 90% treatment success.(2)

Heeding early recognition that *Mycobacterium tuberculosis* develops resistance rapidly in response to single-drug therapy, TB has been treated with combination therapy for more than 50 years.(3) Aside from weight-based dosing, the World Health Organization (WHO) and other TB guidelines authorities recommend a standardized approach for treatment of almost all TB patients.(4–6) The current recommendation for treatment of drug-susceptible TB includes 2 months of isoniazid, rifampin, pyrazinamide, and ethambutol, followed by 4 months of isoniazid and rifampin. However, actual treatment regimens may vary due to differences in drug tolerability, and other individual-level factors that can affect TB treatment outcomes.

Due to the long duration of TB treatment, it would be beneficial for TB outcome studies to identify early treatment predictors of unsuccessful TB treatment outcomes to identify patients needing tailored treatment approaches, such as directly observed therapy (DOT) or extended treatment course. Research suggests that individual characteristics, such as HIV, age, undernutrition, diabetes, TB disease severity, extrapulmonary TB, history of TB, adherence, alcohol use, and adverse drug reactions, are associated with unsuccessful TB treatment outcomes, but results vary by setting and patient population.(7–10)

BMJ Open

Prediction models are defined as any combination or equation of two or more predictors, such as demographic factors, medical history, physical examination, and lab tests, used for estimating an individualized probability of a specific endpoint within a defined period of time.(11) The large number of prediction models for TB outcomes published in recent years highlights a common desire to identify TB patients at greatest risk of an unsuccessful treatment outcome in order to tailor treatment strategies and promote cure. However, to date, there has not been a formal synthesis or quality assessment of existing prediction models for TB treatment outcomes, which is essential to determine which models should inform clinical practice. This could also guide development of future models. Thus, we conducted a systematic review to identify, describe, compare, and synthesize clinical prediction models designed to predict TB treatment outcomes among persons with pulmonary TB.

METHODS AND ANALYSIS

All steps of the systematic review were carried out according to guidelines set by Cochrane Prognosis Methods Group (PMG) and PROGnosis RESearch Strategy (PROGRESS).(12–14) Reporting adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (**Supplemental File 1**). This study was pre-registered on OSF (<u>https://osf.io/rz3wp)</u>.

Study eligibility criteria

The review question was defined according to the PICOTS framework (**Supplemental File 2**). In brief, the goal was to identify prognostic models developed to predict TB treatment outcomes among pulmonary TB cases. The main outcome was unsuccessful TB treatment outcome, defined by the WHO as the combination of death, treatment failure, loss to follow-up, and/or not evaluated, as compared to successful TB treatment outcome, defined as the combination of cure or treatment completion (**Table 1**). Loss to follow-up was sometimes referred to as default or treatment abandonment.

Inclusion criteria were: 1) prognostic model studies with or without external validation(15); 2) study population included adult, drug-susceptible, pulmonary, TB cases; 3) written in English, Spanish, Portuguese, and French; 4) published between January 1, 1995 and January 9, 2020; 5) treatment outcome was one of the following: cure, treatment completion, death, treatment failure, loss to follow-up, or not evaluated.

Exclusion criteria were: 1) predictive value of more than one variable was evaluated but not combined in a prediction model; 2) study population was only multi-drug resistant (MDR) TB cases, only extrapulmonary TB cases, or only children (< 18 years-old); 3) outcome was evaluated during treatment such as: two-month smear/culture conversion, acquired resistance,

BMJ Open

adverse events, quality of life; 4) long-term outcomes, such as relapse, recurrence, or posttreatment mortality.

The decision to include only articles in English, Spanish, Portuguese, and French was based on study team capabilities. The dates reflect modern TB treatment practice; first-line TB treatment regimens were not available until the early 1990s.(16,17) Articles that included a combination of drug-susceptible and drug-resistant cases, or a combination of children and adults were included.

Search strategy and selection criteria

The following electronic databases were searched on January 9, 2020: PubMed, Embase, Web of Science, and the first 200 references from Google Scholar. This combination of databases achieved best overall recall for systematic reviews in a recent study.(18) Clinicaltrials.gov and retractiondatabase.org were also searched for unpublished research. Reference lists of retrieved articles were checked to identify eligible studies.

Search terms relating to the "prediction model" component of the search were adapted from a PubMed search strategy that captured prediction model studies with sensitivity of 98%.(19) That component was combined with terms relating to TB treatment outcomes. The search strategy, developed in PubMed, was adapted for all other databases with assistance from a reference librarian (**Supplemental File 3**).

Article selection was conducted in three stages. The first stage was de-duplication and title screening, carried out using *revtools* in RStudio (version 1.2).(20) Remaining articles were imported into Covidence, a web-based software platform that streamlines systematic reviews, where abstracts (Stage 2) and full text (Stage 3) were screened.(21) Stages 2 and 3 were carried out by two independent reviewers (LSP and FMR). Discordance was discussed between

reviewers, and if consensus was not reached, a third party arbitrated (one of TRS, VCR, PFR, DL). In stage 3, reasons for exclusion were documented according to PRISMA.

Data analysis

Data from selected studies were recorded using a database designed in REDCap (Vanderbilt University).(22,23) Data extraction was informed by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the Prediction Model Risk of Bias Assessment Tool (PROBAST).(15,24,25) CHARMS checklist and PROBAST are in **Supplemental Files 4 and 5**, respectively.

Quality assessment and applicability of included studies was assessed using PROBAST.(15,25) PROBAST was specifically designed to assess risk of bias of prediction model studies, which included identifying deficiencies in study design, conduct, or analysis that led to inaccurate estimates of predictive performance. PROBAST has 4 domains: participants, predictors, outcome, and analysis with 20 total signaling questions. Each question was answered on the scale: yes, probably yes, no, probably no, no information. Domains were scored as low, high, and unclear risk of bias. PROBAST also guides assessment of applicability of participants, predictors, and outcomes from each included study to the review question.

Results were summarized narratively and in tables and figures. Meta-analysis was not possible due to lack of external validation and use of disparate predictors, outcome definitions, and modeling methods. For studies that presented multiple models with the same set of predictors and outcomes, but different methods, the best-performing method was included in data synthesis. For studies presenting multiple models with different sets of predictors (i.e. baseline data vs. longitudinal data), the model developed using only baseline data was included. If studies developed multiple models for different outcomes or with different populations, all models were included.

Patient and public involvement

Neither patients nor the public were involved in the design, conduct, or reporting of the research, as it was not feasible or appropriate for this systematic review. The study protocol is publicly available at https://osf.io/rz3wp.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Study selection

The search identified 14,739 unique studies. After excluding irrelevant titles, 6,426 abstracts were screened, 536 articles underwent full-text review, and 33 model development studies presenting 37 prediction models were included (**Figure 1**).

Study characteristics

Of the 33 studies, most were retrospective cohorts (n=25, 76%), three (9%) were prospective cohort studies, two (6%) were case-control studies, and three (9%) were nested case-control studies. Data from nearly half of studies (n=16, 48%) were collected from surveillance systems; eleven (33%) studies used a data collection form developed specifically for their study and six studies (18%) extracted data from medical records. Median sample size was 803 (interquartile range (IQR): 291-4167). Full details on included studies are in **Table 2**.

Thirteen studies (41%) took place in Asia, eight (25%) in Africa, six (19%) in Europe, four (12%) in North America, and one (3%) included sites in Europe and Argentina. Fewer than half (n=14, 45%) of the studies took place in high-burden TB settings.¹ One study did not report study location. (**Tables 2 and 3**).

Reporting of population characteristics varied by study (**Table 4**). Among 18 studies that reported a measure of central tendency (mean or median) for age, the median of those measures of central tendency was 41 years (IQR: 37-49). Eighteen studies reported including persons living with HIV (PLWH); 5 of these included only TB/HIV patients. Twelve studies reported including persons with diabetes; one of which includes only TB/DM. Eight studies reported including participants with MDR, ten studies included only hospitalized patients, and in 14 studies, all participants were on directly observed therapy (DOT).

BMJ Open

Model characteristics

Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=8, 23%) (**Tables 2 and 5**). The complete outcome definition for all included studies is in **Supplemental File 6**.

Most models were developed using clinical/epidemiologic predictors (n=34, 92%), two (6%) used multiple biomarkers, and one (3%) used adherence data. The most common candidate predictors were age, sex, extrapulmonary TB, smear result, BMI, x-ray findings, and previous TB. The most common predictors retained in the final models were age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV (**Figure 2**).

Only three models (8%) used survival analysis; most models used logistic regression (n=29, 78%) and five (14%) used a machine learning approach. More than half of studies (n=19, 51%) considered variables for inclusion in the multivariable model based on unadjusted associations with the outcome. Model building methods varied widely between models (**Table 5**).

Only 19 (51%) models were internally validated, including ten (53%) split-sample validation, five (26%) bootstrap resampling, and four (21%) cross-validation. Six (16%) models were externally validated.

Many models (n=30, 81%) reported discrimination with c-statistic (concordance statistic) or area under the receiver operating characteristic (AUROC), which are equivalent and quantify the ability of the model to distinguish between patients who do and do not develop an outcome. Only 17 (46%) reported calibration, the agreement between observed and predicted outcomes. Most studies assessed calibration with Hosmer-Lemeshow tests (n=13, 77%); only two studies provided a calibration plot, the preferred reporting method for prediction model

studies,(15,26,27) and one reported the calibration slope (Table 2). Models were presented a variety of ways, the most common of which was a weighted risk score (n=16, 43%); details on model presentation are in Supplemental File 7.

Quality assessment

Grading of PROBAST signaling questions is summarized in **Figure 3**, and the summary risk of bias for the participants, predictors, outcome, and analysis domains and assessment of applicability are shown in **Figure 4**. More than half of the studies were at low risk of bias for the population and outcomes domains, but all studies were at high risk of bias in the analysis domain.

Common sources of population bias included use of non-nested case-control design(28,29), nested case-control design without proper estimation of baseline risk,(30,31) or inappropriate inclusion/exclusion criteria.(32,33) Sources of predictor bias included lack of standardized assessment of key predictors (i.e. HIV, diabetes, chest x-ray scoring)(9,28,30,33– 35) or timing of data collection/availability that would limit the intended use of the model.(9,28,36) Within the outcomes domain, sources of bias included subjective(34) or nonstandard(31,37) outcome measures and inconsistent outcome ascertainment.(28)

Bias in the analysis domain was widespread. More than half of the models included were likely overfit due to low events per variable (EPV) ratios (**Table 5**). Only 6 studies handled continuous and categorical variables appropriately (i.e., didn't dichotomize continuous variables, considered non-linearity of continuous variables).(30,38–42) Most studies used complete caseanalysis or did not mention missing data; no study used multiple imputation in their main analysis. One study with low amounts of missing data (<5%) conducted sensitivity analysis with multiple imputation.(43) A different study excluded only two people out of a total sample size of

BMJ Open

1007 with missing data, which would have little impact on model performance.(44) Fewer than half (n=14) of studies avoided univariable predictor selection, and only three studies used survival analysis, appropriately accounting for censoring.(35,44,45) Performance measures were appropriately reported (i.e. calibration assessed with plot and discrimination assessed with c-statistic/AUROC) in three studies.(40,43,46) Only two studies estimated optimism (degree to which data are overfit) or accounted for potential overfitting with penalization of model parameters.(34,40) Ten studies appropriately presented their model with model coefficients or nomograms, which prevents bias from rounding or transforming model coefficients to generate a risk score.(29,32,50–54,34,36,37,44,46–49)

About half of the models (n=19, 51%) were applicable to the review question in all domains. However, unclear reporting of target population or predictor and outcome definitions limited assessment of applicability for several studies.(37,48,49,55,56) Additionally, studies that included only hospitalized patients with specific laboratory parameters may not be routinely available in the clinical setting.(38,39,41)

DISCUSSION

In this comprehensive, systematic review of prediction models for pulmonary TB treatment outcomes, we identified 33 model development studies presenting 37 prediction models. These prediction models were developed for predicting death, treatment failure, default, or a composite unfavorable outcome during TB treatment. Most models reported good performance (c-statistic/AUROC>0.7), but all were evaluated to have high risk of bias due to poor reporting, exclusion of missing data, weak methodologic approaches, lack of calibration assessment, and limited validation. Predictor and outcome definitions varied by study and limited comparisons between models.

More than half of the models included in the review were developed in low burden TB settings, and none were developed specifically in South America. Prediction of TB treatment outcome is especially important in high burden TB settings, where resources may be limited, and risk assessment can guide resource allocation toward patients who need the most involved care protocols.

Common risk factors included in the models were consistent with well-established risk factors for poor TB treatment outcomes, including age, sex, HIV, extrapulmonary TB, baseline smear results, and previous TB treatment. Among studies that included PLWH, only three considered factors related to management/severity of HIV, such as receipt of antiretroviral therapy, CD4 cell count, or viral load, which likely impact TB treatment outcomes.(39,45,50) Laboratory values or metabolic biomarkers, such as hemoglobin, hemoglobin A1c or random blood glucose, may also be associated with treatment outcome and worth considering as candidate predictors. There is increasing evidence that diabetes impacts TB treatment outcomes, but caution is warranted about how to best define diabetes in the context of a prediction model to ensure consistency and reproducibility across studies.(57) Behavioral characteristics, such as

Page 17 of 66

BMJ Open

tobacco use, alcohol use, and drug use were rarely included in final prediction models and are difficult to collect objectively, suggesting their role in prediction models for TB treatment outcomes may be limited.

Additionally, several studies excluded participants with HIV, diabetes, extrapulmonary TB, or MDR TB, because these factors negatively influence treatment outcomes. However, careful consideration should be given to inclusion/exclusion criteria in prediction model studies. Information necessary to carry out inclusion/exclusions should be available at the of intended use of the model, which may not always hold for these aforementioned factors.(58) This point is especially questionable for MDR, given that conventional drug-susceptibility testing results are not available for several weeks after TB diagnosis; though more recent advances in rapid molecular methods such as GeneXpert or line-probe assays offer rapid screening for drug resistance.(59)

TB researchers should thoughtfully consider how to appropriately handle complexities of censoring and competing risks in TB outcomes research. Only three studies in this review used survival analysis, despite the long duration of TB treatment outcome assessment and relatively high rates of losses to follow-up across studies. Losses to follow-up were frequently excluded, which can lead to selection bias. Additionally, all studies that included death as the outcome considered all-cause mortality. Also, for studies that predict losses to follow-up/default, death (even due to TB) is a competing risk. Competing risk analyses are common in cardiovascular research, research in elderly populations, and there are specific recommendations for competing risk methods in prognostic research.(60,61)

Though all included studies were at high risk of bias in the analysis domain, we want to highlight two studies with some exemplary characteristics.(40,43) Pefura-Yone et al.(40) provide

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

clear explanations of study design, inclusion/exclusion criteria, and data collection procedures: TB diagnosis and treatment outcome definitions were standard. (62) Non-linearity of continuous variables was considered with restricted cubic splines, and no continuous variables were categorized or dichotomized; the final model includes four predictors that are easy to collect and routinely assessed in most TB control programs, especially those in high burden settings. The performance of the model was internally validated with bootstrap validation, and the discrimination (c-statistic=0.808) was corrected for optimism. Model calibration was presented graphically with calibration plots. The final model was presented as a nomogram with instructions for use, which facilitates use in external validation studies. Gupta-Wright and colleagues developed and externally validated a clinical risk score to predict mortality in highburden, low-resource settings.⁴³ They used clinical trial data with very low amounts of missing data for model development, and externally validated the clinical risk score with data collected independently from two other studies (a clinical trial and a prospective cohort). Given high amounts (42%) of missing data in the validation cohort, they conducted sensitivity analysis using multiple imputation for missing data; the c-statistic differed slightly between complete case and multiple-imputation analyses in the validation cohort (0.68 vs. 0.64). Candidate predictors were based on *a priori* clinical knowledge, previous literature, and required variables were objective, reproducible, and available in low-resource settings, consistent with recommended approaches. (25,58,63) Additionally, they reported model performance with the c-statistics and calibration plots for development and validation cohorts, and reported results according to TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidance. (26,27) Regardless, each of these models requires external validation prior to use in clinical practice.

Page 19 of 66

BMJ Open

There are several limitations of this study. First, data extraction was subject to reporting the primary study, which varied widely across studies. Most studies reported discrimination, and several reported sensitivity and specificity; TRIPOD recommends all studies report, at minimum, calibration with a calibration plot and discrimination with c-statistic.(27) Measures of sensitivity and specificity require dichotomization of risks, which then only pertain to a specific risk stratum, rather than quantifying the overall model performance.(14,63) We did not include external validation studies, which is an essential step for translation to clinical practice. However, several studies in the review did not include the full model equation, which inhibits their ability to be externally validated. Upon searching for studies that externally validated prediction models in this review, we found three studies (64–66) that evaluated the same model (TBscore).(35) Briefly, these studies evaluated the ability of TBscore to monitor treatment response in a new setting(64), refined the instrument (TBccoreII) using exploratory factor analysis(65), and then evaluated TBscoreII for use in patients with TB/HIV.(66) To our knowledge, no other studies included in the review were externally validated by other sources. Finally, we excluded 10 studies that were not available in English, Spanish, Portuguese, or French; all abstracts were available in English, and none reported model performance metrics, so they likely would have been excluded for different reasons regardless.

The findings of this review not only serve as a comprehensive overview of existing TB outcome prediction models but can act as a resource for future model development and validation of prediction models for TB treatment outcomes. We encourage researchers to focus future TB outcome prediction models on easily collected and readily available predictors that are widely generalizable. We highlight age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV as common predictors of TB treatment outcomes. Additionally, when

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

building a new prediction model, it is recommended to first prune the set of considered predictors based on expert opinion and previous literature, rather than univariable analysis or variable selection processes(25,58,63) Future model development or validation studies should adhere to the TRIPOD guidelines, which provide a 22-item checklist and aims to improve the reporting of prediction model development studies.(26,27) We also encourage researchers consider the PROBAST criteria when developing their model to limit sources of bias in design and conduct of prediction model studies.

Prediction models are an important tool in TB management, as they can lay the foundation for future intervention studies or clinical decision making by providing risk prediction that can aid in targeted treatment, resource allocation, or intensive case management at patients who are least likely to achieve cure and most likely to benefit from some form of intervention, especially in high-burden and low-resources areas. Though our findings suggest that none of the existing models are ready for clinical application without extensive external validation, we hope they direct future researchers to make use of guidelines for development and reporting of prediction models.

BMJ Open

FOOTNOTES

Ethics approval: Not required.

Transparency statement: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported. No important aspects of the study have been omitted, and any discrepancies from the study as planned were explained.

Contributorship: LSP conceptualized the research question, designed the protocol, and drafted the manuscript. LSP and FMR screened studies. FMR, PFR, DL, VCR and TRS provided feedback on the research design, original protocol, and revised successive drafts of the manuscript. All authors approved the final version of the manuscript.

Funding: This work was supported by the National Center for Advancing Translational Sciences [CTSA Award No. TL1TR000447 to L.S.P.]. Its contents are solely the responsibility of the authors and do not necessarily represent the official views the National Center for Advancing Translational Sciences or the National Institutes of Health.

Competing interests: None declared.

Data sharing: The study protocol is available online at https://osf.io/rz3wp. Most included studies are publicly available. Additional data and code are available upon request.

Exclusive license: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.

REFERENCES

- Global Tuberculosis Report 2019. In Geneva: World Health Organization; 2019. p. Licence: CC BY-NC-SA 3.0 IGO.
- 2. The END TB Strategy. Geneva: World Health Organization; 2015.
- Kerantzas CA, Jacobs WR. Origins of Combination Therapy for Tuberculosis: Lessons for Future Antimicrobial Development and Application. MBio. 2017 May 3;8(2):e01586-16.
- Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clin Infect Dis. 2016;63(7):e147–95.
- Guildelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update.
 In Geneva: World Health Organization; 2017. p. Licence: CC BY-NC-SA 3.0 IGO.
- WHO consolidated guidelines on drug-resistant tuberculosis treatment. In Geneva: World Health Organization; 2019. p. Licence: CC BY-NC-SA 3.0 IGO.
- Vasankari T, Holmström P, Ollgren J, Liippo K, Kokki M, Ruutu P. Risk factors for poor tuberculosis treatment outcome in Finland: A cohort study. BMC Public Health. 2007;7:1–9.
- Ramachandran G, Agibothu K. Factors Influencing Tuberculosis Treatment Outcome in Adult Patients Treated with Thrice-Weekly Regimens. 2017;61(5):1–12.
- Abdelbary BE, Garcia-Viveros M, Ramirez-Oropesa H, Rahbar MH, Restrepo BI.
 Predicting treatment failure, death and drug resistance using a computed risk score among newly diagnosed TB patients in Tamaulipas, Mexico. Epidemiol Infect. 2017;145(14):3020–34.
- 10. Torres NMC, Rodríguez JJQ, Andrade PSP, Arriaga MB, Netto EM. Factors predictive of

BMJ Open

2 3		the success of tuberculosis treatment: A systematic review with meta-analysis. PLoS One.
4 5		2010-14(12)-1 24
6 7		2019,14(12).1–24.
8 9	11.	Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al.
10 11		Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med.
12 13		2013 Feb 5;10(2):e1001381.
14 15	12.	Riley R, Ridley G, Williams K, Altman D, Hayden J, De V. Prognosis research: towards
16 17		evidence-based results and a Cochrane methods group. 2014;60:863-5.
18 19 20	13.	Moons KG, Hooft L, Williams K, Hayden JA, Damen JA, Riley RD. Implementing
21 22		systematic reviews of prognosis studies in Cochrane. Cochrane database Syst Rev.
23 24		2018;10:ED000129.
25 26 27	14.	Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to
28 29		systematic review and meta-analysis of prediction model performance. BMJ. 2017;356.
30 31	15.	Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al.
32 33 34		PROBAST: A tool to assess the risk of bias and applicability of prediction model studies.
35 36		Ann Intern Med [Internet]. 2019;170(1):51-8. Available from:
37 38		http://annals.org/article.aspx?doi=10.7326/M18-1376
39 40	16.	Iseman MD. Tuberculosis therapy: past, present and future. Eur Resp J [Internet]. 2002
41 42 43		[cited 2019 Feb 25];20(36):87s-94s. Available from:
44 45		https://erj.ersjournals.com/content/erj/20/36_suppl/87S.full.pdf
46 47	17.	Council STSMR. Clinical trial of six-month and four-month regimens of chemotherapy in
48 49 50		the treatment of pulmonary tuberculosis: the results up to 30 months. Tubercle. 1981;95-
50 51 52		102.
53 54	18.	Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for
55 56		
57		
59		22

literature searches in systematic reviews: A prospective exploratory study. Syst Rev. 2017;6(1):1-12. 19. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS One. 2012;7(2):3-8. 20. Westgate MJ. revtools: An R package to support article screening for evidence synthesis. Res Synth Methods. 2019;10(4):606–14. 21. Veritas Health Innovation, Melbourne A. Covidence systematic review sofware. Covidence. Melbourne, Australia: Veritas Health Innovation; 2016. 22. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform [Internet]. 2019;95(April):103208. Available from: https://doi.org/10.1016/j.jbi.2019.103208 23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform [Internet]. 2009;42(2):377– 81. Available from: http://dx.doi.org/10.1016/j.jbi.2008.08.010 24. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med. 2014;11(10). 25. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Ann Intern Med [Internet]. 2019 Jan 1;170(1):W1–33.

Page 25 of 66

BMJ Open

2			
3		Available from: http://annals.org/article.aspx?doi=10.7326/M18-1376	
5 6	26.	Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et	: al.
7 8		Transparent reporting of a multivariable prediction model for individual prognosis or	
9 10 11		diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1-	73.
12 13	27.	Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a	
14 15		multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The	
16 17		TRIPOD Statement. 2015;162(1).	
19 20	28.	Cherkaoui I, Sabouni R, Ghali I, Kizub D, Billioux AC, Bennani K, et al. Treatment	
21 22		default amongst patients with tuberculosis in urban Morocco: Predicting and explaining	g
23 24		default and post-default sputum smear and drug susceptibility results. PLoS One.	
25 26 27		2014;9(4).	
28 29	29.	Keane VP, De Klerk N, Krieng T, Hammond G, Musk W. Risk factors for the	
30 31		development of non-response to first-line treatment for tuberculosis in Southern Vietna	ım.
32 33 34		Int J Epidemiol [Internet]. 1997;26(5):1115–20. Available from: NA	
35 36	30.	Chang KC, Leung CC, Tam CM. Risk factors for defaulting from anti-tuberculosis	
37 38		treatment under directly observed treatment in Hong Kong. Int J Tuberc Lung Dis.	
39 40 41		2004;8(12):1492–8.	
41 42 43	31.	Chee CBE, Boudville IC, Chan SP, Zee YK, Wang YT. Patient and disease	
44 45		characteristics, and outcome of treatment defaulters from the Singapore TB control uni	t -
46 47		A one-year retrospective survey. Int J Tuberc Lung Dis [Internet]. 2000;4(6):496–503.	
48 49 50		Available from: NA	
51 52	32.	Luies L, Reenen M Van, Ronacher K, Walzl G, Loots DT. Predicting tuberculosis	
53 54		treatment outcome using metabolomics. Biomark Med [Internet]. 2017;11(12):1057-67	7.
55 56 57			
58 59			24

BMJ Open

Available from: http//www.futuremedicine.com/loi/bmm 33. Killian JA, Wilder B, Sharma A, Choudhary V, Dilkina B, Tambe M. Learning to Prescribe Interventions for Tuberculosis Patients Using Digital Adherence Data. Knowl Discov DATA Min [Internet]. 2019;NA(NA):2430-8. Available from: NA 34. Belilovsky EM, Borisov SE, Cook EF, Shaykevich S, Jakubowiak WM, Kourbatova E V. Treatment interruptions among patients with tuberculosis in Russian TB hospitals. Int J Infect Dis [Internet]. 2010;14(8):e698–703. Available from: http://dx.doi.org/10.1016/j.ijid.2010.03.001 35. Wejse C, Gustafson P, Nielsen J, Gomes VF, Aaby P, Andersen PL, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. Scand J Infect Dis. 2008;40(2):111–20. 36. Nguyen DT, Graviss EA. Development and validation of a risk score to predict mortality during TB treatment in patients with TB-diabetes comorbidity. BMC Infect Dis [Internet]. 2019;19(1):10. Available from: http://www.biomedcentral.com/bmcinfectdis/ 37. Kalhori SRN, Zeng X. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. Lect Notes Eng Comput Sci. 2009;2179(1):774–8. 38. Horita N, Miyazawa N, Yoshiyama T, Kojima R, Omori N, Kaneko T, et al. Poor performance status is a strong predictor for death in patients with smear-positive pulmonary TB admitted to two Japanese hospitals. Trans R Soc Trop Med Hyg. 2013;107(7):451-6. 39. Koegelenberg CFN, Balkema CA, Jooste Y, Taljaard JJ, Irusen EM. Validation of a severity-of-illness score in patients with tuberculosis requiring intensive care unit admission. South African Med J. 2015;105(5):389–92.

BMJ Open

40.	Pefura-Yone EW, Kuaban C, Assamba-Mpom SA, Moifo B, Kengne AP. Derivation,
	validation and comparative performance of a simplified chest X-ray score for assessing
	the severity and outcome of pulmonary tuberculosis. Clin Respir J. 2015;9(2):157-64.
41.	Valade S, Raskine L, Aout M, Malissin I, Brun P, Deye N, et al. Tuberculosis in the
	intensive care unit: A retrospective descriptive cohort study with determination of a
	predictive fatality score. Can J Infect Dis Med Microbiol. 2012;23(4):173-8.
42.	Wang Q, Han W, Niu J, Sun B, Dong W, Li G. Prognostic value of serum macrophage
	migration inhibitory factor levels in pulmonary tuberculosis. Respir Res [Internet].
	2019;20(1):50. Available from: http//respiratory-research.com/home/
43.	Gupta-Wright A, Corbett EL, Wilson D, Van Oosterhout JJ, Dheda K, Huerga H, et al.
	Risk score for predicting mortality including urine lipoarabinomannan detection in
	hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation
	and external validation cohort study. PLoS Med [Internet]. 2019;16(4):1-20. Available
	from: http://dx.plos.org/10.1371/journal.pmed.1002776
44.	Zhang Z, Xu L, Pang X, Zeng Y, Hao Y, Wang Y, et al. A Clinical scoring model to
	predict mortality in HIV/TB co-infected patients at end stage of AIDS in China: An
	observational cohort study. Biosci Trends. 2019;13(2):136-44.
45.	Podlekareva DN, Grint D, Post FA, Mocroft A, Panteleev AM, Miller RF, et al. Health
	care index score and risk of death following tuberculosis diagnosis in HIV-positive
	patients. Int J Tuberc Lung Dis [Internet]. 2013;17(2):198-206+i. Available from:
	http://www.ncbi.nlm.nih.gov/pubmed/23317955
46.	Baussano I, Pivetta E, Vizzini L, Abbona F, Bugiani M. Predicting tuberculosis treatment
	outcome in a low-incidence area. Int J Tuberc Lung Dis. 2008;12(12):1441–8.

47.	Costa-Veiga A, Briz T, Nunes C. Unsuccessful treatment in pulmonary tuberculosis:	
	Factors and a consequent predictive model. Eur J Public Health. 2018;28(2):252-8.	
48.	Niakan Kalhori SR, Nasehi M, Zeng XJ. A logistic regression model to predict high risl	k
	patients to fail in tuberculosis treatment course completion. IAENG Int J Appl Math.	
	2010;40(2):1–6.	
49.	Kalhori SRN, Zeng X-J. PREDICTING THE OUTCOME OF TUBERCULOSIS	
	TREATMENT COURSE IN FRAME OF DOTS - From Demographic Data to Logistic	
	Regression Model. In: Proceedings of the International Conference on Health Informati	cs
	[Internet]. SciTePress - Science and and Technology Publications; 2009. p. 129–34.	
	Available from:	
	http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0001431401290134	
50.	Madan C, Chopra KK, Satyanarayana S, Surie D, Chadha V, Sachdeva KS, et al.	
	Developing a model to predict unfavourable treatment outcomes in patients with	
	tuberculosis and human immunodeficiency virus co-infection in Delhi, India. Dholakia	
	YN, editor. PLoS One [Internet]. 2018 Oct 3;13(10):e0204982. Available from:	
	http://dx.doi.org/10.1371/journal.pone.0204982	
51.	Nguyen DT, Jenkins HE, Graviss EA. Prognostic score to predict mortality during TB	
	treatment in TB / HIV co-infected patients. PLoS One. 2018;13(4):1-12.	
52.	Nguyen DT, Graviss EA. Development and validation of a prognostic score to predict	
	tuberculosis mortality. J Infect. 2018;77(4):283-90.	
53.	Pefura-Yone EW, Balkissou AD, Poka-Mayap V, Fatime-Abaicho HK, Enono-Edende	
	PT, Kengne AP. Development and validation of a prognostic score during tuberculosis	
	treatment. BMC Infect Dis. 2017;17(1):1–9.	
	For poor roview only, http://bmicpon.hmi.com/site/about/suidelines.yhtml	27
	r or peer review only - http://binjopen.binj.com/site/about/guidennes.xittini	

BMJ Open

3	54.	Rodrigo T, Caylà JA, Casals M, García-García JM, Caminero JA, Ruiz-Manzano J, et al.
5 6		A predictive scoring instrument for tuberculosis lost to follow-up outcome. Respir Res.
7 8		2012;13:1–9.
9 10	55.	Hussain OA, Junejo KN. Predicting treatment outcome of drug-susceptible tuberculosis
11 12 13		patients using machine-learning models. Informatics Heal Soc Care. 2019;44(2):135-51.
13 14 15	56.	Sauer CM, Sasson D, Paik KE, McCague N, Celi LA, Fernández IS, et al. Feature
16 17		selection and prediction of treatment failure in tuberculosis. PLoS One. 2018:13(11):1–14
18 19	<i></i>	
20 21	57.	Huangfu P, Ugarte-Gil C, Golub J, Pearson F, Critchley J. The effects of diabetes on
22		tuberculosis treatment outcomes: an updated systematic review and meta-analysis. Int J
23 24 25		Tuberc Lung Dis [Internet]. 2019;23(7):783–96. Available from: NA
25 26 27	58.	Steyerberg EW. Clinical Prediction Models [Internet]. Springer. New York, NY: Springer
28 29		New York; 2009. (Statistics for Biology and Health). Available from:
30 31		http://link.springer.com/10.1007/978-0-387-77244-8
32 33	59.	Sharma SK, Dheda K. What is new in the WHO consolidated guidelines on drug-resistant
34 35		tuberculosis treatment? Vol. 149. The Indian journal of medical research. 2019. p. 309–
36 37		12
38 39		12.
40 41	60.	Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with
42 43		competing risks methods and application to coronary risk prediction. Epidemiology.
44 45		2009;20(4):555–61.
46 47	61.	Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence
48 49 50		of Competing Risks. Circulation. 2016;133(6):601–9.
51 52	62.	National Tuberculosis Control Program. Manual for health personnel. Yaounde; 2012.
53 54 55	63.	Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research:
56 57		
58 59		Ear poor raviou antu, http://bmianan.hmi.com/cita/ahaut/guidalings.yhtml
60		TOT PEETTEVIEW ONLY - HTTP://DHTJOPEN.DHTJ.COM/SITE/dDOUT/QUIGEMIES.XHTM

Developing a prognostic model. BMJ. 2009;338(7707):1373-7.

- 64. Janols H, Abate E, Idh J, Senbeto M, Britton S, Alemu S, et al. Early treatment response evaluated by a clinical scoring system correlates with the prognosis of pulmonary tuberculosis patients in Ethiopia: A prospective follow-up study. Scand J Infect Dis. 2012;44(11):828–34.
- 65. Rudolf F, Lemvik G, Abate E, Verkuilen J, Schön T, Gomes VF, et al. TBscore II: Refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. Scand J Infect Dis. 2013 Nov;45(11):825–36.
- 66. Wejse C, Patsche CB, Kühle A, Bamba FJV V, Mendes MS, Lemvik G, et al. Impact of HIV-1, HIV-2, and HIV-1+2 dual infection on the outcome of tuberculosis. Int J Infect Dis. 2015 Mar;32:128–34.
- 67. Definitions and reporting framework for tuberculosis 2013 revision. In Geneva: World Health Organization; 2013.
- Aljohaney AA. Mortality of patients hospitalized for active tuberculosis in King Abdulaziz University Hospital, Jeddah, Saudi Arabia. Saudi Med J. 2018;39(3):267–72.
- 69. Bastos HN, Osório NS, Castro AG, Ramos A, Carvalho T, Meira L, et al. A prediction rule to stratify mortality risk of patients with pulmonary tuberculosis. PLoS One. 2016;11(9):1–14.
- 70. Gupta-Wright A, Corbett EL, Wilson D, Van Oosterhout JJ, Dheda K, Huerga H, et al. Risk score for predicting mortality including urine lipoarabinomannan detection in hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation and external validation cohort study. PLoS Med [Internet]. 2019;16(4):1–20. Available from: http://dx.doi.org/10.1371/journal.pmed.1002776

BMJ Open

/		
3	71.	Horita N, Miyazawa N, Yoshiyama T, Sato T, Yamamoto M, Tomaru K, et al.
5		Development and validation of a tuberculosis prognostic score for smear-positive in-
7 8		patients in Japan. Int J Tuberc Lung Dis. 2013;17(1):54-60.
9 10 11	72.	Podlekareva DN, Grint D, Post FA, Mocroft A, Panteleev AM, Miller RF, et al. Health
12 13		care index score and risk of death following tuberculosis diagnosis in HIV-positive
14 15		patients. Int J Tuberc Lung Dis [Internet]. 2013 Feb 1;17(2):198–206. Available from:
16 17 19		http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L36810
19 20		6792%5Cnhttp://docstore.ingenta.com/cgi-
21 22		bin/ds_deliver/1/u/d/ISIS/72353575.1/iuatld/ijtld/2013/00000127/00000002/art00122/1F2
23 24		1E2D7FD6A26A51357873269D96871D3E6FFE5E3.pdf?link=ht
25 26 27	73.	Wang Q, Han W, Niu J, Sun B, Dong W, Li G. Prognostic value of serum macrophage
28 29		migration inhibitory factor levels in pulmonary tuberculosis. Respir Res [Internet].
30 31		2019;20(1):50. Available from: http//respiratory-research.com/home/
32 33 34	74.	Wejse C, Gustafson P, Nielsen J, Gomes VF, Aaby P, Andersen PL, et al. TBscore: Signs
35 36		and symptoms from tuberculosis patients in a low-resource setting have predictive value
37 38		and may be used to assess clinical course. Scand J Infect Dis. 2008;40(2):111-20.
39 40 41	75.	Mburu JW, Kingwara L, Ester M, Andrew N. Use of classification and regression tree
41 42 43		(CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor
44 45		tuberculosis treatment outcomes and associated risk factors. J Clin Tuberc Other
46 47		Mycobact Dis. 2018;11(January):10-6.
48 49 50	76.	Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, et al. Host blood RNA
51 52		signatures predict the outcome of tuberculosis treatment. Tuberculosis. 2017;107:48-58.
53 54	77.	Chee CBE, Boudville IC, Chan SP, Zee YK, Wang YT. Patient and disease
55 56 57		
58		

characteristics, and outcome of treatment defaulters from the Singapore TB control unit -A one-year retrospective survey. Int J Tuberc Lung Dis. 2000;4(6):496–503.

- Rodrigo T, Caylà JA, Casals M, García-García JM, Caminero JA, Ruiz-Manzano J, et al. A predictive scoring instrument for tuberculosis lost to follow-up outcome. Respir Res. 2012;13:1–9.
- 79. Kalhori SRN, Zeng X-J. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. In: Lecture Notes in Engineering and Computer Science [Internet]. NA; 2009. p. 774–8. Available from: NA

Page 33 of 66

BMJ Open

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
10	
12	
10	
20	
20	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
4Z 12	
45 11	
44	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	

60

Table 1. World Health Organization definition of treatment outcomes for TB patients(67)

Outcome	Definition
	Completion of treatment without evidence of failure, but without
Treatment	documentation of a negative sputum smear or culture in the last month of
completion	treatment and/or on at least one previous occasion, either because tests
	were not done or because results are unavailable
Cure	Bacteriologic confirmation of a negative smear or culture at the end of TB
	treatment and on at least one previous occasion
Treatment success	Composite of cured and treatment completed
Treatment failure	Sputum smear or culture is positive at month 5 or later during treatment
Death	TB patient who dies for any reason before starting or during the course of
	treatment
Loss to follow-up	TB patient who did not start treatment or whose treatment was interrupted
Loss to ronow up	for 2 consecutive months or more
Not evaluated	TB patient for whom no treatment outcome was assigned, which includes
(transfer out)	cases who "transferred out" to another treatment unit as well as cases for
(numbrer out)	whom the treatment outcome is unknown to the reporting unit

Figure 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow chart of inclusion process



to beet eview only
Table 2. Study characteristics

2										
2 3 4 First author, year 5	Population	Study years	Study design	Location	Validation	No. with outcome / sample size (%)	Predictors in final model	Performance measures	Model presentation	Risk of bias (population, predictor, outcome, analysis)
6 Death										
7 8 Abdelbary(9) / 2017	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 261/4216 (6%) Validation: 260/4215 (6%)	Age (<41, 41-65, ≥65), sex, MDR, HIV, malnutrition, alcoholism, diabetes, pulmonary TB	c-statistic = 0.70 Sensitivity = 60% Specificity = 71%	Risk score	Low, High, Low, High
10 Abdelbary(9) / 11 2017 (TB-DM) 12	TB-DM cases	2006 - 2013	Retrospective cohort	Mexico	None	88/2121 (4%)	Sex, malnutrition, BCG vaccinated, AFB smear (positive vs. negative)	c-statistic = 0.68	Risk score	Unclear, High, Low, High
13 _{Aljohaney(68)} / 14 2018	Hospitalized TB patients	Dec 2011 – Dec 2016	Retrospective cohort	Saudi Arabia	None	41/291 (14%)	<u>Clinical model:</u> Age, congestive heart failure <u>Clinical + lab model:</u> * Age > 65, congestive heart failure, bilateral disease on chest xray	<u>Clinical model:</u> Accuracy = 86% <u>Clinical & lab model:*</u> Accuracy = 90%	Odds ratios	Unclear, Unclear, Unclear, High
15 16 17 ^{Bastos(69) / 2016} 18	Inpatient and outpatient TB cases on DOT	2007 - 2013	Retrospective cohort	Portugal	External (setting)	Development: 121/681 (18%) Validation: 24/103 (23%)	Hypoxemic respiratory failure, age (≥50 vs. <50), bilateral involvement, comorbidities (at least one of HIV, diabetes, liver at least one of: HIV, diabetes, liver failure/cirrhosis, congestive heart failure, chronic respiratory disease), hemoglobin (<12 vs. ≥12)	AUROC = 0.84 (95% CI: 0.76-0.93) Sensitivity = 41.8% Specificity = 92.1%	Risk score	Low, Unclear, Low, High
19 Gupta-Wright(70) / 20 2019	Hospitalized TB- HIV patients	Oct 2015 – Sept 2017	Retrospective cohort	Malawi and South Africa	External (setting)	Development: 94/315 (30%) Validation: 147/644 (23%)	Sex, age 55+, currently taking ART, ability to walk unaided, severe anemia, positive TB-LAM	c-statistic = 0.68 (95% CI: 0.61-0.74) HL test: p=0.13 Calibration plot	Risk score	Low, Low, Low, High
22 Horita(71) / 2013 23	Hospitalized TB patients	Jan 2008 – Jul 2011	Retrospective cohort	Japan	External (setting)	Development: 36/179 (20%) Validation: 48/244 (20%)	Age, oxygen requirement, albumin, activities of daily living	AUROC = 0.893 Sensitivity = 0.92 Specificity = 0.73	Risk score	Low, Low, Low, High
24 25 _{Koegelenberg} (39) / 26 ²⁰¹⁵ 27	Hospitalized TB patients	Jan 2012 – May 2013	Retrospective cohort	South Africa	None	38/83 (46%)	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	Mean score in survivors: 2.27 (SD=1.47) Mean score in non- survivors: 3.58 (SD=1.08)	Risk score	Low, Low, Low, High
28 Nguyen(52) 29 (general pop) / 2018	TB cases	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (split-sample)	Development: 253/3378 (7%) Validation: 270/3377 (8%)	Age group (15-44, 44-64, >64), US born, homeless, resident of long term care facility, chronic kidney failure, meningeal TB, miliary TB, HIV positive, HIV unknown	AUROC = 0.80 (95% CI: 0.77-0.82) HL test:X ² =6.3, p=0.613	Risk score	Low, Unclear, Unclear, High
BU B1Nguyen(36) (TB- DM) / 2019 B2	TB-DM patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	112/1227 (9%)	Age ≥65, US-born, homeless, IDU, chronic kidney failure, TB meningitis, Miliary TB, AFB positive smear, HIV positive	AUROC = 0.82 (95% CI: 0.78-0.87) HL test: X ² =4.54, p=0.81 Brier score=0.07	Risk score	Unclear, Unclear, Unclear, High
B3 B4 ^{Nguyen(51) (TB- HIV) / 2018 B5}	TB-HIV patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	57/450 (13%)	Age ≥ 45, resident of LTCF, meningeal TB, abnormal CXR, diagnosis confirmed by positive culture of NAA, culture not converted or unknown	AUROC = 0.79 (95% CI 0.70-0.87) HL test: X ² =4.25, p=0.51 Brier score: 0.09	Risk score	Low, High, Unclear, High
36 37 ^{Pefura-Yone(53) /} 2017 38	TB patients	Jan 2012 – Dec 2013	Retrospective cohort	Cameroon	Internal (bootstrap)	213/2250 (9%)	Age, adjusted BMI, clinical form (PTB+, PTB-, EPTB), HIV	C-statistic: 0.808 HL test: X ² =6.44, p=0.60 Sensitivity = 80.7% Specificity = 68.2% Calibration plot	Model coefficients	Low, Low, Low, High
40 ^{Podlekareva(72) /} 2013 41	TB/HIV patients	Jan 2004 – Dec 2006	Retrospective cohort	52 cities in Europe and Argentina	None	995†	DST performed, treatment with RHZ, and cART at/near TB diagnosis	Crude RH = 0.62 (95% CI: 0.64-0.84)	Risk score	Low, Unclear, Low, High
42 43 44	Hospitalized TB patients	Mar 2000 – Jul 2009	Retrospective cohort	France	Internal (bootstrap)	20/53 (38%)	Miliary TB, catecholamine infusion, mechanical ventilation on admission	AUROC = 0.92 (95% CI: 0.85-0.98) Brier score = 0.13	Risk score	Unclear, Low,
44										54

Page	36	of	66
------	----	----	----

1 2								Optimism = 0.03 Accuracy = 85% Sensitivity - 75% Specificity = 91%		Low, High
3 4 Wang(73) / 2019 5	HIV-negative, culture-confirmed, pulmonary TB cases	Jan 2014 – Dec 2016	Prospective cohort	China	External (setting)	Development: 36/287 (13%) Validation: 15/104 (14%)	Age, cavitary lesion, pleural effusion, drug resistance, disseminated, albumin, c-reactive protein, white blood cell count, IL-6, MIF	AUROC = 0.85 ± 0.028	Odds ratios	Low, Low, Low, High
6 Wejse(74) / 2008	Pulmonary TB patients on DOT	1996 - 2001	Retrospective cohort	Guinea Bissau	None	100/698 (14%)	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	AUROC = 0.65 (95% CI: 0.6-0.7) Sensitivity = 0.45 Specificity = 0.75	Risk score	Low, High, Low, High
5 9 Zhang(44) / 2019 10	TB/HIV patients at end stage of AIDS	Aug 2009 – Jan 2018	Retrospective cohort	China	Internal (split-sample)	Development: 157/807 (19%) Validation: 40/200 (20%)	Anemia, TB meningitis, severe pneumonia, hypoalbuminemia, unexplained infection or space-occupying lesions, malignancy	AUROC = 0.867 (95% CI: 0.832-0.902) Sensitivity = 79.6% Specificity = 82.9%	Risk score	Low, Low, Low, High
11 _{Treatment failure}										5
12 13 Abdelbary(9) / 14 ²⁰¹⁷	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 2109† Validation: 6322†	Education (no or low vs. higher than primary school), MDR, AFB smear (>+2, +1, negative)	c-statistic = 0.65 Sensitivity = 52% Specificity = 66%	Risk score	Low, High, Low, High
15 16 Kalhori(48) (logistic) / 2010 17	TB cases at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 828/4836 (17%) Validation: 2418†	Gender, age, weight nationality, prison, case type	AUROC = 0.70 Accuracy = 81.64% HL test: X ² =11.935, df=8, p=0.154	Model coefficients	Unclear, Unclear, Unclear, High
18 19 _{Keane(29) / 1997} 20 21	Smear-positive TB patients on standard first-line regimen with DOT	1990 - 1995	Non-nested case control	Vietnam	None	130/803 (16%)	<u>3 month model:</u> Extensive lesions, mediastinal shift, average smear score 3rd month, weight, progressive x-ray, any previous treatment <u>Baseline model:</u> Mediastinal shift, average smear score, extensive lesions, any previous treatment, cavities, weight	$\frac{3 \text{ month:}}{\text{Sensitivity} = 80\%}$ $\frac{3}{\text{Specificity} = 80\%}$ $\frac{3}{\text{Baseline:}}$ $\frac{3}{\text{Sensitivity} = 70\%}$ $\frac{3}{\text{Specificity} = 80\%}$	Model coefficients	High, Unclear, Unclear, High
22 23 ^{Luies(32) / 2017} 24	Smear-positive pulmonary TB cases on DOT	May 1999 – Jul 2002	Nested case- control	South Africa	Internal (cross-validation)	10/31 (32%)	3,5,-Dihydroxybenzoic acid, (3-(4-Hydroxy-3-methoxyphenyl) propionic acid	AUROC = 0.89 (95% CI: 0.7-1.00)	Model coefficients	High, Unclear, Unclear, High
25 Mburu(75) / 2018 26	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	13/321 (4%)	HbA1c, regimen (retreatment), age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine	AUROC = 0.56 ± 0.07	Relative score	Low, Low, Low, High
27 <u>Default</u>										
28 29 Thompson(76) / 30 ²⁰¹⁷	HIV uninfected adults with newly diagnosed pulmonary TB	Apr 2010 – Apr 2013	Retrospective cohort	South Africa	Internal (cross-validation) and external (setting)	6/99 (6%)	18 splice junctions and 13 genes	AUROC (internal) = 0.87 AUROC (external) = 0.63	Heatmap of differentially expressed genes	Low, Low, Low, High
31 32 Abdelbary(9) / 2017 (TB-DM) 33	TB cases	2006 - 2013	Retrospective cohort	Mexico	None	93/2121 (4%)	Age (<40 vs. ≥40), sex, HIV	c-statistic= 0.62	Risk score	Unclear, High, Unclear, High
34 35 _{Belilovsky(34) /} 36 2010 37	Hospitalized TB patients	1993 - 2002	Retrospective cohort	Russia	External (geographical)	Development: 1326/3904 (34%) Validation: 4662/12803 (36%)	Sex, unemployment, retreatment case, alcohol abuse (yes, no, no data), severe TB form, residence (urban vs. rural), age (25-50 vs. other), pulmonary TB (vs extrapulmonary), prison history	Belgrood: AUROC = 0.75 Orel: AUROC = 0.75 Pskov: AUROC = 0.78 Yaroslavi: AUROC = 0.75 Calibration table	Model coefficients	Unclear, High, High, High
38 39 40 _{Chang(30) / 2004} 41 42 43	All tuberculosis patients	Jan 1999 – Mar 1999	Nested case- control	China	None	102/408 (25%)	Baseline:* Ever smoker (current, former, never), retreatment (history of default, no history of default, not) Longitudinal: Smoking status (current, former, never), retreatment (with history of default, without history of default, never), unsatisfactory adherence in first two months (good, poor, fair, unknown), subsequent hospitalization, treatment side effects in last month of treatment	$\frac{\text{Baseline:}^{*}}{\text{AUROC} = 0.70 (95\% \text{ CI:} 0.63-0.76)}$ HL test: X ² = 1.448, df=5, p=0.919 <u>Longitudinal:</u> AUROC = 0.85 (95% CI: 0.80-0.90)	Odds ratios	High, High, Low, High
44										35

BMJ Open

1								HL test: $X^2 = 5.887$, df=6, p=0.436		
2 Chee(77) / 2000	TB cases	1996	Nested case- control	Singapore	None	38/71 (54%)	Chinese race, extent of family support, treatment duration	Accuracy = 74.6%	Model coefficients	High, Unclear, High, High
Cherkaoui(28) / 2014	TB patients with definite or probable pulmonary or extrapulmonary TB	Jun 2010 – Oct 2011	Non-nested case-control	Morocco	None	91/277 (33%)	Age <50, work interfering with ability to take TB treatment, retreatment regimen, daily DOT, moderate or severe side effects, told friends about TB, current smoker, never smoker, symptom resolution in <2 months, knowledge of TB treatment duration	AUROC = 0.85 (95% CI: 0.80-0.90) Sensitivity = 82.4% Specificity = 87.6% HL test: X ² =0.77, p- value=1.00	Survey tool	High, High, High, High, High
3 9 _{Rodrigo(78) / 2012} 10	New TB cases	Jan 2006 – Dec 2009	Prospective cohort	Spain	Internal (split-sample)	Development: 92/1490 (6%) Validation: 103/1589 (6%)	Immigrant, living alone, living in an institution, previous TB treatment, linguistic barriers (poor understanding), IV drug use, unknown IV drug use	AUROC = 0.67 (95% CI: 0.65-0.70) Sensitivity = 65.05% Specificity = 67.36%	Risk score	Low, Low, Low, High
11 <u>Unfavorable</u> outcome										
13 Kalhori(49) 14 ^{predicting) / 2009[†]}	TB patients at DOT registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 6920† Validation: 2966†	Age, gender, nationality, prison, area, weight	Classification rate = 89.8% R2 = 0.45	Model coefficients	Unclear, Unclear, Unclear, High
5 6 17 18 19 20 21 22 23 24 25 26 27 28 29Sauer(56) / 2018 [†] 30 31 32 33 34 35 36 37 38 39 40 41 42 43	TB cases	Data available through March 2018	Retrospective cohort	Azerbaijan, Belarus, Georgia, Moldova, Romania	Internal (split-sample)	Development: 103/411 (25%) Validation: 44/176 (25%)	<u>Forward selection (FS):*</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Backwards elimination (BE):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Stepwise selection (SS):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Lasso:</u> Country, employment, extrapulmonary, cavity size, decrease in lung capacity, smear microscopy, drug sensitivity, chest imaging <u>Random forest (RF):</u> Top 5 by mean decrease accuracy: lung cavity size, type of resistance, employment status, country, total cavities Top 5 by mean decrease Gini index: Age of onset, drug regimen, lung cavity size, number of daily contacts, culture	$\frac{FS:*}{AUROC = 0.74}$ (95% CI: 0.66-0.82) Sensitivity = 0.36 Specificity = 0.89 Misclassification = 0.24 <u>BE:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.3 Specificity = 0.88 Misclassification = 0.27 <u>SS</u> : AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>Lasso</u> : AUROC = 0.72 (95% CI: 0.64-0.80) Sensitivity = 0.21 Specificity = 0.96 Misclassification = 0.23 <u>RF</u> : AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.96 Misclassification = 0.23 <u>RF</u> : AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>SVM linear</u> : AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0.21 Specificity = 0.94 Misclassification = 0.24 <u>SVM polynomial</u> : AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0 Specificity = 1 Misclassification = 0.25	List	Unclear, Unclear, Unclear, High

Baussano(46) /	Pulmonary TB cases	2001 - 2005	Retrospective	Italy	Internal	576/1242 (46%)	Residency (residential vs. homeless), sex, geographic origin (non-EU vs. EU), case definition (other than definite vs. definite),	AUROC= 0.75 Calibration slope = 0.98	Nomogram	Low, Unclear,
2008 ^s 2	,		cohort		(bootstrap)		treatment setting (inpatient and unknown vs. outpatient), age (continuous)	$R^2 = 0.24$		Low, High
B 4 Costa-Veiga(47) / 2017§	Pulmonary TB cases	2000 - 2012	Retrospective cohort	Portugal	External (temporal)	Development: 1152/10766 (11%) Validation: 4714†	HIV, previous treatment, age class (25-44, 15-24, 45-64, >64), IV drug use, pathologies (other disease comorbidity)	AUROC= 75.9% (95% CI: 74.1-77.7) Sensitivity = 71% Specificity = 73%	Nomogram	Low, Low, Low, High
6 7 8 9 Killian(33) / 2019 [§] 10 11	TB patients (99DOTS program)	Feb 2017 – Sep 2018	Retrospective cohort	India	None	433/4167 (10%)	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 <u>w-misses</u> : missed doses in last week t-misses: total missed doses in 35 days units and a dense layer with 48 units for the dense layer and 4 units for the penultimate layer <u>Random forest</u> : 150 trees and no max depth based on DAT from first 35 day	<u>LEAP*</u> AUROC = 0.743 <u>lw-misses:</u> AUROC = 0.607 <u>t-misses:</u> AUROC = 0.630 <u>Random forest:</u> AUROC = 0.722	None	High, High, Unclear, High
12 13 14 ^{Madan(50) / 2018§} 15	TB-HIV patients on DOT with first-line TB treatment	2015	Retrospective cohort	India	None	78/448 (17%)	Sputum smear grade, previous TB,; disease classification, HIV status, ART status, CD4 cell count, sex and age group (with interaction terms between age group and sex; sputum smear status and type of TB; HIV status at TB diagnosis and CD4 cell category).	AUROC = 0.783 HL test p-value = 0.149	Model coefficients	Low, Low, Low, High
16 17 ^{Mburu(75) / 2018§}	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	32/340 (9%)	HbA1c, treatment regimen (retreatment), creatinine, BMI, BUN, weight, age, random blood glucose, HIV positive result, male gender	AUROC = 0.65 ± 0.06	Relative score	Low, Low, Low, High
19 Other outcome										
20 2 Kalhori(79) (fuzzy) / 20091 22	TB patients at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 7254† Validation: 2418†	Case type, treatment category, risky sex, prison, sex, recent TB infection, diabetes, low body weight, TB type, length, previous imprisonment, age, area, HIV	Mean absolute percentage error = 1.24	Learned parameters	Unclear, Unclear, High, High
23 24. Hussain(55) / 2019 25 26	Pulmonary and extrapulmonary TB patients (TB Reach)	2011 - 2014	Retrospective cohort	Unknown	Internal (split-sample)	Development: 3371† Validation: 842†	Random forest*, artificial neural networks, and SVM	Random forest:* Accuracy = 76.32%	None	Unclear, Unclear, Unclear, High
Abbreviations 7 *Indicates best 28 validation, inti- 29 *Outcome is c 30 *Outcome is c *Outcome is a 31 #Outcome is tr 32 33 34 35	: AUROC=Area under re t-performing/most relevant ernal validation, no valida nber unknown omposite of death and tre omposite of death, treatm value from 1 to 5 (1= patt eatment completion	ecciver operating nt model, which ation). If internal eatment failure (luent failure, loss tient completed t	characteristic; c-s is included throug and external valid osses to follow-up o follow-up, and n he treatment cours	tatistic=concord hout the manusc lation were perf and not evaluat not evaluated ie in frame of D	lance statistic; DOTS= rript (see methods sect formed, both are report ed (unknown) outcom OTS, 2=cured, 3= quit	Directly Observed Tl ion for details). Perfo ed. es were excluded) treatment, 4=failed t	herapy, DM=Diabetes; HL=Hosmer-Lemeshow; TB=Tuberculosis; rmance measures are reported for highest level of validation performe reatment and 5=death)	d (ranked from strongest to we	akest: external	

Table 3. Characteristics of patient populations in the 33 included studies with prediction models for tuberculosis treatment outcomes

reporting characteristic, n (% of total)Median [IQSample size33 (11)-Study duration, years32 (97)-Study design33 (100)Prospective cohort3 (9)Retrospective cohort25 (76)Nested case-control3 (9)Data source33 (100)Medical recordMedian [IQNested case-control3 (9)Non-nested case-control2 (6)Data source33 (100)Median [IQNested case-control2 (6)Data source33 (100)Medical record6 (18)	,
characteristic, n (% of total)characteristic, n (% of total)n (% of total)Sample size33 (11)-803 [291, 410]Study duration, years32 (97)-4 [2,7]years4 [2,7]Study design33 (100)Prospective cohort3 (9)Retrospective cohort25 (76)Nested case-control3 (9)Non-nested case-control3 (9)Data source33 (100)Medical record6 (18)Netional resistance resultance12 (20))R]
n (% of total) 803 [291, 410] Sample size 33 (11) - 803 [291, 410] Study duration, 32 (97) - 4 [2,7] years - - 4 [2,7] Study design 33 (100) Prospective cohort 3 (9) Retrospective cohort 25 (76) 3 (9) Nested case-control 3 (9) Non-nested case-control 2 (6) Data source 33 (100) Medical record 6 (18)	
Sample size $33 (11)$ - $803 [291, 410]$ Study duration, years $32 (97)$ - $4 [2,7]$ Study design $33 (100)$ Prospective cohort $3 (9)$ Retrospective cohort $25 (76)$ Nested case-control $3 (9)$ Non-nested case-control $2 (6)$ Data source $33 (100)$ Medical record $6 (18)$	
Study duration, years32 (97) (97)-4 [2,7]Study design33 (100)Prospective cohort3 (9)Retrospective cohort25 (76)Nested case-control3 (9)Non-nested case-control3 (9)Data source33 (100)Medical record6 (18)Nested case-control12 (20)	67]
yearsProspective cohort3 (9)Study design33 (100)Prospective cohort3 (9)Retrospective cohort25 (76)Nested case-control3 (9)Non-nested case-control2 (6)Data source33 (100)Medical record6 (18)Nested case-control12 (20)	
Study design33 (100)Prospective cohort3 (9)Retrospective cohort25 (76)Nested case-control3 (9)Non-nested case-control2 (6)Data source33 (100)Medical record6 (18)Nested case-control12 (20)	
Retrospective cohort25 (76)Nested case-control3 (9)Non-nested case-control2 (6)Data source33 (100)Medical record6 (18)Notice less interest and pressible response inter	
Nested case-control3 (9)Non-nested case-control2 (6)Data source33 (100)Medical record6 (18)Netional matrixes as pressillance12 (20)	
Non-nested case-control2 (6)Data source33 (100)Medical record6 (18)National resistance12 (20)	
Data source33 (100)Medical record6 (18)Netice of environmentation12 (20)	
$\mathbf{N}_{\mathbf{r}} = 1_{\mathbf{r}} = $	
National registry of surveillance 13 (39)	
system	
Local registry or surveillance 1 (3)	
system	
Regional registry or surveillance2 (6)	
system	
Data collect form for study11 (33)	
purposes	
Study region 32 (97) Africa 8 (25)	
Asia 13 (41)	
Europe 6 (19)	
North America 4 (12)	
South America 0 (0)	
Global 1 (3)	
High burden TB 31 (94) All 143(42)	
setting*	
Some 1 (3)	
None 17 (55)	
Missing data18 (54)Complete case-analysis9 (50)	
Missing indicator method 4 (22)	
Heckman's method 1 (6)	
Simple imputation 2 (12)	
Sensitivity analysis with 1 (6)	
imputation	
Other 1 (5)	
Number of models 33 (100) 1 25 (76)	
2 4(12)	
3 1(3)	

		4	2 (6)
		7	1 (3)
Reasons for multiple models developed	8 (24)	Different outcomes	1 (12)
		Different predictors considered	4 (50)
		Different methods	2 (25)
		Different outcomes	1 (12)
		Different populations and	1 (12)
		outcomes	

*Determined based on study location and WHO list of 30 high-burden TB countries in the 2019 Global Tuberculosis Report (1).

		Included?		
Characteristic	Yes	No	Unknown	Median [IQR] [‡] , n
Age*	18	-	15	41 [37, 49],
HIV	18	7	8	23% [10-100], n=17
Diabetes	12	2	19	12% [5-21], n=11
MDR	8	7	18	1% [1-3], n=8
Other drug resistance	12	1	20	6% [4-12], n=10
Extrapulmonary TB [†]	22	4	7	11% [4-17], n=16
Previous TB	20	1	12	19% [9-30], n=17
DOT	14	0	19	100% [100-100], n=14
Hospitalized patients	13	1	19	100% [100-100], n=10

Table 4. Study population characteristics of 33 included studies

Abbreviations: DOT=directly observed therapy; IQR=interquartile range; MDR=multi-drug resistance; TB=tuberculosis

*Based on the measure of central tendency reported in the study (mean: n=11; median: n=7) *Forms of extrapulmonary TB differ by study but included some of the following: Miliary, meningeal, pleural, peritoneal, disseminated, blood/bone, abdominal

[‡]Other than age (which is reported in years), this is the percentage of the population that has the characteristic among studies that include patients with the characteristic. For example, among the 18 studies that include persons with HIV, 17 report how many people had HIV and among those, the median percentage of the population with HIV is 23%.

Table 5. Methods reported for the 37 models of the 33 included studies with prediction models for tuberculosis treatment outcomes

Characteristic	Studies reporting	Categories	N(%) or median
	characteristic,		[IQR]
	n (%)		
Type of outcome	37 (100)	Single	29 (78)
		Composite	8 (22)
Outcome	37 (100)	Death	16 (43)
		Treatment failure	6 (16)
		Default, Loss to follow-up,	6 (16)
		or treatment interruption	
		Unfavorable outcome	6 (16)
		Treatment success	2 (6)
	6	Other [‡]	1 (3)
Number - prevalence of	32 (87)	-	94 [38-171]
outcome*			15% [9-26]
Events per candidate variable [†]	30 (81)	-	6 [3-11]
Events per variable (in final model)	29 (78)	-	14 [9-26]
Predictor types		Clinical/epidemiologic	34 (92)
		Adherence	1 (3)
		Biomarker	2 (5)
Analysis	37 (100)	Logistic regression	29 (78)
		Survival analysis	3 (8)
		Machine learning	5 (14)
Method for considering predictors in multivariable models	36 (97)	All candidate predictors	12 (32)
		Based on unadjusted association with outcome	19 (51)
		Based on clinical relevance	1 (3)
		Other§	4 (14)
Selection of predictors during modeling	31 (84)	Full model approach	2 (6)
		Forward selection	7 (23)
		Backwards elimination	5 (16)
		Stepwise selection	8 (26)
		Random Forest	1 (3)
		Hosmer-Lemeshow model	4 (13)
		building criteria	
		Bayesian model averaging	3 (10)
		Pairwise selection	1 (3)

1	
2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
10	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
20	
20	
51	
32	
33	
34	
35	
36	
37	
38	
39	
40	
/1	
40	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
52 52	
22	
54 55	
55	
56	
57	
58	
59	

60

P-value for consideration in	17 (46)	0.01	2 (12)
model	17 (40)		2 (12)
		0.05	3 (18)
		0.11	1 (6)
		0.5	6 (35)
		0.25	5 (29)
P-value for retention in MV model	20 (54)	0.02	9 (45)
		0.1	9 (45)
		0.12	1 (5)
		0.2	1 (5)
Internal validation	19 (51)	Split-sample	10 (53)
		Bootstrap	5 (26)
		Cross-validation	4 (21)
External validation	6 (16)	Temporal	1 (17)
	6	Geographic	1 (4)
-		Setting	4 (67)
Calibration	17 (46)	Calibration plot [¶]	2 (12)
		Calibration slope [¶]	1 (6)
		Hosmer-Lemeshow	13 (77)
		goodness of fit p-value [¶]	0.51 [0.20,
		-	0.79]
		Calibration table [¶]	2 (12)
		Mean absolute error [¶]	1 (6)
Discrimination	30 (81)	C-statistic (AUROC) [¶]	30 (100)
			0.75 [0.68-
			0.84]
		Log rank test [¶]	2 (5)
Classification	18 (49)	Sensitivity	14 (78)
			70 [54, 78]
		Specificity	13 (72)
		1 5	75 [71, 88]
		Accuracy	2 (11)
		Other**	2 (11)
Model presentation	34 (92)	Risk score	16 (43)
		Model coefficient	8 (22)
		Nomogram	2 (6)
		Odds ratios/relative scores	4 (12)
			• ()

Abbreviations: AUROC=area under receiver operating characteristic; c-statistic=concordance statistic

*Prevalence of outcome in the population used to develop the prediction model (i.e.

derivation/development subset if split-sample technique was used or full sample if the model was not validated or if bootstrap/cross-validation was used)

[†]Only 5 studies report the exact number of predictors considered. Otherwise, the number of candidate predictors was estimated from the provided tables or lists of candidate predictors in the source paper.

[‡]Outcome is a value from 1 to 5 (1= patient completed the treatment course in frame of DOTS, 2=cured, 3= quit treatment, 4=failed treatment and 5=death)

[§]Other methods of determining which variables to consider for prediction model include: principal components analysis (n=1), screening for multi-collinearity via correlation coefficient (n=1), one study used a combination of a priori and selection via univariable association, and the other used machine learning pre-processing (n=1)

[¶]Sums to more than 100%, because some studies report multiple measures of calibration or discrimination

^IBased on the following cut-off methods: Youden (n=4) concordance probability (n=1), estimated at nearest 0,1 for studies that present a range of sensitivity and specificity in a table or figure (n=4), or unknown (n=5)

**Other includes one study that reports false positive rate and one study that includes a graph of sensitivity vs. specificity.

Figure 2. Most common predictors considered and included

[See Figure 2]

Figure 2 legend:

Considered: the predictor as evaluated as a candidate predictor prior to multivariable modeling Included: the predictor was considered and subsequently included in the final multivariable model

Figure 3. Heatmap of signaling questions from risk of bias assessment with PROBAST

[See Figure 3]

Figure 3 legend:

PROBAST questions (additional details in Supplemental File 5)

- Participants 1: What study design was used and was it appropriate?
- Participants 2: Were all inclusion and exclusion criteria appropriate?
- Predictors 1: Were predictors defined as assessed the same way for all participants?
 - Predictors 2: Were predictor assessments made without knowledge of data outcome?
 - Predictors 3: Are all predictors available at the time the model was intended to be used?
- Outcome 1: Was the outcome determined appropriately?
- Outcome 2: Was the outcome pre-specified or standard?
- Outcome 3: Were predictors excluded from outcome definition?
- Outcome 4: Was the outcome defined and determined in a similar way for all participants?
- Outcome 5: Was the outcome determined without predictor information?
- Outcome 6: Was the time interval between predictor assessment and outcome determination appropriate?
- Analysis 1: Were there a reasonable number of participants with the outcome?
- Analysis 2: Were continuous and categorical variables handled appropriately?
- Analysis 3: Were all enrolled participants included in the analysis?
- Analysis 4: Were participants with missing data handled appropriately?
 - Analysis 5: Was selection of predictors based on univariable analysis avoided?
 - Analysis 6: Were complexities in data (censoring, competing risks, sampling of control participants) accounted for appropriately?
 - Analysis 7: Were relevant model performance measures evaluated appropriately?

Analysis 8: Were model overfitting, underfitting, and optimism in the model performance accounted for?

Analysis 9: Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?

Figure 4. Summary of risk of bias and applicability assessment with PROBAST

[See Figure 4]

tor oper teries only

BMJ Open





203x203mm (300 x 300 DPI)



203x203mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml





203x203mm (300 x 300 DPI)

BMJ Open

Supplemental File 1. PRISMA Checklist

Section/topic	#	Checklist item	Reported of #
TITLE			-
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			<u>I</u>
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).	Supplement 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
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.	Supplement 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; Suppleme Files 4 and
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; Suppleme 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		<u>.</u>	-
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; Tabl 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; Figu and 4

	-		
Results of individual studies	Results of ndividual 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	thesis of results 21 Present results of each meta-analysis done, including confidence intervals and measures of consistency.		N/A
Risk of bias across studies	sk of bias across 22 Present results of any assessment of risk of bias across studies (see Item 15). udies 22		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.	21

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
	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
Outcome	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.
Timina	The timespan of prediction may vary between studies, depending on the duration of treatment and follow-up, but we expect most
1 mmg	studies will evaluate endpoints around 6-9 months.
Sotting	Model designed for use in clinical or hospital setting at the time of TB treatment initiation to aid in targeted treatment or programmatic
Setting	support for individuals at greatest risk for unsuccessful TB treatment outcomes.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Supplemental File 3. Search Strategy

	1. ((validat*[tiab] OR predict*[ti] OR rule*[tiab])
PubMed	 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 clinical*[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 "Logistic Models"[Mesh])) OR (prognostic[tiab] AND (history[tiab] OR variable*[tiab] OR criteria[tiab] OR scor*[tiab] OR characteristic*[tiab] OR finding*[tiab] factor*[tiab] OR model*[tiab]) OR (prognostic[tiab] AND (history[tiab] OR variable*[tiab] OR criteria[tiab] OR scor*[tiab] OR characteristic*[tiab] OR finding*[tiab] 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]) 2. (stratification[tiab] OR "ROC Curve"[Mesh] 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 success*[tiab] OR unsuccess*[tiab] OR die[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	 (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. (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"/ tuberculosis/ or tuberculosis.ti,ab (outcome\$ or mortality\$ or death\$ or fail\$ or recur\$ or relapse\$ or default\$ or abandon\$ or loss\$ or cure\$ or success\$ or unsuccess\$ or do or died or dies).ti,ab. 1 or 2 3 and 4 5 and 6 limit 7 to (human and yr="1995 -Current")
Web of Science	 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 identity 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*)) TS=(stratification or discrimination or discriminate or c-statistic or "c statistic" or "area under the curve" or AUC or calibration or indice algorithm or multivariable or "receiver operating characteristic") TS=(tuberculosis) TS=(outcome* or mortality* or death* or fail* or recur* or relapse* or default* or abandon* or loss* or cure* or success* or unsuccess* die or died or dies) 1 or 2 3 and 4 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	on pa				
SOURCE OF DATA	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)					
	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting,					
	inclusion and exclusion criteria)					
PADTICIPANTS	Participant description					
TARITCH ANTS	Details of treatments received, if relevant					
	Study dates					
	Definition and method for measurement of outcome					
	Was the same outcome definition (and method for measurement) used in all patients?					
OUTCOME(S) TO BE	Type of outcome (e.g., single or combined endpoints)					
PREDICTED	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					
	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease					
	characteristics)					
	Definition and method for measurement of candidate predictors					
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)?					
(OR INDEX TESTS)	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)					
	Number of participants with any missing value (include predictors and outcomes)					
MISSING DATA	Number of participants with missing data for each predictor					
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)					
	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)					
MODEL DEVELOPMENT	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)					
	Calibration (solibration plate solibration plane Harman Lamothers test) and Discrimination (Casteristic D					
	Canoration (canoration piot, canoration stope, Hosmer-Lemesnow test) and Discrimination (C-statistic, D- statistic log-rank) measures with confidence intervals					
PERFORMANCE	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a- priori cut points were used					
	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)					
MODEL EVALUATION	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)					
	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)					
RESULTS	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance					

BMJ Open

	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/probabl	y yes	No/probably no		No information	
1	What study design was used and was it appropriate?	Yes: If a cohort design or proper registry data) you have confidence in participant enrollment is described Probably yes: a nested of case-cohort design (with adjustment of the baselin the analysis) has been u design was used but par enrollment was data qua	(including RCT was used and data quality and s clearly case–control or h proper ne risk/hazard in sed or a cohort ticipant ality is unclear	No: If a non-nested case–control de has been used Probably no: a nested case-control s was used without proper adjustment baseline risk/hazard	sign Ii is study t of	f the method of participant sampling s unclear.	
2	Were all inclusion and exclusion criteria appropriate?	enrollment was data quality is unclear 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.	
				6.	1		
	Low risk of	bias		High risk of bias		Unclear risk of bias	
If the "Prob ≥1 of could	answer to all signaling quest bably yes," then risk of bias c the answers is "No" or "Prol still be "Low risk of bias" b	tions is "Yes" or an be considered low. If bably no," the judgment ut specific reasons should	If the answer to a "No" or "Probab except if defined	any of the signaling questions is oly no," there is a potential for bias, I at low risk of bias above.	If releva the sign signalin domain	ant information is missing for some of haling questions and none of the ng questions is judged to put this at high risk of bias.	

Domain 2: Predictors

be provided why the risk of bias can be considered low.

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 l outcome information during predictor asses entirely clear (retrosp collection/surveillanc , All included predicto	ikely that was not used ssment, but not ective data e data) ors would be	Predictors would not be avail	able at	No information on whether
3	Are all predictors available a the time the model was inten to be used?	t available at the time t intended to be used for	he model is or prediction	the time the model is intended used for prediction.	to be	predictors would be available a time the model is intended to b used for prediction.
	Low risk of	bias	H	ligh risk of bias		Unclear risk of bias
If th "Pro ≥1 o cou be p	ne answer to all signaling ques obably yes," then risk of bias of of the answers is "No" or "Pro ld still be "Low risk of bias" b provided why the risk of bias of	tions is "Yes" or an be considered low. If bably no," the judgment ut specific reasons should an be considered low.	If the answer to a "No" or "Probabl bias, except if de	ny of the signaling questions is ly no," there is a potential for fined at low risk of bias above.	If relevar the signal signaling at high ri	nt information is missing for some ling questions and none of the questions is judged to put this do sk of bias.
Don Bias outo met dete	main 3: Outcome s in model performance can oc come determination can result thods are inconsistently applied ermination can also result in bi	cur when methods used to do from use of suboptimal meth d across participants, or wher as.	etermine outcomes ods, tests, or criter h knowledge of pre	s incorrectly classify participants we ria that lead to unacceptably high redictors influence outcome determ	vith or with levels of err ination. Inc	out the outcome. Bias in methods rors in outcome determination, wl correct timing of outcome
	Signaling question	Yes/probably	yes	No/probably no		No information
1	Was the outcome determined appropriately?	If a method of outcome determination ha 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 outcom determined
2	Was the outcome pre- specified or standard?	of the outcome method is appropriate). Yes: If the method of outcome determination is objective, or if a standar 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		No: If the outcome definition wa standard and not prespecified Probably no: a non-standard or n prespecified outcome was used, a unclear whether the outcome deficult could introduce bias. *Caution with composite outcome favor a better model by excluding outcome components or including atypical events	s not on- and it is inition ees that g typical g	No information on whether the outcome definition was prespec or standard
3	Were predictors excluded from outcome definition?	inadvertently alter study results 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 outcome definition	of the	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 determined in a different way for participants	and some	No information on whether out were defined or determined in a similar way for all participants
5	Was the outcome determined without predictor information	Yes: If predictor informati known when determining t status, or outcome status d clearly reported as determi knowledge of predictor inf Probably yes: predictor inf have been available at time assessment, but outcome d objective and knowing info predictors would not influe	on was not the outcome etermination is ned without formation. Formation might e of outcome efinition is ormation about ence outcome	No: If it is clear that predictor inf was used when determining the c status Probably no: it is likely predictor information was available at the outcome assessment, and outcom definition is subjective and know predictors could influence outcor determination.	formation putcome time of lee ledge of ne	No information on whether out was determined without knowle of predictor information

6	Was the time interval between predictor assessment and outcome determination appropriate	ased on culture results, f the time interval betwee sessment and outcome vas appropriate to enabl nd representative numb utcomes to be recorded afformation on the time is equired to allow a repre- f the relevant outcome redictor assessment and etermination were from the within an appropri-	etc) een predictor determination e the correct type er of relevant , or if no interval is sentative number occur or if loutcome information ate time interval.	If the tin assessme too short type and relevant	ne interval between predi ent and outcome determi t or too long to enable the l representative number o outcomes to be recorded	ictor nation is e correct f	If no information was provide the time interval between pre- assessment and outcome determination.
	Low risk of bia	s	Н	ligh risk	of bias		Unclear risk of bias
If th "Pr ≥1 cou sho low	he answer to all signaling question obably yes," then risk of bias can of the answers is "No" or "Probab Id still be "Low risk of bias" but s ould be provided why the risk of b 7.	is is "Yes" or be considered low. If ly no," the judgment pecific reasons as can be considered	If the answer to an "No" or "Probably bias, except if def	ny of the s y no," the fined at lo	signaling questions is rre is a potential for w risk of bias above.	If relevar the signal signaling at high ri	It information is missing for sor ling questions and none of the questions is judged to put this sk of bias.
Do Sta bia see	main 4: Analysis tistical analysis is a critical part of s in reported model performance i k statistical advice when completi	Prediction model deven neasures. Model develo	lopment and validati pment studies includ	ion. The u de many s	use of inappropriate statis steps where flawed metho	tical analys ods can dist	is methods increases the potent ort results. We recommend revi
	Signaling question	Yes/	probably yes	numbor	No/probably	y no	No information
1	Were there a reasonable number of participants with the outcome	of participants with the number of can- is ≥20 (EPV ≥20). ? For model validati participants with th	the outcome relative didate predictor para * on studies, if the nume outcome is ≥100.	we to ameters mber of	the number of participa outcome relative to the candidate predictor par <10 (EPV <10).* For model validation st number of participants outcome is <100.	ants with the number of ameters is tudies, if the with the	 no information on the nu candidate predictor parar or number of participants the outcome, such that th cannot be calculated. For model validation studinformation on the numb participants with the outcome.
		* For EPVs betwee frequency, overall 145 to 147.	en 10 and 20, the ite model performance	em should , and distr	be rated as either probab ribution of the predictors	ly yes or pr in the mode	obably no, depending on the ou el. For more guidance, see refer
2	Were continuous and categorica predictors handled appropriately	Yes: If continuous continuous or if co- examined as linear restricted cubic spi polynomials. Probably yes: If co- converted into >2 the model (i.e., dio using a prespecifie avoids sparse data improve statistical For model validati predictors are inclu- definitions or trans- variables are categ points, ascompare- study.	predictors are kept a ntinuous predictors or non-linear using lines or fractional ontinuous predictors categories when incl hotomized or catego d method or in a wa 'would not intention significance. on studies, if continu- uded using the same formations, and cate orized using the same d with the developm	as are luded in orized) ny that ally uous egorical ne cut ent	No: For model develop if continuous predictor converted into 2 catego included in the model. Probably no: If categon group definitions do no prespecified method or variables were split int but the decision of how variables is unclear. For model validation si continuous predictors a using different definition transformations, or cate variables are categorizz different cut points, as with the development s	ment studie s are ories when ical predict t use a continuous o >2 groups t to split tudies, if ture included ons or egorical ed using compared study.	 No information on wheth continuous predictors are examined for nonlinearity no information on how categorical predictor grou defined. For model validation stud information on whether ti same definitions or transformations and the s cut points are used, as co with the development stud
3	Were all enrolled participants included in the analysis?	If all participants e included in the dat	nrolled in the study a analysis.	are	If some or a subgroup are inappropriately exc the analysis (because the missing data, unknown outliers)	of participa luded from ney were outcome,	nts No information on wheth enrolled participants are included in the analysis.
		Yes: If there are no predictors or outco explicitly reports t	o missing values of omes and the study hat participants are r	not	No: If participants with are omitted from the ar the method of handling is clearly flawed e a	n missing da nalysis, or in g missing da missing	ta If there is insufficient information to determine method of handling miss is appropriate

		Probably yes: If a small percenta with missing data were excluded provide comparison of included	age of persons l and authors vs. excluded	if the study had no explicit me of methods to handle missing	ention data.	
		participants or if sensitivity anal imputation methods are convinc is low	ysis with ing that bias	Probably no: If authors provid comparison of included vs. ex participants or if sensitivity ar with imputation methods are reported, but the results are no convincing to rule out bias fro excluding missing data	le acluded nalysis ot om	
5	Was selection of predictors based on univariable analysis avoided?	If the predictors are not selected of univariable analysis prior to n modeling.	on the basis nultivariable	If the predictors are selected or basis of univariable analysis p multivariable modeling.	on the prior to	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 accoun for appropriately, or if it is clear that any potential data complexities have been identified appropriately as unimportant.		If complexities in the data tha affect model performance are ignore. For example, case-cor studies that do not estimate ba risk or studies with censoring competing risks that do not us survival analysis or other appropriate methods.	t could ntrol aseline or se	No information is provided on whether complexities in the da 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 evaluat if only goodness-of-fit tests (Hosmer-Lemeshow test), are to evaluate calibration or if fo models predicting survival ou performance measures accour for censoring are not used, or classification measures (like sensitivity, specificity, or pred values) were presented using predicted probability threshol derived from the data set at he but calibration is not otherwis evaluated.	ed, or used r tcomes ting if dictive ds und, e	Either calibration or discrimination are not reported or no information is provided a to whether appropriate performance measures for survival outcomes are used (e., references to relevant literature or specific mention of methods such as using Kaplan–Meier estimates), or no information of thresholds for estimating classification measures is give
8	Were model overfitting, underfitting, and optimism in model performance accounted for?	Yes: If internal validation techni (bootstrapping and cross-validat all model development procedur to account for any optimism in r and subsequent adjustment of th performance estimates were app Probably yes: If internal validati and optimism was estimated as then optimism-corrected perform measures were not appropriately (accounting for all model develop procedures)	ques ion) including es, were used nodel fitting, e model lied. on was used very low, and nance calculated opment	No: If no internal validation h been performed, or if internal validation consists only of a s random split-sample of partic data, Probably no: Internal validatio bootstrapping or cross-validat was conducted but did not inc all model development proced including any variable selection were not used to correct mode performance measures.	as ingle ipant on with ion hude hures on or el	No information: No informatio is provided on whether interna validation techniques, includir 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 in the final model correspond to results from multivariable analys	coefficients reported sis.	If the predictors and regressio coefficients in the final model correspond to reported results multivariable analysis. (i.e. ro of model coefficients to creat "risk score" are inappropriate determined).	n from unding e a ly	If it is unclear whether the regression coefficients in the final model correspond to reported results from multivariable analysis.
	Li-lf	ki				
If tl risk "Pr spe con	Low FISK OF ne answer to all signaling questions is to f bias can be considered low. If ≥ 1 obably no," the judgment could still li- cific reasons should be provided why sidered low.	s "Yes" or "Probably yes," then of the answers is "No" or be "Low risk of bias" but y the risk of bias can be	If the answer questions is "] a potential for risk of bias ab	to any of the signaling No" or "Probably no," there is bias, except if defined at low bove.	If relev some o of the s this dor	ant information is missing for f the signaling questions and nor ignaling questions is judged to p main at high risk of bias.
Ap	plicability					

Page 61 of 66

BMJ Open

<u>Participants</u> : 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.
<u>Predictors</u> : 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.

to been even only

Supplemental File 6. Model outcome definitions

Study ID	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
Aliohaney / 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 /	Douth	Found that occurred during the most of months area anglious were emissived as 1D doubt
2019	Death	The outcome was mortality risk at 2 months after admission. '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
Koegelenberg / 2015	Death	any cause. Patients were categorised as either ICU/hospital survivors or non-survivors
Nguyen (general	Death	
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
2013	Death	Death within 12 months of TB diagnosis
Valade / 2012	Death	Final outcomes of survival or death were recorded
Wang / 2019	Death	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 has 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) Default	Never defined
Belilovsky / 2010	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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	(interruption >2	
	months)	
	Default,	
	Abandon, or	
	LTF	Defaulter or cases were defined as patients on anti-tuberculosis treatment at the TBCU who failed to turn up for the
	(interruption >2	scheduled appointments despite usual attempts to recall them by phone or mail, as described below, and from whom
Chee / 2000	months)	least one home visit during the study was recorded
	Default,	
	Abandon, or	
	LTF	
C1 1	(interruption >2	
Cherkaoui / 2014	months)	I reatment default was defined as an interruption in 1B treatment for >=2 consecutive months.
	Default,	
	Abandon, or	
	LIF (intermention >2	Intermetion of tractment for our proper for more than 2 months non-completion of tractment within 0 months when
Dadriga / 2012	(interruption > 2	Interruption of treatment for any reason for more than 2 months, non-completion of treatment within 9 months where $r_{1}^{(1)}$ and $r_{2}^{(1)}$ is also an equation of the interval of $r_{2}^{(2)}$ (b) the presented data.
Kourigo / 2012	Treatment	patient is placed on a 6 month regiment of drug intake of ~80% the prescribed dose.
(predicting) /	success (cure +	
2009	completion)	For each nationt dependent variable was recorded whether or not the nationt finished the treatment course and get out
2007	Unfavorable	To reach parton dependent variable was recorded whener of not the patient mission the reachefft course and get ou
	outcome (death	
Sauer / 2018	+ failure)	The primary outcome was treatment failure, which we defined as failure of therapy or death
	Unfavorable	in the second we dealed the second of the second se
	outcome (death	
	failure. LTF.	Treatment interruption or default, treatment failure, transferred out cases and those lost to follow-up were grouped
Baussano / 2008	NE)	'unsuccessful outcomes
	,	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 a
		declared cured, including both negative smear microscopy at the end of treatment at least one previous follow-up tes
	Unfavorable	in case of not providing sputum samples, cure is declared if treatment completed and absent of disease clinical evide
	outcome (death,	(categories 1 and 2). (ii) Unsuccessful outcome-if treatment of PTB patients resulted in failure (i.e. remaining smear
Costa-Veiga /	failure, LTF,	positive after 5 months of treatment, cat. 3), default (i.e. patients who interrupted their treatment for two consecutive
2017	NE)	months or more after registration, cat. 4), death (cat. 5) or were transferred-out (cat. 6)
	Unfavorable	
	outcome (death,	
	failure, LTF,	We label 'Cured' and 'Treatment Complete' to be favorable outcomes and 'Died', 'Treatment failed', and 'Lost to follo
Killian / 2019	NE)	up' to be unfavorable outcomes
	Unfavorable	
	outcome (death,	
NC 1 / 0010	failure, LTF,	Favourable treatment outcomes included cure and treatment completed. Unfavourable treatment outcomes included
Madan / 2018	NE)	death, loss to follow-up, treatment failure, transfer out, or a switch to MDR TB treatment.
	Untavorable	
Missing / 2010	outcome (death,	
Mburu / 2018 -	tailure, LTF,	
Untavorable	NE)	I he primary analyses compared tavorable versus unfavorable outcomes at end of treatment
TZ 11 1 (2))	Other	The values of outcomes might be any values from 1 to 5 which means different outcomes. Value 1 means patient
Kalhori (fuzzy) /	composite	completed the treatment course in trame of DOTS, 2 means the patient has been cured, 3 means patients has quitted
2009	outcome	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

2	Study ID	Final model
3	Abdelbary / 2017 -	2 + 2*(Age + 1-65) + 5*(Age = 65) + 2*(Male gender) + 4*(MDR TB) + 3*(HIV) + 3*(Malnutrition) + 2*(Alcoholism) + 2*(Alcohol
4	Death (2017	2*(Male*diabetes) + 3*(HIV*pulmonary TB) - 1*(diabetes) - 1*(pulmonary TB)
5	Abdelbary / 201 / - Failure	$8*(N_0 \text{ or low education}) + 40*(MDR) + 10*(AFB \text{ smear} + 2) + 15*(AFB \text{ smear} + 3)$
6	Abdelbary / 2017 -	
7	TB-DM / Death	2 + 3*(Male gender) + 3*(Malnutrition) - 1*(BCG vaccinated) - 1*(AFB smear positive)
8	Abdelbary / 2017 -	$2 + 2 + (4 - z_{1} + 2) + 2 + (4 + 1 - 1) + 4 + (1113)$
9	IB-DM / Default	$2 + 2^{(Age<40)} + 2^{(Male gender)} + 4^{(HIV)}$
10	Aljohaney / 2018	congestive heart failure (1.231), bilateral disease on chest x-ray (1.192)
11		3*(Hypoxemic respiratory failure) + 2*(Age>=50) + 1*(Bilateral involvement) + 1*(At least one of: HIV, diabetes, liver failure/cirrhosis,
12	Bastos / 2016	congestive heart failure, chronic respiratory disease) + 1*(Hemoglobin<12)
13	Baussano / 2008	vs. definite), treatment setting (inpatient and unknown vs. outpatient), age (continuous)
14	D 111 1 (2010	-3.2 + 0.8*(male gender) + 0.7*(unemployment) + 0.4*(retreatment case) + 1.1*(alcohol abuse) + 0.6*(no data about alcohol)
15	Belilovsky / 2010	0.8*(severe TB form) - 0.3*(urban residence) + 0.4*(age 25-50) + 0.8*(pulmonary TB) + 0.5*(prison history)
16	Chang / 2004	Current smokers (3.44), ex-smokers (2.48), history of default (10.74), no history of default (0.80),
17		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
18	CT (0000	friends (0.08), Treatment duration (1.85). Treatment duration is categorical as 6 months, 9 months, and >9 months, but only one OR is
19	Chee / 2000	presented.
20		take TB treatment? Are you taking a retreatment regimen for TB? Do you or doctor think you are having moderate or severe side effects
21		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
22	C1 1 : / 2014	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
23	Cherkaoui / 2014	know how long your 1B treatment is supposed to last (1 point for no) Have you ever smoked cigarettes (-1 point for no)
24	Costa-Veiga / 2017	yes/no)
25	0	9*(Male sex) + 7*(patient aged 55+) + 6*(currently taking ART) + 7*(unable to walk unaided) + 7*(hemoglobin <80, severe anemia) +
26	Gupta-Wright / 2019	6*(positive on urine TB-LAM)
27	Horita / 2013	1*Age (years) + 10*(oxygen requirement) - 20*(albumin) + 5*(semi-dependent, ADL) + 10*(total dependent, ADL)
28	Hussain / 2019	None
29		Learned parameters by training set for each predictor written as predictor (learned parameter): Case type (0.467), treatment category (-
30	Kalhori (fuzzy) /	(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
31	Z009 Kalhori (logistic) /	(0.950), length (-0.255), previous imprisonment (2.398), age (0.237), area (0.8895), HIV (0.731)
32	2010	exp(-0.93 - 0.71*(gender) + 0.02*(age) - 0.02*(weight) + 0.5*(nationality) + 0.99*(prison) + 0.16*(case type))
33	Kalhori (predicting) /	
34	2009	$\exp(-(1.58 - 0.12^{*}(age) + 0.807^{*}(gender) - 0.039^{*}(nationality) - 0.263^{*}(prison) + 0.15^{*}(area) + 0.021^{*}(weight))$
35	Keane / 1997	(3.6), any previous treatment (2.3), cavities (1.7), weight (0.98)
36		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
37		dense layer and 4 units for the penultimate layer
38	Killian / 2019	
39	Koegelenberg / 2015	radiograph showing miliary pattern/parenchymal infiltrates, absence of TB treatment at admission
40	Luies / 2017	Written as predictor (OR): 3.5Dihydroxybenzoic acid (25.6), 3-(4-Hydroxy-3-methoxyphenyl) propionic acid (1.3)
41		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),
42		New TB without smear (1.67), New TB with smear positive, unknown grade (1.00), Previously treated, smear negative TB (1.35),
43		previously treated with scanty smear (4. /4), 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 sputtum smear (2.46), previously treated with unknown grade
44		(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
45		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
46		(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>500 (2.40) HIV+ diagnosed after TB
47		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-
48	Madan / 2018	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)
49	Mburu / 2018 -	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
50	Failure	noarc, regimen, age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine, respectively Present relative scores for each covariate included not sure if this was how it should be used. Palative scores are 100, 70, 32, 70, 00, 62, 02
51	Mburu / 2018 -	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
52	Unfavorable	result, male gender, respectively
53 54	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]
54 55	Nguyen (TB-DM) /	16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[IDU] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliary TB] + 10*[Miliary TB] + 10*[Milia
56	2019	6*[AFB positive smear] + 24*[Positive HIV] Promotic score S*[A co >= 65] + 12*[Position of LTCE] + 0*[Maningsol TD] + 6*[store == 1 CVD] + 0*[discore = 1 id]
57	2018	positive culture or NAA] + 10° [culture not converted or unknown]
58	·	11

Page 65 of 66

BMJ Open

1		$\label{eq:constraint} \begin{aligned} \text{Model: -6.994499} + 1.069024*[\text{Age} >= 65] + 2.541147*[\text{Resident of LTCF}] + 1.998852*[\text{Meningeal TB}] + 1.37995*[\text{abnormal CXR}] + 1.899108*[\text{diagnosis confirmed with positive culture or NAA}] + 2.186305*[\text{culture not converted or unknown}] \end{aligned}$
2	Pefura-Yone / 2017	1/(1 + exp(-1.3120 + 0.0474*[age] - 0.1866*[adjusted BMI] + 1.1637*[PTB-] + 0.5418*[ETB] + 1.3820*[HIV]
3	Podlekareva / 2013	1*[DST performed] + 2*[Initial treatment with RHZ] + 2*[cART started before or up to 1 month after TB diagnosis]
4 5	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]
6 7	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)
, 8	Thompson / 2017	Heatmap of differentially expressed genes
9	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).
10	Wang / 2019	Unknown
11 12	Weise / 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
13 14	Zhang / 2019	2*[Anemia (HGB < 90g/L)]+ 2*[Tuberculous meningitis] + 5*[Severe pneumonia] + 2*[Hypoalbuminemia] + 7* [Unexplained infections or space-occupying lesions] + 5* [Malignancies]
14	6	
15 16		
17		
12		
10		
20		
20 21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32 22		
33 24		
24 25		
36		
37		
38		
39		
40		
41		
42		
43		
44		

2	
3	
4	
5	
5	
6	
7	
8	
٥ ٥	
20	
10	
11	
12	
12	
1.4	
14	
15	
16	
17	
10	
18	
19	
20	
21	
22	
22	
23	
24	
25	
25	
26	
27	
28	
20	
29	
30	
31	
32	
22	
24	
34	
35	
36	
37	
20	
38	
39	
40	
41	
12	
42	
43	
44	
45	
10	
46	
47	
48	
49	
50	
51	
52	
53	
55	
54	
55	
56	
57	
50	
20	
59	
60	

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.	
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	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	0 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.	
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

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
10	
11	
12	
13	
14	
15	
16	
10	
17	
18	
19	
20	
21	
21	
22	
23	
24	
25	
26	
27	
27	
28	
29	
30	
31	
32	
22	
22	
34	
35	
36	
37	
38	
30	
72	
40	
41	
42	
43	
44	
15	
40	
46	
47	
48	
49	
50	
50	
21	
52	
53	
54	
55	

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

BMJ Open

A systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults

Journal:	BMJ Open
Manuscript ID	bmjopen-2020-044687.R1
Article Type:	Original research
Date Submitted by the Author:	30-Dec-2020
Complete List of Authors:	Peetluk, Lauren; Vanderbilt University School of Medicine, Epidemiology Ridolfi, Felipe; Instituto Nacional de Infectologia Evandro Chagas Rebeiro, Peter; Vanderbilt University School of Medicine, Epidemiology; Vanderbilt University School of Medicine, Division of Infectious Diseases Liu, Dandan; Vanderbilt University School of Medicine, Biostatistics Rolla, Valeria; Instituto Nacional de Infectologia Evandro Chagas Sterling, Timothy; Vanderbilt University School of Medicine, Division of Infectious Diseases
Primary Subject Heading :	Infectious diseases
Secondary Subject Heading:	Global health, Patient-centred medicine, Public health
Keywords:	Tuberculosis < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, STATISTICS & RESEARCH METHODS

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

reliez oni

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults

Lauren S. Peetluk, MPH,¹ Felipe M. Ridolfi, MD, MSc,² Peter F. Rebeiro, PhD, MHS,^{1,3} Dandan

Liu, PhD,⁴ Valeria C. Rolla, MD, PhD,² Timothy R. Sterling, MD³

¹Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine,

Nashville, Tennessee, USA

²Instituto Nacional de Infectologia Evandro Chagas (INI) – Fiocruz, Rio de Janeiro, Brazil

³Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of

Medicine, Nashville, TN, USA

⁴Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA

teliezon

Corresponding author: Lauren S. Peetluk, MPH A2209 Medical Center North 1161 21st Avenue South Nashville, TN 37203 E-mail: lauren.s.peetluk@vanderbilt.edu

Word count main text: 3676
BMJ Open

ABSTRACT

Objective: To systematically review and critically evaluate prediction models developed to predict tuberculosis (TB) treatment outcomes among adults with pulmonary tuberculosis. **Design:** Systematic review

Data sources: PubMed, Embase, Web of Science, and Google Scholar were searched for studies published January 1, 1995 - January 9, 2020.

Study selection and data extraction: Studies that developed a model to predict pulmonary TB treatment outcomes were included. Study screening, data extraction, and quality assessment were conducted independently by two reviewers. Study quality was evaluated using the Prediction model Risk Of Bias Assessment Tool (PROBAST). Data were synthesized with narrative review and in tables and figures.

Results: 14,739 articles were identified, 536 underwent full-text review, and 33 studies presenting 37 prediction models were included. Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=9, 25%). Most models (n=29, 78%) measured discrimination (median c-statistic=0.75; interquartile range: 0.68-0.84), and 17 (46%) reported calibration, often the Hosmer-Lemeshow test (n=13). Nineteen (51%) models were internally validated, and six (16%) were externally validated. Eighteen studies (54%) mentioned missing data, and of those, half (n=9) used complete case analysis. The most common predictors included age, sex, extrapulmonary TB, body mass index (BMI), chest x-ray results, previous TB, and HIV. Risk of bias varied across studies, but all studies had high risk of bias in their analysis.

Conclusions: TB outcome prediction models are heterogeneous with disparate outcome definitions, predictors, and methodology. We do not recommend applying any in clinical settings

without external validation, and encourage future researchers adhere to guidelines for developing and reporting of prediction models.

Registration: The study was registered on the international prospective register of systematic reviews PROSPERO (CRD42020155782)

to pretterien ont

ARTICLE SUMMARY:

Strengths and limitations

- Prediction models for tuberculosis treatment outcomes have the potential to inform interventions or treatment management protocols to promote cure among tuberculosis patients at the greatest risk of unsuccessful treatment outcomes, but the methods and clinical utility of existing models had not been formally evaluated.
- This was the first systematic review of prediction models for tuberculosis treatment outcomes.
- The review used a comprehensive search strategy, conducted thorough bias assessment with the Prediction Model Risk of Bias Assessment Tool (PROBAST) tool, and offers recommendations for future model development and validation studies for predicting tuberculosis treatment outcomes.
- Evidence synthesis and quality assessment were limited by incomplete reporting in primary studies
- External validation studies or studies written in languages other than English, Spanish, Portuguese, or French were excluded.

BACKGROUND

Tuberculosis (TB) is one of the top ten causes of death worldwide and a leading cause of death from an infectious disease. In 2018, 10 million people developed TB and 1.45 million people died from it globally, despite widespread availability of curative treatment.[1] Global treatment success was 85% for all new and relapse TB patients in 2018. For HIV-associated TB, it was 75%. These proportions are lower than the End TB Strategy target of \geq 90% treatment success.[2]

Heeding early recognition that *Mycobacterium tuberculosis* develops resistance rapidly in response to single-drug therapy, TB has been treated with combination therapy for more than 50 years.[3] Aside from weight-based dosing, the World Health Organization (WHO) and other TB guidelines authorities recommend a standardized approach for treatment of almost all TB patients.[4–6] The current recommendation for treatment of drug-susceptible TB includes 2 months of isoniazid, rifampin, pyrazinamide, and ethambutol, followed by 4 months of isoniazid and rifampin. However, actual treatment regimens may vary due to differences in drug tolerability, and other individual-level factors that can affect TB treatment outcomes.

Due to the long duration of TB treatment, it would be beneficial for TB outcome studies to identify early treatment predictors of unsuccessful TB treatment outcomes to identify patients needing tailored treatment approaches, such as directly observed therapy (DOT) or extended treatment course. Research suggests that individual characteristics, such as HIV, age, undernutrition, diabetes, TB disease severity, extrapulmonary TB, history of TB, adherence, alcohol use, and adverse drug reactions, are associated with unsuccessful TB treatment outcomes, but results vary by setting and patient population.[7–10]

BMJ Open

Prediction models are defined as any combination or equation of two or more predictors, such as demographic factors, medical history, physical examination, and lab tests, used for estimating an individualized probability of a specific endpoint within a defined period of time.[11] The large number of prediction models for TB outcomes published in recent years highlights a common desire to identify TB patients at greatest risk of an unsuccessful treatment outcome in order to tailor treatment strategies and promote cure. However, to date, there has not been a formal synthesis or quality assessment of existing prediction models for TB treatment outcomes, which is essential to determine which models should inform clinical practice. This could also guide development of future models. Thus, we conducted a systematic review to identify, describe, compare, and synthesize clinical prediction models designed to predict TB treatment outcomes among persons with pulmonary TB.

15 WILL PULLER

METHODS AND ANALYSIS

All steps of the systematic review were carried out according to guidelines set by Cochrane Prognosis Methods Group (PMG) and PROGnosis RESearch Strategy (PROGRESS).[12–14] Reporting adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (**Supplemental File 1**). This study was pre-registered on Open Science Framework (OSF) (https://osf.io/rz3wp) and the international prospective register of systematic reviews (PROSPERO; CRD42020155782).

Study eligibility criteria

The review question was defined according to the PICOTS (Population, Intervention, Comaparator, Outcomes, Timing, Setting) framework (**Supplemental File 2**). In brief, the goal was to identify prognostic models developed to predict TB treatment outcomes among pulmonary TB cases. The main outcome was unsuccessful TB treatment outcome, defined by the WHO as the combination of death, treatment failure, loss to follow-up, and/or not evaluated, as compared to successful TB treatment outcome, defined as the combination of cure or treatment completion (**Table 1**) [15]. Loss to follow-up was sometimes referred to as default or treatment abandonment.

Inclusion criteria were: 1) prognostic model studies with or without external validation[16]; 2) study population included adult, drug-susceptible, pulmonary, TB cases; 3) written in English, Spanish, Portuguese, and French; 4) published between January 1, 1995 and January 9, 2020; 5) treatment outcome was one of the following: cure, treatment completion, death, treatment failure, loss to follow-up, or not evaluated.

Exclusion criteria were: 1) predictive value of more than one variable was evaluated but not combined in a prediction model; 2) study population was only multi-drug resistant (MDR) TB cases, only extrapulmonary TB cases, or only children (< 18 years-old); 3) outcome was

BMJ Open

evaluated during treatment such as: two-month smear/culture conversion, acquired resistance, adverse events, quality of life; 4) long-term outcomes, such as relapse, recurrence, or post-treatment mortality.

The decision to include only articles in English, Spanish, Portuguese, and French was based on study team capabilities. The dates reflect modern TB treatment practice; first-line TB treatment regimens were not available until the early 1990s.[17,18] Articles that included a combination of drug-susceptible and drug-resistant cases, or a combination of children and adults were included.

Search strategy and selection criteria

The following electronic databases were searched on January 9, 2020: PubMed, Embase, Web of Science, and the first 200 references from Google Scholar. This combination of databases achieved best overall recall for systematic reviews in a recent study.[19] Clinicaltrials.gov and retractiondatabase.org were also searched for unpublished research. Reference lists of retrieved articles were checked to identify eligible studies.

Search terms relating to the "prediction model" component of the search were adapted from a PubMed search strategy that captured prediction model studies with sensitivity of 98%.[20] That component was combined with terms relating to TB treatment outcomes. The search strategy, developed in PubMed, was adapted for all other databases with assistance from a reference librarian (**Supplemental File 3**).

Article selection was conducted in three stages. The first stage was automatic deduplication and title screening, carried out using *revtools* in RStudio (version 1.2).[21] Remaining articles were imported into Covidence, a web-based software platform that streamlines systematic reviews, where abstracts (Stage 2) and full text (Stage 3) were manually screened.[22] Stages 2 and 3 were carried out by two independent reviewers (LSP and FMR).

Discordance was discussed between reviewers, and if consensus was not reached, a third party arbitrated (one of TRS, VCR, PFR, DL). In stage 3, reasons for exclusion were documented according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Data analysis

Data from selected studies were recorded using a database designed in REDCap (Vanderbilt University).[23,24] Data extraction was informed by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the Prediction Model Risk of Bias Assessment Tool (PROBAST).[16,25,26] CHARMS checklist and PROBAST are in **Supplemental Files 4 and 5**, respectively.

Quality assessment and applicability of included studies was assessed using PROBAST by dual independent review.[16,26] PROBAST was specifically designed to assess risk of bias of prediction model studies, which included identifying deficiencies in study design, conduct, or analysis that led to inaccurate estimates of predictive performance. PROBAST has 4 domains: participants, predictors, outcome, and analysis with 20 total signaling questions. Each question was answered on the scale: yes, probably yes, no, probably no, no information. Domains were scored as low, high, and unclear risk of bias. PROBAST also guides assessment of applicability of participants, predictors, and outcomes from each included study to the review question.

Results were summarized narratively and in tables and figures. Meta-analysis was not possible due to lack of external validation and use of disparate predictors, outcome definitions, and modeling methods. For studies that presented multiple models with the same set of predictors and outcomes, but different methods, the best-performing method was included in data synthesis. For studies presenting multiple models with different sets of predictors (i.e. baseline data vs. longitudinal data), the model developed using only baseline data was included. If studies

2	
3	
4	
5	
6	
7	
8	
q	
10	
10	
11	
12	
13	
14	
15	
16	
17	
18	
10	
20	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
27	
22	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
11	
44	
45	
40	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
50	
5/	
58	
59	
60	

developed multiple models for different outcomes or with different populations, all models were included.

Patient and public involvement

Neither patients nor the public were involved in the design, conduct, or reporting of the research, as it was not feasible or appropriate for this systematic review. The study protocol is publicly available at <u>https://osf.io/rz3wp</u>.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Study selection

The search identified 14,739 unique studies. After excluding irrelevant titles, 6,426 abstracts were screened, 536 articles underwent full-text review, and 33 model development studies presenting 37 prediction models were included (**Figure 1**).

Study characteristics

Of the 33 studies, most were retrospective cohorts (n=25, 76%), three (9%) were prospective cohort studies, two (6%) were case-control studies, and three (9%) were nested case-control studies. Data from nearly half of studies (n=16, 48%) were collected from surveillance systems; eleven (33%) studies used a data collection form developed specifically for their study and six studies (18%) extracted data from medical records. Median sample size was 803 (interquartile range (IQR): 291-4167). Full details on included studies are in **Table 2**.

Thirteen studies (41%) took place in Asia, eight (25%) in Africa, six (19%) in Europe, four (12%) in North America, and one (3%) included sites in Europe and Argentina. Fewer than half (n=14, 45%) of the studies took place in high-burden TB settings.¹ One study did not report study location. (**Tables 2 and 3**).

Reporting of population characteristics varied by study (**Table 4**). Among 18 studies that reported a measure of central tendency (mean or median) for age, the median of those measures of central tendency was 41 years (IQR: 37-49). Of 17 studies that reported the minimum age of participants, seven (41%) had a minimum age of 15, one (6%) had a minimum age of 16, one (6%) had a minimum age of 17, and the remainder had minimum age of 18. Eighteen studies reported including persons living with HIV (PLWH); 5 of these included only TB/HIV patients. Twelve studies reported including persons with diabetes; one of which includes only TB/DM.

BMJ Open

Eight studies reported including some participants with MDR, though prevalence of MDR was low in all studies. Ten studies included only hospitalized patients, and in 14 studies, all participants were on directly observed therapy (DOT).

Model characteristics

Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=8, 23%) (**Tables 2 and 5**). The complete outcome definition for all included studies is in **Supplemental File 6**.

Most models were developed using clinical/epidemiologic predictors (n=34, 92%), two (6%) used multiple biomarkers, and one (3%) used adherence data. The most common candidate predictors were age, sex, extrapulmonary TB, smear result, BMI, x-ray findings, and previous TB. The most common predictors retained in the final models were age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV (**Figure 2**).

Only three models (8%) used survival analysis; most models used logistic regression (n=29, 78%) and five (14%) used a machine learning approach. More than half of studies (n=19, 51%) considered variables for inclusion in the multivariable model based on unadjusted associations with the outcome. Model building methods varied widely between models (**Table 5**).

Only 19 (51%) models were internally validated, including ten (53%) split-sample validation, five (26%) bootstrap resampling, and four (21%) cross-validation. Six (16%) models were externally validated.

Many models (n=30, 81%) reported discrimination with c-statistic (concordance statistic) or area under the receiver operating characteristic (AUROC), which are equivalent and quantify the ability of the model to distinguish between patients who do and do not develop an outcome.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Only 17 (46%) reported calibration, the agreement between observed and predicted outcomes. Most studies assessed calibration with Hosmer-Lemeshow tests (n=13, 77%); only two studies provided a calibration plot, the preferred reporting method for prediction model studies,[16,27,28] and one reported the calibration slope (**Table 2**). Models were presented a variety of ways, the most common of which was a weighted risk score (n=16, 43%); details on model presentation are in **Supplemental File 7**.

Quality assessment

Grading of PROBAST signaling questions is summarized in **Figure 3**, and the summary risk of bias for the participants, predictors, outcome, and analysis domains and assessment of applicability are shown in **Figure 4**. More than half of the studies were at low risk of bias for the population and outcomes domains, but all studies were at high risk of bias in the analysis domain.

Common sources of population bias included use of non-nested case-control design[29,30], nested case-control design without proper estimation of baseline risk,[31,32] or inappropriate inclusion/exclusion criteria.[33,34] Sources of predictor bias included lack of standardized assessment of key predictors (i.e. HIV, diabetes, chest x-ray scoring)[9,29,31,34–36] or timing of data collection/availability that would limit the intended use of the model.[9,29,37] Within the outcomes domain, sources of bias included subjective[35] or non-standard[32,38] outcome measures and inconsistent outcome ascertainment.[29]

Bias in the analysis domain was widespread. More than half of the models included were likely overfit due to low events per variable (EPV) ratios (**Table 5**). Only 6 studies handled continuous and categorical variables appropriately (i.e., didn't dichotomize continuous variables, considered non-linearity of continuous variables).[31,39–43] Most studies used complete case-

BMJ Open

analysis or did not mention missing data; no study used multiple imputation in their main analysis. One study with low amounts of missing data (<5%) conducted sensitivity analysis with multiple imputation.[44] A different study excluded only two people out of a total sample size of 1007 with missing data, which would have little impact on model performance.[45] Fewer than half (n=14) of studies avoided univariable predictor selection, and only three studies used survival analysis, appropriately accounting for censoring.[36,45,46] Performance measures were appropriately reported (i.e. calibration assessed with plot and discrimination assessed with cstatistic/AUROC) in three studies.[41,44,47] Only two studies estimated optimism (degree to which data are overfit) or accounted for potential overfitting with penalization of model parameters.[35,41] Ten studies appropriately presented their model with model coefficients or nomograms, which prevents bias from rounding or transforming model coefficients to generate a risk score.[30,33,35,37,38,45,47–55]

About half of the models (n=19, 51%) were applicable to the review question in all domains. However, unclear reporting of target population or predictor and outcome definitions limited assessment of applicability for several studies.[38,49,50,56,57] Additionally, studies that included only hospitalized patients with specific laboratory parameters may not be routinely available in the clinical setting.[39,40,42]

DISCUSSION

In this comprehensive, systematic review of prediction models for pulmonary TB treatment outcomes, we identified 33 model development studies presenting 37 prediction models. Although diagnostic prediction models for prevalent TB were previously systematically reviewed, this is the first systematic review of TB treatment outcomes.[58] The included prediction models were developed for predicting death, treatment failure, default, or a composite unfavorable outcome during TB treatment. Most models reported good performance (c-statistic/AUROC>0.7), but all were evaluated to have high risk of bias due to poor reporting, exclusion of missing data, weak methodologic approaches, lack of calibration assessment, and limited validation. Predictor and outcome definitions varied by study and limited comparisons between models.

More than half of the models included in the review were developed in low burden TB settings, and none were developed specifically in South America. Prediction of TB treatment outcome is especially important in high burden TB settings, where resources may be limited, and risk assessment can guide resource allocation toward patients who need the most involved care protocols.

Common risk factors included in the models were consistent with well-established risk factors for poor TB treatment outcomes, including age, sex, HIV, extrapulmonary TB, baseline smear results, and previous TB treatment. Among studies that included PLWH, only three considered factors related to management/severity of HIV, such as receipt of antiretroviral therapy, CD4 cell count, or viral load, which likely impact TB treatment outcomes.[40,46,51] Laboratory values or metabolic biomarkers, such as hemoglobin, hemoglobin A1c or random blood glucose, may also be associated with treatment outcome and worth considering as candidate predictors. There is increasing evidence that diabetes impacts TB treatment outcomes,

Page 17 of 62

BMJ Open

but caution is warranted about how to best define diabetes in the context of a prediction model to ensure consistency and reproducibility across studies.[59] Behavioral characteristics, such as tobacco use, alcohol use, and drug use were rarely included in final prediction models and are difficult to collect objectively, suggesting their role in prediction models for TB treatment outcomes may be limited.

Additionally, several studies excluded participants with HIV, diabetes, extrapulmonary TB, or MDR TB, because these factors negatively influence treatment outcomes. However, careful consideration should be given to inclusion/exclusion criteria in prediction model studies. Information necessary to carry out inclusion/exclusions should be available at the of intended use of the model, which may not always hold for these aforementioned factors.[60] This point is especially questionable for MDR, given that conventional drug-susceptibility testing results are not available for several weeks after TB diagnosis; though more recent advances in rapid molecular methods such as GeneXpert or line-probe assays offer rapid screening for drug resistance.[61]

TB researchers should thoughtfully consider how to appropriately handle complexities of censoring and competing risks in TB outcomes research. Only three studies in this review used survival analysis, despite the long duration of TB treatment outcome assessment and relatively high rates of losses to follow-up across studies. Losses to follow-up were frequently excluded, which can lead to selection bias. Additionally, all studies that included death as the outcome considered all-cause mortality. Also, for studies that predict losses to follow-up/default, death (even due to TB) is a competing risk. Competing risk analyses are common in cardiovascular research, research in elderly populations, and there are specific recommendations for competing risk methods in prognostic research.[62,63]

Though all included studies were at high risk of bias in the analysis domain, we want to highlight two studies with some exemplary characteristics.[41,44] Pefura-Yone et al.[41] provide clear explanations of study design, inclusion/exclusion criteria, and data collection procedures; TB diagnosis and treatment outcome definitions were standard.[64] Non-linearity of continuous variables was considered with restricted cubic splines, and no continuous variables were categorized or dichotomized; the final model includes four predictors that are easy to collect and routinely assessed in most TB control programs, especially those in high burden settings. The performance of the model was internally validated with bootstrap validation, and the discrimination (c-statistic=0.808) was corrected for optimism. Model calibration was presented graphically with calibration plots. The final model was presented as a nomogram with instructions for use, which facilitates use in external validation studies. Gupta-Wright and colleagues developed and externally validated a clinical risk score to predict mortality in highburden, low-resource settings.⁴³ They used clinical trial data with very low amounts of missing data for model development, and externally validated the clinical risk score with data collected independently from two other studies (a clinical trial and a prospective cohort). Given high amounts (42%) of missing data in the validation cohort, they conducted sensitivity analysis using multiple imputation for missing data; the c-statistic differed slightly between complete case and multiple-imputation analyses in the validation cohort (0.68 vs. 0.64). Candidate predictors were based on *a priori* clinical knowledge, previous literature, and required variables were objective, reproducible, and available in low-resource settings, consistent with recommended approaches. [26,60,65] Additionally, they reported model performance with the c-statistics and calibration plots for development and validation cohorts, and reported results according to TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or

diagnosis) guidance.[27,28] Regardless, each of these models requires external validation prior to use in clinical practice.

There are several limitations of this study. First, data extraction was subject to reporting the primary study, which varied widely across studies. Most studies reported discrimination, and several reported sensitivity and specificity; TRIPOD recommends all studies report, at minimum, calibration with a calibration plot and discrimination with c-statistic.[28] Measures of sensitivity and specificity require dichotomization of risks, which then only pertain to a specific risk stratum, rather than quantifying the overall model performance. [14,65] We did not include external validation studies, which is an essential step for translation to clinical practice. However, several studies in the review did not include the full model equation, which inhibits their ability to be externally validated. Upon searching for studies that externally validated prediction models in this review, we found three studies [66–68] that evaluated the same model (TBscore).[36] Briefly, these studies evaluated the ability of TBscore to monitor treatment response in a new setting[66], refined the instrument (TBccoreII) using exploratory factor analysis[67], and then evaluated TBscoreII for use in patients with TB/HIV.[68] To our knowledge, no other studies included in the review were externally validated by other sources. Finally, we excluded 10 studies that were not available in English, Spanish, Portuguese, or French; all abstracts were available in English, and none reported model performance metrics, so they likely would have been excluded for different reasons regardless.

The findings of this review not only serve as a comprehensive overview of existing TB outcome prediction models but can act as a resource for future model development and validation of prediction models for TB treatment outcomes. We encourage researchers to focus future TB outcome prediction models on easily collected and readily available predictors that are

widely generalizable. We highlight age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV as common predictors of TB treatment outcomes. Additionally, when building a new prediction model, it is recommended to first prune the set of considered predictors based on expert opinion and previous literature, rather than univariable analysis or variable selection processes[26,60,65] Future model development or validation studies should adhere to the TRIPOD guidelines, which provide a 22-item checklist and aims to improve the reporting of prediction model development studies.[27,28] We also encourage researchers consider the PROBAST criteria when developing their model to limit sources of bias in design and conduct of prediction model studies.

Prediction models are an important tool in TB management, as they can lay the foundation for future intervention studies or clinical decision making by providing risk prediction that can aid in targeted treatment, resource allocation, or intensive case management at patients who are least likely to achieve cure and most likely to benefit from some form of intervention, especially in high-burden and low-resources areas. Use of prediction models can potentially help guide tuberculosis treatment practices to achieve the End TB Strategy target of >90% treatment success, but methodologic rigor and detailed reporting must be improved. Though our findings suggest that none of the existing models are ready for clinical application without extensive external validation, we hope they direct future researchers to make use of guidelines for development and reporting of prediction models.

BMJ Open

FOOTNOTES

Ethics approval: Not required.

Transparency statement: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported. No important aspects of the study have been omitted, and any discrepancies from the study as planned were explained.

Contributorship: LSP conceptualized the research question, designed the protocol, and drafted the manuscript. LSP and FMR screened studies. FMR, PFR, DL, VCR and TRS provided feedback on the research design, original protocol, and revised successive drafts of the manuscript. All authors approved the final version of the manuscript.

Funding: This work was supported by the National Center for Advancing Translational Sciences [CTSA Award No. TL1TR000447 to L.S.P.] and the National Institutes of Allergy and Infectious Diseases [F31AI152614-01A1 to L.S.P],. Its contents are solely the responsibility of the authors and do not necessarily represent the official views the National Center for Advancing Translational Sciences or the National Institutes of Health.

Competing interests: None declared.

Data sharing: The study protocol is available online at https://osf.io/rz3wp. Most included studies are publicly available. Additional data and code are available upon request.

Exclusive license: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.

REFERENCES

- 1 Global Tuberculosis Report 2019. Geneva: World Health Organization 2019. Licence: CC BY-NC-SA 3.0 IGO.
- 2 The END TB Strategy. Geneva: World Health Organization 2015.
- 3 Kerantzas CA, Jacobs WR. Origins of Combination Therapy for Tuberculosis: Lessons for Future Antimicrobial Development and Application. *mBio* 2017;**8**:e01586-16. doi:10.1128/MBIO.01586-16
- 4 Nahid P, Dorman SE, Alipanah N, *et al.* Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clinical Infectious Diseases* 2016;**63**:e147–95. doi:10.1093/cid/ciw376
- 5 Guildelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. Geneva: World Health Organization 2017. Licence: CC BY-NC-SA 3.0 IGO. doi:WHO/HTM/TB/2017.05
- 6 WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: World Health Organization 2019. Licence: CC BY-NC-SA 3.0 IGO.
- 7 Vasankari T, Holmström P, Ollgren J, et al. Risk factors for poor tuberculosis treatment outcome in Finland: A cohort study. BMC Public Health 2007;7:1–9. doi:10.1186/1471-2458-7-291
- 8 Ramachandran G, Agibothu K. Factors Influencing Tuberculosis Treatment Outcome in Adult Patients Treated with Thrice-Weekly Regimens. 2017;**61**:1–12.
- 9 Abdelbary BE, Garcia-Viveros M, Ramirez-Oropesa H, et al. Predicting treatment failure, death and drug resistance using a computed risk score among newly diagnosed TB patients in Tamaulipas, Mexico. *Epidemiology and Infection* 2017;145:3020–34. doi:10.1017/S0950268817001911
- 10 Torres NMC, Rodríguez JJQ, Andrade PSP, *et al.* Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis. *PLoS ONE* 2019;14:1–24. doi:10.1371/journal.pone.0226507
- 11 Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Medicine 2013;10:e1001381. doi:10.1371/journal.pmed.1001381
- 12 Riley R, Ridley G, Williams K, *et al.* Prognosis research: towards evidence-based results and a Cochrane methods group. 2014;**60**:863–5.
- 13 Moons KG, Hooft L, Williams K, *et al.* Implementing systematic reviews of prognosis studies in Cochrane. *The Cochrane database of systematic reviews* 2018;**10**:ED000129. doi:10.1002/14651858.ED000129

2	
3	
4	
5	
6	
7	
, 8	
0	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
10	
19	
20	
21	
22	
23	
24	
25	
26	
27	
22	
20	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
30	
70	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

- 14 Debray TPA, Damen JAAG, Snell KIE, *et al.* A guide to systematic review and metaanalysis of prediction model performance. *BMJ (Online)* 2017;**356**. doi:10.1136/bmj.i6460
- 15 Definitions and reporting framework for tuberculosis 2013 revision. In: *Annex 2, TB case and treatment outcome definitions*. Geneva: World Health Organization 2014.
- 16 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**:51–8. doi:10.7326/M18-1376
- 17 Iseman MD. Tuberculosis therapy: past, present and future. *Eur Resp J* 2002;**20**:87s–94s. doi:10.1183/09031936.02.00309102
- 18 Council STSMR. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis: the results up to 30 months. *Tubercle* 1981;:95– 102.
- 19 Bramer WM, Rethlefsen ML, Kleijnen J, *et al.* Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews* 2017;6:1–12. doi:10.1186/s13643-017-0644-y
- 20 Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS ONE 2012;7:3–8. doi:10.1371/journal.pone.0032844
- 21 Westgate MJ. revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods* 2019;**10**:606–14. doi:10.1002/jrsm.1374
- 22 Veritas Health Innovation, Melbourne A. Covidence systematic review sofware. Covidence. 2016. doi:10.1016/j.carbon.2012.08.062
- 23 Harris PA, Taylor R, Minor BL, *et al.* The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019;**95**:103208. doi:10.1016/j.jbi.2019.103208
- 24 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009;**42**:377–81. doi:10.1016/j.jbi.2008.08.010
- 25 Moons KGM, de Groot JAH, Bouwmeester W, *et al.* Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine* 2014;**11**. doi:10.1371/journal.pmed.1001744
- 26 Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Annals of Internal Medicine* 2019;**170**:W1–33. doi:10.7326/M18-1377

27 Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine* 2015;**162**:W1–73. doi:10.7326/M14-0698

- 28 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. 2015;**162**. doi:10.7326/M14-0697
- 29 Cherkaoui I, Sabouni R, Ghali I, *et al.* Treatment default amongst patients with tuberculosis in urban Morocco: Predicting and explaining default and post-default sputum smear and drug susceptibility results. *PLoS ONE* 2014;9. doi:10.1371/journal.pone.0093574
- 30 Keane VP, De Klerk N, Krieng T, et al. Risk factors for the development of non-response to first-line treatment for tuberculosis in Southern Vietnam. International Journal of Epidemiology 1997;26:1115–20. doi:10.1093/ije/26.5.1115
- 31 Chang KC, Leung CC, Tam CM. Risk factors for defaulting from anti-tuberculosis treatment under directly observed treatment in Hong Kong. *International Journal of Tuberculosis and Lung Disease* 2004;**8**:1492–8.
- 32 Chee CBE, Boudville IC, Chan SP, *et al.* Patient and disease characteristics, and outcome of treatment defaulters from the Singapore TB control unit A one-year retrospective survey. *International Journal of Tuberculosis and Lung Disease* 2000;**4**:496–503. doi:NA
- 33 Luies L, Reenen M Van, Ronacher K, *et al.* Predicting tuberculosis treatment outcome using metabolomics. *Biomarkers in Medicine* 2017;**11**:1057–67. doi:10.2217/bmm-2017-0133
- 34 Killian JA, Wilder B, Sharma A, *et al.* Learning to Prescribe Interventions for Tuberculosis Patients Using Digital Adherence Data. *KNOWLEDGE DISCOVERY AND DATA MINING* 2019;NA:2430–8. doi:10.1145/3292500.3330777
- 35 Belilovsky EM, Borisov SE, Cook EF, *et al.* Treatment interruptions among patients with tuberculosis in Russian TB hospitals. *International Journal of Infectious Diseases* 2010;**14**:e698–703. doi:10.1016/j.ijid.2010.03.001
- 36 Wejse C, Gustafson P, Nielsen J, *et al.* TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. *Scandinavian Journal of Infectious Diseases* 2008;**40**:111–20. doi:10.1080/00365540701558698
- 37 Nguyen DT, Graviss EA. Development and validation of a risk score to predict mortality during TB treatment in patients with TB-diabetes comorbidity. *BMC Infectious Diseases* 2019;**19**:10. doi:10.1186/s12879-018-3632-5
- 38 Kalhori SRN, Zeng X. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. *Lecture Notes in Engineering and Computer Science* 2009;**2179**:774–8.

ว
2
3
4
5
6
7
/
8
9
10
11
12
12
1.0
14
15
16
17
18
19
20
∠∪ ⊃1
21
22
23
24
25
26
20
27
20
29
30
31
32
33
34
25
22
30
37
38
39
40
41
42
43
11
 / -
45
46
47
48
49
50
51
57
52
22
54
55
56
57
58
59
60
00

- 39 Horita N, Miyazawa N, Yoshiyama T, *et al.* Poor performance status is a strong predictor for death in patients with smear-positive pulmonary TB admitted to two Japanese hospitals. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2013;**107**:451–6. doi:10.1093/trstmh/trt037
- 40 Koegelenberg CFN, Balkema CA, Jooste Y, *et al.* Validation of a severity-of-illness score in patients with tuberculosis requiring intensive care unit admission. *South African Medical Journal* 2015;**105**:389–92. doi:10.7196/SAMJ.9148
- 41 Pefura-Yone EW, Kuaban C, Assamba-Mpom SA, *et al.* Derivation, validation and comparative performance of a simplified chest X-ray score for assessing the severity and outcome of pulmonary tuberculosis. *Clinical Respiratory Journal* 2015;**9**:157–64. doi:10.1111/crj.12112
- 42 Valade S, Raskine L, Aout M, *et al.* Tuberculosis in the intensive care unit: A retrospective descriptive cohort study with determination of a predictive fatality score. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2012;**23**:173–8. doi:10.1155/2012/361292
- 43 Wang Q, Han W, Niu J, *et al.* Prognostic value of serum macrophage migration inhibitory factor levels in pulmonary tuberculosis. *Respiratory Research* 2019;**20**:50. doi:10.1186/s12931-019-1004-3
- 44 Gupta-Wright A, Corbett EL, Wilson D, *et al.* Risk score for predicting mortality including urine lipoarabinomannan detection in hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation and external validation cohort study. *PLoS Medicine* 2019;**16**:1–20. doi:10.1371/journal.pmed.1002776
- 45 Zhang Z, Xu L, Pang X, *et al.* A Clinical scoring model to predict mortality in HIV/TB coinfected patients at end stage of AIDS in China: An observational cohort study. *BioScience Trends* 2019;**13**:136–44. doi:10.5582/bst.2018.01309
- 46 Podlekareva DN, Grint D, Post FA, *et al.* Health care index score and risk of death following tuberculosis diagnosis in HIV-positive patients. *International Journal of Tuberculosis and Lung Disease* 2013;**17**:198-206+i. doi:10.5588/ijtld.12.0224
- 47 Baussano I, Pivetta E, Vizzini L, *et al.* Predicting tuberculosis treatment outcome in a lowincidence area. *International Journal of Tuberculosis and Lung Disease* 2008;**12**:1441–8.
- 48 Costa-Veiga A, Briz T, Nunes C. Unsuccessful treatment in pulmonary tuberculosis: Factors and a consequent predictive model. *European Journal of Public Health* 2018;**28**:252–8. doi:10.1093/eurpub/ckx136
- 49 Niakan Kalhori SR, Nasehi M, Zeng XJ. A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion. *IAENG International Journal of Applied Mathematics* 2010;**40**:1–6.
- 50 Kalhori SRN, Zeng X-J. PREDICTING THE OUTCOME OF TUBERCULOSIS TREATMENT COURSE IN FRAME OF DOTS - From Demographic Data to Logistic

Regression Model. In: *Proceedings of the International Conference on Health Informatics*. SciTePress - Science and and Technology Publications 2009. 129–34. doi:10.5220/0001431401290134

- 51 Madan C, Chopra KK, Satyanarayana S, *et al.* Developing a model to predict unfavourable treatment outcomes in patients with tuberculosis and human immunodeficiency virus co-infection in Delhi, India. *PLoS ONE* 2018;**13**:e0204982. doi:10.1371/journal.pone.0204982
- 52 Nguyen DT, Jenkins HE, Graviss EA. Prognostic score to predict mortality during TB treatment in TB / HIV co-infected patients. *PLoS ONE* 2018;**13**:1–12. doi:10.1371/journal.pone.0196022
- 53 Nguyen DT, Graviss EA. Development and validation of a prognostic score to predict tuberculosis mortality. *Journal of Infection* 2018;77:283–90. doi:10.1016/j.jinf.2018.02.009
- 54 Pefura-Yone EW, Balkissou AD, Poka-Mayap V, *et al.* Development and validation of a prognostic score during tuberculosis treatment. *BMC Infectious Diseases* 2017;**17**:1–9. doi:10.1186/s12879-017-2309-9
- 55 Rodrigo T, Caylà JA, Casals M, *et al.* A predictive scoring instrument for tuberculosis lost to follow-up outcome. *Respiratory Research* 2012;**13**:1–9. doi:10.1186/1465-9921-13-75
- 56 Hussain OA, Junejo KN. Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Informatics for Health and Social Care* 2019;**44**:135–51. doi:10.1080/17538157.2018.1433676
- 57 Sauer CM, Sasson D, Paik KE, *et al.* Feature selection and prediction of treatment failure in tuberculosis. *PLoS ONE* 2018;**13**:1–14. doi:10.1371/journal.pone.0207491
- 58 Wyk SSV, Lin HH, Claassens MM. A systematic review of prediction models for prevalent pulmonary tuberculosis in adults. *Int J Tuberc Lung Dis*;**21**.
- 59 Huangfu P, Ugarte-Gil C, Golub J, *et al.* The effects of diabetes on tuberculosis treatment outcomes: an updated systematic review and meta-analysis. *The International Journal of Tuberculosis and Lung Disease* 2019;**23**:783–96. doi:10.5588/ijtld.18.0433
- 60 Steyerberg EW. *Clinical Prediction Models*. New York, NY: : Springer New York 2009. doi:10.1007/978-0-387-77244-8
- 61 Sharma SK, Dheda K. What is new in the WHO consolidated guidelines on drug-resistant tuberculosis treatment? The Indian journal of medical research. 2019;**149**:309–12. doi:10.4103/ijmr.IJMR_579_19
- 62 Wolbers M, Koller MT, Witteman JCM, *et al.* Prognostic models with competing risks methods and application to coronary risk prediction. *Epidemiology* 2009;**20**:555–61. doi:10.1097/EDE.0b013e3181a39056

- 63 Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016;**133**:601–9. doi:10.1161/CIRCULATIONAHA.115.017719
- 64 National Tuberculosis Control Program. Manual for health personnel. Yaounde: 2012.
- 65 Royston P, Moons KGM, Altman DG, *et al.* Prognosis and prognostic research: Developing a prognostic model. *BMJ (Online)* 2009;**338**:1373–7. doi:10.1136/bmj.b604
- 66 Janols H, Abate E, Idh J, *et al.* Early treatment response evaluated by a clinical scoring system correlates with the prognosis of pulmonary tuberculosis patients in Ethiopia: A prospective follow-up study. *Scandinavian Journal of Infectious Diseases* 2012;44:828–34. doi:10.3109/00365548.2012.694468
- 67 Rudolf F, Lemvik G, Abate E, *et al.* TBscore II: Refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. *Scandinavian Journal of Infectious Diseases* 2013;**45**:825–36. doi:10.3109/00365548.2013.826876
- 68 Wejse C, Patsche CB, Kühle A, *et al.* Impact of HIV-1, HIV-2, and HIV-1+2 dual infection on the outcome of tuberculosis. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2015;**32**:128–34. doi:10.1016/j.ijid.2014.12.015
- 69 Aljohaney AA. Mortality of patients hospitalized for active tuberculosis in King Abdulaziz University Hospital, Jeddah, Saudi Arabia. *Saudi Medical Journal* 2018;**39**:267–72. doi:10.15537/smj.2018.3.22280
- 70 Bastos HN, Osório NS, Castro AG, *et al.* A prediction rule to stratify mortality risk of patients with pulmonary tuberculosis. *PLoS ONE* 2016;**11**:1–14. doi:10.1371/journal.pone.0162797
- 71 Gupta-Wright A, Corbett EL, Wilson D, *et al.* Risk score for predicting mortality including urine lipoarabinomannan detection in hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation and external validation cohort study. *PLoS Medicine* 2019;**16**:1–20. doi:10.1371/journal.pmed.1002776
- 72 Horita N, Miyazawa N, Yoshiyama T, *et al.* Development and validation of a tuberculosis prognostic score for smear-positive in-patients in Japan. *International Journal of Tuberculosis and Lung Disease* 2013;**17**:54–60. doi:10.5588/ijtld.12.0476
- 73 Podlekareva DN, Grint D, Post FA, *et al.* Health care index score and risk of death following tuberculosis diagnosis in HIV-positive patients. *The International Journal of Tuberculosis and Lung Disease* 2013;**17**:198–206. doi:10.5588/ijtld.12.0224
- 74 Wang Q, Han W, Niu J, *et al.* Prognostic value of serum macrophage migration inhibitory factor levels in pulmonary tuberculosis. *Respiratory Research* 2019;**20**:50. doi:10.1186/s12931-019-1004-3

75 Wejse C, Gustafson P, Nielsen J, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. Scandinavian Journal of Infectious Diseases 2008;40:111–20. doi:10.1080/00365540701558698

- 76 Mburu JW, Kingwara L, Ester M, et al. Use of classification and regression tree (CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor tuberculosis treatment outcomes and associated risk factors. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* 2018;11:10–6. doi:10.1016/j.jctube.2018.01.002
- 77 Thompson EG, Du Y, Malherbe ST, *et al.* Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis* 2017;**107**:48–58. doi:10.1016/j.tube.2017.08.004
- 78 Chee CBE, Boudville IC, Chan SP, *et al.* Patient and disease characteristics, and outcome of treatment defaulters from the Singapore TB control unit A one-year retrospective survey. *International Journal of Tuberculosis and Lung Disease* 2000;4:496–503.
- 79 Rodrigo T, Caylà JA, Casals M, *et al.* A predictive scoring instrument for tuberculosis lost to follow-up outcome. *Respiratory Research* 2012;**13**:1–9. doi:10.1186/1465-9921-13-75
- 80 Kalhori SRN, Zeng X-J. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. In: *Lecture Notes in Engineering and Computer Science*. NA 2009. 774–8. doi:NA

1	
2	
2	
7	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
20	
21 22	
∠∠ วว	
∠⊃ ⊃4	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
- + ⊿5	
رب ۸۸	
40 17	
4/ 40	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	

Table 1. World Health Organization definition of treatment outcomes for T	B patients
---	------------

Outcome	Definition
	Completion of treatment without evidence of failure, but without
Treatment	documentation of a negative sputum smear or culture in the last month of
completion	treatment and/or on at least one previous occasion, either because tests
	were not done or because results are unavailable
Cure	Bacteriologic confirmation of a negative smear or culture at the end of TB
	treatment and on at least one previous occasion
Treatment success	Composite of cured and treatment completed
Treatment failure	Sputum smear or culture is positive at month 5 or later during treatment
Death	TB patient who dies for any reason before starting or during the course of
Douin	treatment
Loss to follow-up	TB patient who did not start treatment or whose treatment was interrupted
Loss to follow up	for 2 consecutive months or more
Not evaluated	TB patient for whom no treatment outcome was assigned, which includes
	cases who "transferred out" to another treatment unit as well as cases for
(transfer out)	whom the treatment outcome is unknown to the reporting unit

Figure 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow chart of inclusion process



tor ocer texics only

Table 2. Study characteristics

And-temport Periodicion Stady scars Stady design Location Number of sample size sample size (%) Periodicion in final model Periodicion final model Periodicion in final mod		. Study charact									
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	2 3 4 First author, year 5	Population	Study years	Study design	Location	Validation	No. with outcome / sample size (%)	Predictors in final model	Performance measures	Model presentation	Risk of bias (population predictor, outcome, analysis)
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	5 <u>Death</u>										
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	7 3 Abdelbary[9] / 2017	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 261/4216 (6%) Validation: 260/4215 (6%)	Age (<41, 41-65, ≥65), sex, MDR, HIV, malnutrition, alcoholism, diabetes, pulmonary TB	c-statistic = 0.70 Sensitivity = 60% Specificity = 71%	Risk score	Low, High, Low, High
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	10 Abdelbary[9] / 11 2017 (TB-DM) 12	TB-DM cases	2006 - 2013	Retrospective cohort	Mexico	None	88/2121 (4%)	Sex, malnutrition, BCG vaccinated, AFB smear (positive vs. negative)	c-statistic = 0.68	Risk score	Unclear, High, Low, High
$\frac{1}{7} [Jestics] [0] / 2016 augustent TB cases augustent TB cagesting tages augustent TB cases augustent $	13 _{Aljohaney[69]} / 14 2018	Hospitalized TB patients	Dec 2011 – Dec 2016	Retrospective cohort	Saudi Arabia	None	41/291 (14%)	<u>Clinical model:</u> Age, congestive heart failure <u>Clinical + lab model:</u> * Age > 65, congestive heart failure, bilateral disease on chest xray	<u>Clinical model:</u> Accuracy = 86% <u>Clinical & lab model:*</u> Accuracy = 90%	Odds ratios	Unclear, Unclear, Unclear, High
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	15 16 17 ^{Bastos[70] / 2016} 18	Inpatient and outpatient TB cases on DOT	2007 - 2013	Retrospective cohort	Portugal	External (setting)	Development: 121/681 (18%) Validation: 24/103 (23%)	Hypoxemic respiratory failure, age (≥50 vs. <50), bilateral involvement, comorbidities (at least one of HIV, diabetes, liver at least one of: HIV, diabetes, liver failure/cirrhosis, congestive heart failure, chronic respiratory disease), hemoglobin (<12 vs. ≥12)	AUROC = 0.84 (95% CI: 0.76-0.93) Sensitivity = 41.8% Specificity = 92.1%	Risk score	Low, Unclear, Low, High
$ \begin{array}{c} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 $	19 Gupta-Wright[71] / 20 2019	Hospitalized TB- HIV patients	Oct 2015 – Sept 2017	Retrospective cohort	Malawi and South Africa	External (setting)	Development: 94/315 (30%) Validation: 147/644 (23%)	Sex, age 55+, currently taking ART, ability to walk unaided, severe anemia, positive TB-LAM	c-statistic = 0.68 (95% CI: 0.61-0.74) HL test: p=0.13 Calibration plot	Risk score	Low, Low, Low, High
$ \begin{array}{c} 24 \\ 25 \ cogelenberg[40] / \\ 25 \ cogelenberg[40] / \\ 26 \ 2015 \end{array} \\ \begin{array}{c} 36 \ 28 \ 2012 \\ 27 \ (2m-1) \\ 27 \ (2m-1) \\ 27 \ (2m-1) \\ 27 \ (2m-1) \\ 28 \ (2m-1) \\ 2018 \end{array} \\ \begin{array}{c} 3m \ 2012 \\ 2m \ 2$	22 22 Horita[72] / 2013 23	Hospitalized TB patients	Jan 2008 – Jul 2011	Retrospective cohort	Japan	External (setting)	Development: 36/179 (20%) Validation: 48/244 (20%)	Age, oxygen requirement, albumin, activities of daily living	AUROC = 0.893 Sensitivity = 0.92 Specificity = 0.73	Risk score	Low, Low, Low, High
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	24 25 _{xoegelenberg[40]} / 26 ²⁰¹⁵ 27	Hospitalized TB patients	Jan 2012 – May 2013	Retrospective cohort	South Africa	None	38/83 (46%)	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	Mean score in survivors: 2.27 (SD=1.47) Mean score in non- survivors: 3.58 (SD=1.08)	Risk score	Low, Low, Low, High
50 31 Neuven[37] (TB- 2 DM) / 2019TB-DM patientsJan 2010 - Dec 2016Retrospective cohortTexasInternal (bootstrap)112/1227 (9%)Age ≥ 65 , US-born, homeless, IDU, chronic kidney failure, TB meningitis, Miliary TB, AFB positive smear, HIV positive meningitis, Miliary TB, AFB positive smear, HIV positive meningeal TB, abnormal CXR, diagnosis confirmed by positive culture of NAA, culture not meningeal TB, abnormal CXR, HI Est: X=4, 42, 5, p=0.51 HI Est: X=4, 42, p=0.62 Sensitivity 80, 7% Specificity = 68, 2% Calibration plot <td>28 Nguyen[53] 29 (general pop) / 2018</td> <td>TB cases</td> <td>Jan 2010 – Dec 2016</td> <td>Retrospective cohort</td> <td>Texas</td> <td>Internal (split-sample)</td> <td>Development: 253/3378 (7%) Validation: 270/3377 (8%)</td> <td>Age group (15-44, 44-64, >64), US born, homeless, resident of long term care facility, chronic kidney failure, meningeal TB, miliary TB, HIV positive, HIV unknown</td> <td>AUROC = 0.80 (95% CI: 0.77-0.82) HL test:X²=6.3, p=0.613</td> <td>Risk score</td> <td>Low, Unclear, Unclear, High</td>	28 Nguyen[53] 29 (general pop) / 2018	TB cases	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (split-sample)	Development: 253/3378 (7%) Validation: 270/3377 (8%)	Age group (15-44, 44-64, >64), US born, homeless, resident of long term care facility, chronic kidney failure, meningeal TB, miliary TB, HIV positive, HIV unknown	AUROC = 0.80 (95% CI: 0.77-0.82) HL test:X ² =6.3, p=0.613	Risk score	Low, Unclear, Unclear, High
$\frac{33}{34} \underbrace{\operatorname{Nguyen}[52] (TB-HIV patients)}_{HV / 2018} TB-HIV patients} \underbrace{\operatorname{Jan 2010-}_{De c 2016}}_{De c 2016} \underbrace{\operatorname{Retrospective}}_{cohort} \underbrace{\operatorname{Texas}}_{(bootstrap)} \underbrace{\operatorname{Internal}}_{(bootstrap)} \underbrace{57/450 (13\%)}_{57/450 (13\%)} \underbrace{\operatorname{Age} \geq 45, resident of LTCF, meningeal TB, abnormal CXR, dilute not converted or unknown} \underbrace{\operatorname{AUROC} = 0.79}_{(95\% C1 0.70-0.87)} \underbrace{\operatorname{Retrospective}}_{Briter score: 0.09} \underbrace{\operatorname{Retrospective}}_{Briter score: 0.09} \underbrace{\operatorname{Retrospective}}_{Briter score: 0.09} \underbrace{\operatorname{Retrospective}}_{Cameroon} \underbrace{\operatorname{Internal}}_{(bootstrap)} \underbrace{213/2250 (9\%)}_{213/2250 (9\%)} \underbrace{\operatorname{Age} \geq 45, resident of LTCF, meningeal TB, abnormal CXR, dilute not converted or unknown \underbrace{\operatorname{Retrospective}}_{Briter score: 0.09} \underbrace{\operatorname{Retrospective}}_{Cameroon} \underbrace{\operatorname{Retrospective}}_{cohort} \underbrace{\operatorname{Cameroon}}_{(bootstrap)} \underbrace{\operatorname{Internal}}_{(bootstrap)} \underbrace{213/2250 (9\%)}_{213/2250 (9\%)} \underbrace{\operatorname{Age}, adjusted BMI, clinical form (PTB+, PTB-, EPTB), HIV}_{Age, adjusted BMI, clinical form (PTB+, PTB-, EPTB), HIV} \underbrace{\operatorname{Retrospective}}_{Briter score: 0.09} \underbrace{\operatorname{Retrospective}}_{Cohort} \underbrace{\operatorname{S2} cities in}_{Europe and} \operatorname{Retrospective}}_{2013} \underbrace{\operatorname{S2} cities in}_{Briter score: Cohort} \underbrace{\operatorname{Retrospective}}_{Cohort} \operatorname{Retro$	30 31Nguyen[37] (TB- DM) / 2019 32	TB-DM patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	112/1227 (9%)	Age ≥65, US-born, homeless, IDU, chronic kidney failure, TB meningitis, Miliary TB, AFB positive smear, HIV positive	AUROC = 0.82 (95% CI: 0.78-0.87) HL test: X ² =4.54, p=0.81 Brier score=0.07	Risk score	Unclear, Unclear, Unclear, High
$\frac{36}{37} \frac{36}{2017} \frac{36}{2017} \frac{36}{2017} \frac{36}{2017} \frac{38}{2017} \frac{38}{$	33 34 ^{Nguyen[52] (TB- HIV) / 2018 35}	TB-HIV patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	57/450 (13%)	Age ≥ 45, resident of LTCF, meningeal TB, abnormal CXR, diagnosis confirmed by positive culture of NAA, culture not converted or unknown	AUROC = 0.79 (95% CI 0.70-0.87) HL test: X ² =4.25, p=0.51 Brier score: 0.09	Risk score	Low, High, Unclear, High
$\frac{1}{40^{\text{Polekareva[73]}}} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{20200^{\text{chird}}} \\ \frac{1}{20200^{\text{chird}}} \\ \frac{1}{20200^{\text{chird}}} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{20200^{\text{chird}}} \\ \frac{1}{20200^{\text{chird}}} \\ \frac{1}{2000^{\text{chird}}} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2013} \\ \frac{1}{2000^{\text{chird}}} \\ \frac{1}{2000^{\text{chird}}} \\ \frac{1}{2013} \\ \frac{1}{2013$	36 37 ^{Pefura-Yone[54] /} 2017 38	TB patients	Jan 2012 – Dec 2013	Retrospective cohort	Cameroon	Internal (bootstrap)	213/2250 (9%)	Age, adjusted BMI, clinical form (PTB+, PTB-, EPTB), HIV	C-statistic: 0.808 HL test: X ² =6.44, p=0.60 Sensitivity = 80.7% Specificity = 68.2% Calibration plot	Model coefficients	Low, Low, Low, High
42 Valade[42] / 2012 Hospitalized TB patients Mar 2000 – Jul 2009 Retrospective cohort France Internal (bootstrap) 20/53 (38%) Miliary TB, catecholamine infusion, mechanical ventilation on admission AUROC = 0.92 (95% CI: 0.85-0.98) Brier score = 0.13 Risk score Unc	39 40 ^{Podlekareva[73] /} 2013 41	TB/HIV patients	Jan 2004 – Dec 2006	Retrospective cohort	52 cities in Europe and Argentina	None	995†	DST performed, treatment with RHZ, and cART at/near TB diagnosis	Crude RH = 0.62 (95% CI: 0.64-0.84)	Risk score	Low, Unclear, Low, High
	12 Valade[42] / 2012 13	Hospitalized TB patients	Mar 2000 – Jul 2009	Retrospective cohort	France	Internal (bootstrap)	20/53 (38%)	Miliary TB, catecholamine infusion, mechanical ventilation on admission	AUROC = 0.92 (95% CI: 0.85-0.98) Brier score = 0.13	Risk score	Unclear, Low,

1								Optimism = 0.03 Accuracy = 85% Sensitivity - 75% Specificity = 91%		Low, High
3 4 Wang[74] / 2019	HIV-negative, culture-confirmed, pulmonary TB cases	Jan 2014 – Dec 2016	Prospective cohort	China	External (setting)	Development: 36/287 (13%) Validation: 15/104 (14%)	Age, cavitary lesion, pleural effusion, drug resistance, disseminated, albumin, c-reactive protein, white blood cell count, IL-6, MIF	AUROC = 0.85 ± 0.028	Odds ratios	Low, Low, Low, High
Wejse[75] / 2008	Pulmonary TB patients on DOT	1996 - 2001	Retrospective cohort	Guinea Bissau	None	100/698 (14%)	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	AUROC = 0.65 (95% CI: 0.6-0.7) Sensitivity = 0.45 Specificity = 0.75	Risk score	Low, High, Low, High
) Zhang[45] / 2019 0	TB/HIV patients at end stage of AIDS	Aug 2009 – Jan 2018	Retrospective cohort	China	Internal (split-sample)	Development: 157/807 (19%) Validation: 40/200 (20%)	Anemia, TB meningitis, severe pneumonia, hypoalbuminemia, unexplained infection or space-occupying lesions, malignancy	AUROC = 0.867 (95% CI: 0.832-0.902) Sensitivity = 79.6% Specificity = 82.9%	Risk score	Low, Low, Low, High
1 Treatment failure										
2 3 Abdelbary[9] / 4 2017	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 2109† Validation: 6322†	Education (no or low vs. higher than primary school), MDR, AFB smear (>+2, +1, negative)	c-statistic = 0.65 Sensitivity = 52% Specificity = 66%	Risk score	Low, High, Low, High
5 6 Kalhori[49] (logistic) / 2010	TB cases at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 828/4836 (17%) Validation: 2418†	Gender, age, weight nationality, prison, case type	AUROC = 0.70 Accuracy = 81.64% HL test: X ² =11.935, df=8, p=0.154	Model coefficients	Unclear, Unclear, Unclear, High
8 9 _{Keane[30] / 1997} 20 21	Smear-positive TB patients on standard first-line regimen with DOT	1990 - 1995	Non-nested case control	Vietnam	None	130/803 (16%)	3 month model: Extensive lesions, mediastinal shift, average smear score 3rd month, weight, progressive x-ray, any previous treatment Baseline model: Mediastinal shift, average smear score, extensive lesions, any previous treatment, cavities, weight	$\frac{3 \text{ month:}}{\text{Sensitivity} = 80\%}$ $\text{Specificity} = 80\%$ $\frac{\text{Baseline:}}{\text{Sensitivity} = 70\%}$ $\text{Specificity} = 80\%$	Model coefficients	High, Unclear, Unclear, High
22 23 ^{Luies[33] / 2017}	Smear-positive pulmonary TB cases on DOT	May 1999 – Jul 2002	Nested case- control	South Africa	Internal (cross-validation)	10/31 (32%)	3,5,-Dihydroxybenzoic acid, (3-(4-Hydroxy-3-methoxyphenyl) propionic acid	AUROC = 0.89 (95% CI: 0.7-1.00)	Model coefficients	High, Unclear, Unclear, High
25 Mburu[76] / 2018 26	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	13/321 (4%)	HbA1c, regimen (retreatment), age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine	AUROC = 0.56 ± 0.07	Relative score	Low, Low, Low, High
27 Default										
29 Thompson[77] / 30 2017	HIV uninfected adults with newly diagnosed pulmonary TB	Apr 2010 – Apr 2013	Retrospective cohort	South Africa	Internal (cross-validation) and external (setting)	6/99 (6%)	18 splice junctions and 13 genes	AUROC (internal) = 0.87 AUROC (external) = 0.63	Heatmap of differentially expressed genes	Low, Low, Low, High
 Abdelbary[9] / 2017 (TB-DM) 	TB cases	2006 - 2013	Retrospective cohort	Mexico	None	93/2121 (4%)	Age (<40 vs. ≥40), sex, HIV	c-statistic= 0.62	Risk score	Unclear, High, Unclear, High
34 35 _{Belilovsky[35]} / 36 2010 37	Hospitalized TB patients	1993 - 2002	Retrospective cohort	Russia	External (geographical)	Development: 1326/3904 (34%) Validation: 4662/12803 (36%)	Sex, unemployment, retreatment case, alcohol abuse (yes, no, no data), severe TB form, residence (urban vs. rural), age (25-50 vs. other), pulmonary TB (vs extrapulmonary), prison history	Belgrood: AUROC = 0.75 Orel: AUROC = 0.75 Pskov: AUROC = 0.78 Yaroslavi: AUROC = 0.75 Calibration table	Model coefficients	Unclear, High, High, High
8 9 10 _{Chang[31] / 2004} 11 12 13	All tuberculosis patients	Jan 1999 – Mar 1999	Nested case- control	China	None	102/408 (25%)	Baseline:* Ever smoker (current, former, never), retreatment (history of default, no history of default, not) Longitudinal: Smoking status (current, former, never), retreatment (with history of default, without history of default, never), unsatisfactory adherence in first two months (good, poor, fair, unknown), subsequent hospitalization, treatment side effects in last month of treatment	$\frac{\text{Baseline:}^{*}}{\text{AUROC} = 0.70 (95\% \text{ CI:} \\ 0.63-0.76)}$ HL test: X ² = 1.448, df=5, p=0.919 Longitudinal: AUROC = 0.85 (95% CI: 0.80-0.90)	Odds ratios	High, High, Low, High
14										31

BMJ Open

1								HL test: $X^2 = 5.887$, df=6, p=0.436		
2 Chee[78] / 2000	TB cases	1996	Nested case- control	Singapore	None	38/71 (54%)	Chinese race, extent of family support, treatment duration	Accuracy = 74.6%	Model coefficients	High, Unclear, High, High
4 5 6 Cherkaoui[29] / 6 2014 7	TB patients with definite or probable pulmonary or extrapulmonary TB	Jun 2010 – Oct 2011	Non-nested case-control	Morocco	None	91/277 (33%)	Age <50, work interfering with ability to take TB treatment, retreatment regimen, daily DOT, moderate or severe side effects, told friends about TB, current smoker, never smoker, symptom resolution in <2 months, knowledge of TB treatment duration	AUROC = 0.85 (95% CI: 0.80-0.90) Sensitivity = 82.4% Specificity = 87.6% HL test: X ² =0.77, p- value=1.00	Survey tool	High, High, High, High
3 9 _{Rodrigo[79] / 2012 10}	New TB cases	Jan 2006 – Dec 2009	Prospective cohort	Spain	Internal (split-sample)	Development: 92/1490 (6%) Validation: 103/1589 (6%)	Immigrant, living alone, living in an institution, previous TB treatment, linguistic barriers (poor understanding), IV drug use, unknown IV drug use	AUROC = 0.67 (95% CI: 0.65-0.70) Sensitivity = 65.05% Specificity = 67.36%	Risk score	Low, Low, Low, High
11 <u>Unfavorable</u> outcome										
12 <u>Entremi</u> 13 Kalhori[50] 14 ^{predicting) / 2009[†]}	TB patients at DOT registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 6920† Validation: 2966†	Age, gender, nationality, prison, area, weight	Classification rate = 89.8% R2 = 0.45	Model coefficients	Unclear, Unclear, Unclear, High
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29Sauer[57] / 2018 [†] 30 31 32 33 34 35 36 37 38 39 40 41 42 43	TB cases	Data available through March 2018	Retrospective cohort	Azerbaijan, Belarus, Georgia, Moldova, Romania	Internal (split-sample)	Development: 103/411 (25%) Validation: 44/176 (25%)	<u>Forward selection (FS):*</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Backwards elimination (BE):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Stepwise selection (SS):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Lasso:</u> Country, employment, extrapulmonary, cavity size, decrease in lung capacity, smear microscopy, drug sensitivity, chest imaging <u>Random forest (RF):</u> Top 5 by mean decrease accuracy: lung cavity size, type of resistance, employment status, country, total cavities Top 5 by mean decrease Gini index: Age of onset, drug regimen, lung cavity size, number of daily contacts, culture	$\frac{FS:*}{AUROC = 0.74}$ (95% CI: 0.66-0.82) Sensitivity = 0.36 Specificity = 0.89 Misclassification = 0.24 <u>BE:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.3 Specificity = 0.88 Misclassification = 0.27 <u>SS:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>Lasso</u> : AUROC = 0.72 (95% CI: 0.64-0.80) Sensitivity = 0.21 Specificity = 0.96 Misclassification = 0.23 <u>RF:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.96 Misclassification = 0.23 <u>RF:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>SVM linear</u> : AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0.21 Specificity = 0.94 Misclassification = 0.24 <u>SVM polynomial</u> : AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0 Specificity = 1 Misclassification = 0.25	List	Unclear, Unclear, Unclear, High

1 Baussano[47] / 2008§ Pulmonary TB cases 2001 - 20		2001 - 2005	Retrospective Italy Internal (bootstrap)		576/1242 (46%)	Residency (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)	AUROC= 0.75 Calibration slope = 0.98 $R^2 = 0.24$	Nomogram	Low, Unclear, Low, High		
B 4 5	Costa-Veiga[48] / 2017 [§]	Pulmonary TB cases	2000 - 2012	Retrospective cohort	Portugal	External (temporal)	Development: 1152/10766 (11%) Validation: 4714†	HIV, previous treatment, age class (25-44, 15-24, 45-64, >64), IV drug use, pathologies (other disease comorbidity)	AUROC= 75.9% (95% CI: 74.1-77.7) Sensitivity = 71% Specificity = 73%	Nomogram	Low, Low, Low, High
o 7 8 9 Killian[34] / 2019 [§] 10 11		TB patients (99DOTS program)	TB patients (99DOTS program) Feb 2017 – Sep 2018 Retrospective cohort India None 433/4167 (10%) India Mone 433/4167 (10%) India India None 433/4167 (10%) India India India		LEAP* AUROC = 0.743 <u>lw-misses:</u> AUROC = 0.607 <u>t-misses:</u> AUROC = 0.630 <u>Random forest:</u> AUROC = 0.722	None	High, High, Unclear, High				
13 13 14	Madan[51] / 2018§	TB-HIV patients on DOT with first-line TB treatment	2015	Retrospective cohort	India	None	78/448 (17%)	Sputum smear grade, previous TB,; disease classification, HIV status, ART status, CD4 cell count, sex and age group (with interaction terms between age group and sex; sputum smear status and type of TB; HIV status at TB diagnosis and CD4 cell category).	AUROC = 0.783 HL test p-value = 0.149	Model coefficients	Low, Low, Low, High
16 17	Mburu[76] / 2018§	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	32/340 (9%)	HbA1c, treatment regimen (retreatment), creatinine, BMI, BUN, weight, age, random blood glucose, HIV positive result, male gender	AUROC = 0.65 ± 0.06	Relative score	Low, Low, Low, High
10 Other outcome											
20 2 Kalhori[80] (fuzzy) 2 Kalhori[80] (fuzzy) 2 Z TB patients at DOTS registration		TB patients at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 7254† Validation: 2418†	Case type, treatment category, risky sex, prison, sex, recent TB infection, diabetes, low body weight, TB type, length, previous imprisonment, age, area, HIV	Mean absolute percentage error = 1.24	Learned parameters	Unclear, Unclear, High, High
23 24 25	Hussain[56] / 2019	Pulmonary and extrapulmonary TB patients (TB Reach)	2011 - 2014	Retrospective cohort	Unknown	Internal (split-sample)	Development: 3371† Validation: 842†	Random forest*, artificial neural networks, and SVM	Random forest:* Accuracy = 76.32%	None	Unclear, Unclear, Unclear, High
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42	Abbreviations: AUROC=Area under receiver operating characteristic; c-statistic=concordance statistic; DOTS=Directly Observed Therapy, DM=Diabetes; HL=Hosmer-Lemeshow; TB=Tuberculosis; Thaticates best-performing most relevant model, which is included throughout the manuscript (see methods section for details). Performance measures are reported for highest level of validation performed (ranked from strongest to weakest: external validation, internal validation, internal validation), and related validation, internal validation, i										

Table 3. Characteristics of patient populations in the 33 included studies with prediction models for tuberculosis treatment outcomes

Characteristic	Studies	Categories	N(%) or	
	reporting	8	Median [IOR]	
	characteristic.			
	n (% of total)			
Sample size	33 (11)	_	803 [291, 4167]	
Study duration	32 (97)	_	4 [2 7]	
vears	5=(57)		. [_,,]	
Study design	33 (100)	Prospective cohort	3 (9)	
· · · · · · · · · · · · · · · · · · ·		Retrospective cohort	25 (76)	
		Nested case-control	3 (9)	
		Non-nested case-control	2 (6)	
Data source	33 (100)	Medical record	6 (18)	
		National registry or surveillance	13 (39)	
		system		
		Local registry or surveillance	1 (3)	
		system		
		Regional registry or surveillance	2 (6)	
		system		
		Data collect form for study	11 (33)	
		purposes		
Study region	32 (97)	Africa	8 (25)	
		Asia	13 (41)	
		Europe	6 (19)	
		North America	4 (12)	
		South America	0 (0)	
		Global	1 (3)	
High burden TB	31 (94)	All	143(42)	
setting*		2		
		Some	1 (3)	
		None	17 (55)	
Missing data	18 (54)	Complete case-analysis	9 (50)	
		Missing indicator method	4 (22)	
		Heckman's method	1 (6)	
		Simple imputation	2 (12)	
		Sensitivity analysis with	1 (6)	
		imputation		
		Other	1 (5)	
Number of models developed	33 (100)	1	25 (76)	
· · · · · · · · · · · · · · · · · · ·		2	4 (12)	
		3	1 (3)	

		4	2 (6)
		7	1 (3)
Reasons for multiple models developed	8 (24)	Different outcomes	1 (12)
		Different predictors considered	4 (50)
		Different methods	2 (25)
		Different outcomes	1 (12)
		Different populations and	1 (12)
		outcomes	

*Determined based on study location and WHO list of 30 high-burden TB countries in the 2019 Global Tuberculosis Report (1).

2 3 4	Ta
5	
6	
7 o	
9	
10	
11	
12	
13 14	
15	F
16	
17	
18	
19 20	
21	
22	Ab
23	resi
24 25	*Ba
25	†Fo
27	me
28	‡Ot
29	cha
30 31	18
32	the
33	
34	
35	
30	
38	
39	
40	
41 42	
42 43	
44	
45	
46	
47 48	
-+	

		Included?		
Characteristic	Yes	No	Unknown	Median [IQR] [‡] , n
Age*	-	-	15	41 [37, 49], n=18
HIV	18	7	8	23% [10-100], n=17
Diabetes	12	2	19	12% [5-21], n=11
MDR	8	7	18	1% [1-3], n=8
Other drug resistance	12	1	20	6% [4-12], n=10
Extrapulmonary TB [†]	22	4	7	11% [4-17], n=16
Previous TB	20	1	12	19% [9-30], n=17
DOT	14	0	19	100% [100-100], n=14
Hospitalized patients	13	1	19	100% [100-100], n=10

Abbreviations: DOT=directly observed therapy; IQR=interquartile range; MDR=multi-drug resistance; TB=tuberculosis

*Based on the measure of central tendency reported in the study (mean: n=11; median: n=7) [†]Forms of extrapulmonary TB differ by study but included some of the following: Miliary, meningeal, pleural, peritoneal, disseminated, blood/bone, abdominal

[‡]Other than age (which is reported in years), this is the percentage of the population that has the characteristic among studies that include patients with the characteristic. For example, among the 18 studies that include persons with HIV, 17 report how many people had HIV and among those, the median percentage of the population with HIV is 23%.

Table 5. Methods reported for the 37 models of the 33 included studies with prediction models for tuberculosis treatment outcomes

Characteristic	Studies reporting characteristic, n (%)	Categories	N(%) or median [IQR]
Type of outcome	37 (100)	Single	29 (78)
		Composite	8 (22)
Outcome	37 (100)	Death	16 (43)
		Treatment failure	6 (16)
		Default, Loss to follow-up,	6 (16)
		or treatment interruption	
0		Unfavorable outcome	6 (16)
		Treatment success	2 (6)
		Other [‡]	1 (3)
Number - prevalence of outcome*	32 (87)	-	94 [38-171] 15% [9-26]
Events per candidate variable [†]	30 (81)	-	6 [3-11]
Events per variable (in final model)	29 (78)	-	14 [9-26]
Predictor types		Clinical/epidemiologic	34 (92)
		Adherence	1 (3)
		Biomarker	2 (5)
Analysis	37 (100)	Logistic regression	29 (78)
		Survival analysis	3 (8)
		Machine learning	5 (14)
Method for considering predictors in multivariable models	36 (97)	All candidate predictors	12 (32)
		Based on unadjusted association with outcome	19 (51)
		Based on clinical relevance	1 (3)
		Other§	4 (14)
Selection of predictors during modeling	31 (84)	Full model approach	2 (6)
		Forward selection	7 (23)
		Backwards elimination	5 (16)
		Stepwise selection	8 (26)
		Random Forest	1 (3)
		Hosmer-Lemeshow model building criteria	4 (13)
		Bayesian model averaging	3 (10)
		Pairwise selection	1 (3)
1			
--------	--		
2			
2			
1			
4 C			
5			
6			
/			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
45			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			

59

60

P-value for consideration in model	17 (46)	0.01	2 (12)
		0.05	3 (18)
		0.11	1 (6)
		0.2	6 (35)
		0.25	5 (29)
P-value for retention in MV model	20 (54)	0.02	9 (45)
		0.1	9 (45)
		0.12	1 (5)
		0.2	1 (5)
Internal validation	19 (51)	Split-sample	10 (53)
		Bootstrap	5 (26)
		Cross-validation	4 (21)
External validation	6 (16)	Temporal	1 (17)
		Geographic	1 (4)
		Setting	4 (67)
Calibration	17 (46)	Calibration plot [¶]	2 (12)
		Calibration slope [¶]	1 (6)
		Hosmer-Lemeshow	13 (77)
		goodness of fit p-value [¶]	0.51 [0.20, 0.79]
		Calibration table [¶]	2 (12)
		Mean absolute error [¶]	1 (6)
Discrimination	30 (81)	C-statistic (AUROC) [¶]	30 (100)
			0.75 [0.68-0.84]
		Log rank test¶	2 (5)
Classification	18 (49)	Sensitivity	14 (78)
Chubbhhoution	10(1)	Sensitivity	70 [54 78]
		Specificity	13 (72)
		specificity	75 [71 00]
		Accuracy	2(11)
		Accuracy Other**	$\begin{array}{c} 75[71,88] \\ \hline 2(11) \\ \hline 2(11) \end{array}$
Model presentation	34 (92)	Accuracy Other ^{**} Risk score	$ \begin{array}{c} 73[71, 88] \\ 2(11) \\ 2(11) \\ 16(43) \end{array} $
Model presentation	34 (92)	Accuracy Other ^{**} Risk score Model coefficient	$ \begin{array}{c c} 73[71, 88] \\ \hline 2 (11) \\ \hline 2 (11) \\ \hline 16 (43) \\ \hline 8 (22) \\ \end{array} $
Model presentation	34 (92)	Accuracy Other ^{**} Risk score Model coefficient Nomogram	$ \begin{array}{c} 73 [71, 88] \\ 2 (11) \\ 2 (11) \\ 16 (43) \\ 8 (22) \\ 2 (6) \end{array} $
Model presentation	34 (92)	Accuracy Other** Risk score Model coefficient Nomogram Odds ratios/relative scores	$ \begin{array}{c} 73 [71, 88] \\ 2 (11) \\ 2 (11) \\ 16 (43) \\ 8 (22) \\ 2 (6) \\ 4 (12) \end{array} $

Abbreviations: AUROC=area under receiver operating characteristic; c-statistic=concordance statistic

*Prevalence of outcome in the population used to develop the prediction model (i.e.

derivation/development subset if split-sample technique was used or full sample if the model was not validated or if bootstrap/cross-validation was used)

[†]Only 5 studies report the exact number of predictors considered. Otherwise, the number of candidate predictors was estimated from the provided tables or lists of candidate predictors in the source paper.

[‡]Outcome is a value from 1 to 5 (1= patient completed the treatment course in frame of DOTS, 2=cured, 3= quit treatment, 4=failed treatment and 5=death)

[§]Other methods of determining which variables to consider for prediction model include: principal components analysis (n=1), screening for multi-collinearity via correlation coefficient (n=1), one study used a combination of a priori and selection via univariable association, and the other used machine learning pre-processing (n=1)

[¶]Sums to more than 100%, because some studies report multiple measures of calibration or discrimination

Based on the following cut-off methods: Youden (n=4) concordance probability (n=1), estimated at nearest 0,1 for studies that present a range of sensitivity and specificity in a table or figure (n=4), or unknown (n=5)

**Other includes one study that reports false positive rate and one study that includes a graph of sensitivity vs. specificity.

Figure 2. Most common predictors considered and included

[See Figure 2]

Figure 2 legend:

Considered: the predictor as evaluated as a candidate predictor prior to multivariable modeling Included: the predictor was considered and subsequently included in the final multivariable model

Figure 3. Heatmap of signaling questions from risk of bias assessment with PROBAST

[See Figure 3]

Figure 3 legend:

PROBAST questions (additional details in Supplemental File 5)

- Participants 1: What study design was used and was it appropriate?
- Participants 2: Were all inclusion and exclusion criteria appropriate?
- Predictors 1: Were predictors defined as assessed the same way for all participants?
 - Predictors 2: Were predictor assessments made without knowledge of data outcome?
 - Predictors 3: Are all predictors available at the time the model was intended to be used?
- Outcome 1: Was the outcome determined appropriately?
- Outcome 2: Was the outcome pre-specified or standard?
- Outcome 3: Were predictors excluded from outcome definition?
- Outcome 4: Was the outcome defined and determined in a similar way for all participants?
- Outcome 5: Was the outcome determined without predictor information?

Outcome 6: Was the time interval between predictor assessment and outcome determination appropriate?

- Analysis 1: Were there a reasonable number of participants with the outcome?
- Analysis 2: Were continuous and categorical variables handled appropriately?
- Analysis 3: Were all enrolled participants included in the analysis?
- Analysis 4: Were participants with missing data handled appropriately?
 - Analysis 5: Was selection of predictors based on univariable analysis avoided?
 - Analysis 6: Were complexities in data (censoring, competing risks, sampling of control participants) accounted for appropriately?
 - Analysis 7: Were relevant model performance measures evaluated appropriately?

Analysis 8: Were model overfitting, underfitting, and optimism in the model performance accounted for?

Analysis 9: Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?



[See Figure 4]

to been terier only

BMJ Open





203x203mm (300 x 300 DPI)

2	
6	

~	
7	





PROBAST Question

203x203mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml





203x203mm (300 x 300 DPI)

BMJ Open

Supplemental File 1. PRISMA Checklist

Section/topic	#	Checklist item	Reported o #			
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 e criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; co and implications of key findings; systematic review registration number.		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			1			
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).	Supplement 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			
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. 9						
Search	ch 8 Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.					
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).				
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.				
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; Supplem Files 4 and			
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; Supplem 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., l 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			<u>+</u>			
Study selection	tudy selection17Give 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; Tabl 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; Figu and 4			

	1						
Results of individual studies	tesults of ndividual studies20For 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.						
Synthesis of results	ynthesis of results 21 Present results of each meta-analysis done, including confidence intervals and measures of consistency.						
Risk of bias across studies	Risk of bias across 22 Present results of any assessment of risk of bias across studies (see Item 15). studies 22		N/A				
Additional analysis 23 Give results of addition. 16]).		Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A				
DISCUSSION	-						
Summary of evidence24Summarize the main findings including the strength of evidence for e relevance to key groups (e.g., healthcare providers, users, and policy		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 2		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						
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.	21				

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	ention Any prognostic model developed to predict tuberculosis treatment outcome. This includes model development studies with and with external validation				
Comparator Models will be compared to each other, as there is no other relevant comparator for this systematic review					
	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				
Outcome	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				
6	support for individuals at greatest risk for unsuccessful TB treatment outcomes.				

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Supplemental File 3. Search Strategy

	Search terms
PubMed	 ((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]) AN (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] factor*[tiab] OR model*[tiab])) (stratification[tiab] OR "ROC Curve"[Mesh] OR discrimination[tiab] OR discriminate[tiab] OR "c-statistic"[tiab] OR "c statistic"[tiab]) (stratification[tiab] OR "ROC Curve"[Mesh] OR discrimination[tiab] OR indices[tiab] OR algorithm[tiab] OR multivariable[tiab]) (tuberculosis[Mesh] OR tuberculosis[tiab]) (tuberculosis[Mesh] OR mortality*[tiab] OR death*[tiab] OR fail*[tiab] OR recur*[tiab] OR relapse*[tiab] OR default*[tiab] OR abandon*[tiab] OR mortality*[tiab] OR cure*[tiab] OR success*[tiab] OR unsuccess*[tiab] OR die[tiab] OR died[tiab] OR dies[tiab])) 1 OR 2 3 AND 4 5 AND 6 AND (humans[Filter]) AND ("1995"[Date - Publication] : "3000"[Date - Publication])
Embase	 (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. (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. (outcome\$ or mortality\$ or death\$ or fail\$ or recur\$ or relapse\$ or default\$ or abandon\$ or loss\$ or cure\$ or success\$ or unsuccess\$ or or died or dies).ti,ab. 1 or 2 3 and 4 5 and 6 limit 7 to (human and yr="1995 -Current")
Web of Science	 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 ident prognos*)) OR TS=(decision* and ((model* or clinical*). or "statistical model")) 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*)) TS=(stratification or discrimination or discriminate or c-statistic or "c statistic" or "area under the curve" or AUC or calibration or indic algorithm or multivariable or "receiver operating characteristic") TS=(tuberculosis) TS=(outcome* or mortality* or death* or fail* or recur* or relapse* or default* or abandon* or loss* or cure* or success* or unsuccess die or died or dies) 1 or 2 3 and 4 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)			
	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting,			
	inclusion and exclusion criteria)			
PADTICIDANTS	Participant description			
TAKIICITANIS	Details of treatments received, if relevant			
Study dates				
	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)			
PREDICTED	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			
	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease			
	characteristics)			
	Definition and method for measurement of candidate predictors			
CANDIDATE	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)?			
OR INDEX TESTS)	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)			
	Number of participants with any missing value (include predictors and outcomes)			
MISSING DATA	Number of participants with missing data for each predictor			
MISSING DATA	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)			
	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)			
MODEL DEVELOPMENT	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)			
	Calibration (calibration plot calibration slope Hosmer-Lemeshow test) and Discrimination (C-statistic D-			
MODEL	statistic. log-rank) measures with confidence intervals			
PERFORMANCE	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a- priori cut points were used			
	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)			
MODEL EVALUATION	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)			
	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)			
RESULTS	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/probabl	<u>y yes</u>	No/probably no		No information	
1	What study design was used and was it appropriate?Yes: If a cohort design or proper registry data) you have confidence in participant enrollment i describedWhat study design was used and was it appropriate?Probably yes: a nested case-cohort design (wi adjustment of the basel the analysis) has been u design was used but pa enrollment was data ou		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 describedNo: If a non-nested case-control de has been usedhat study design was id and was it propriate?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 unclearNo: If a non-nested case-control de has been used has been used	esign If the method of participant sampling is unclear. study it of			
2	Were all inclusion and exclusion criteria appropriate?	 enrollment was data quality is unclear 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.	
				6	1		
	Low risk of	bias	High risk of bias			Unclear risk of bias	
If the "Prob ≥ 1 of could	answer to all signaling quest bably yes," then risk of bias c the answers is "No" or "Prob I still be "Low risk of bias" bu	ions is "Yes" or an be considered low. If pably no," the judgment at specific reasons should	If the answer to a "No" or "Probab except if defined	any of the signaling questions is oly no," there is a potential for bias, I at low risk of bias above.	If relet the signali domai	vant information is missing for some of gnaling questions and none of the ing questions is judged to put this n at high risk of bias.	

Domain 2: Predictors

be provided why the risk of bias can be considered low.

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 l outcome information during predictor asses entirely clear (retrosp collection/surveillanc	ikely that was not used ssment, but not ective data e data) prs would be	Predictors would not be avail	able at	No information on whether	
3	Are all predictors available a the time the model was inten to be used?	t ded intended to be used for	redictors would not be available the time the model is the time the model is intended to be used for prediction used for prediction.		to be	at No information on whether e predictors would be available at time the model is intended to be used for prediction.	
	Low risk of	bias	Н	ligh risk of bias		Unclear risk of bias	
If th "Pro ≥1 o cou be p	he answer to all signaling ques obably yes," then risk of bias of of the answers is "No" or "Pro Id still be "Low risk of bias" b provided why the risk of bias of	tions is "Yes" or an be considered low. If bably no," the judgment ut specific reasons should an be considered low.	If the answer to a "No" or "Probabl bias, except if def	ny of the signaling questions is ly no," there is a potential for fined at low risk of bias above.	If relevan the signal signaling at high ris	at information is missing for some ling questions and none of the questions is judged to put this do sk of bias.	
Dor Bias outo met dete	main 3: Outcome s in model performance can oc come determination can result thods are inconsistently applied ermination can also result in bi	cur when methods used to de from use of suboptimal meth d across participants, or wher as.	etermine outcomes ods, tests, or criter h knowledge of pre	incorrectly classify participants w ria that lead to unacceptably high l edictors influence outcome determ	vith or with levels of err ination. Inc	out the outcome. Bias in methods fors in outcome determination, where the outcome	
	Signaling question	Yes/probably	yes	No/probably no		No information	
1	Was the outcome determined appropriately?	If a method of outcome de been used which is conside acceptable by guidelines or publications on the topic Note: This is about level o error within the method of the outcome (see concerns applicability about whethe of the outcome method is a	termination has ered optimal or r previous f measurement determining for r the definition uppropriate).	If a clearly suboptimal method ha used that causes unacceptable err determining outcome status in participants	as been or in	No information on how outcom determined	
2	Was the outcome pre- specified or standard?	Yes: If the method of outco determination is objective, outcome definition is used prespecified categories are outcomes. (i.e. outcome as based on previously publis published study protocol, o guidelines) Probably yes: The outcome is not clearly based on guid previous research, but outco assessment is objective and	ome or if a standard , or if used to group sessment is hed studies, or clinical e determination delines or toome d would not	No: If the outcome definition was standard and not prespecified Probably no: a non-standard or no prespecified outcome was used, a unclear whether the outcome defi could introduce bias. *Caution with composite outcom favor a better model by excluding outcome components or including atypical events	s not on- ind it is inition es that g typical g	No information on whether the outcome definition was prespec or standard	
3	Were predictors excluded from outcome definition?	inadvertently alter study re Yes: None of the predictor in the outcome definition (Probably yes: None of the included in the outcome de (assumed)	sults s are included clearly stated) predictors are efinition	If ≥1 of the predictors forms part outcome definition	of the	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 defi determined in a similar wa participants (clearly stated) Probably yes: If outcomes and determined in a simila participants (assumed)	ined and y for all) were defined r way for all	If outcomes were clearly defined determined in a different way for participants	and some	No information on whether out were defined or determined in a similar way for all participants	
5	Was the outcome determined without predictor information	Yes: If predictor informatic known when determining t status, or outcome status de clearly reported as determi knowledge of predictor inf Probably yes: predictor inf have been available at time assessment, but outcome d objective and knowing infor predictors would not influe	on was not he outcome etermination is ned without formation.	No: If it is clear that predictor inf was used when determining the o status Probably no: it is likely predictor information was available at the t outcome assessment, and outcom definition is subjective and know predictors could influence outcom determination.	time of ledge of ne	No information on whether out was determined without knowle of predictor information	

6	Was the time interval between predictor determination appropriate	sased on culture results, assed on culture results, f the time interval betwee issessment and outcome was appropriate to enable and representative numb boutcomes to be recorded information on the time is equired to allow a repre- of the relevant outcome of predictor assessment and letermination were from aken within an approprise	etc) een predictor determination e the correct type er of relevant , or if no interval is sentative number occur or if l outcome information ate time interval.	If the tin assessme too short type and relevant	ne interval between predi ent and outcome determin t or too long to enable the representative number o outcomes to be recorded	ctor nation is e correct f	If no information was provide the time interval between pre- assessment and outcome determination.
	Low risk of bia	15	F	ligh risk	of bias		Unclear risk of bias
If th "Pr ≥1 cou sho low	he answer to all signaling question obably yes," then risk of bias can of the answers is "No" or "Probal and still be "Low risk of bias" but buld be provided why the risk of b 7.	ns is "Yes" or be considered low. If oly no," the judgment specific reasons ias can be considered	If the answer to a "No" or "Probabl bias, except if def	ny of the s y no," the fined at lo	signaling questions is re is a potential for w risk of bias above.	If relevan the signal signaling at high ris	It information is missing for son ling questions and none of the questions is judged to put this sk of bias.
Do Stat bias see	main 4: Analysis tistical analysis is a critical part o s in reported model performance n k statistical advice when completi	f prediction model devel measures. Model develo	lopment and validati pment studies includ	ion. The u de many s	use of inappropriate statis steps where flawed metho	tical analys ods can diste	is methods increases the potent ort results. We recommend revi
	Signaling question	For model develop	probably yes	number	No/probably	no t studies if	No information
1	Were there a reasonable number of participants with the outcome	e there a reasonable number articipants with the outcome? For model validation studies, if the number of s ≥ 20 (EPV ≥ 20).* For model validation studies, if the number of participants with the outcome is ≥ 100 .		we to ameters mber of	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.		 no information on the nu candidate predictor parar or number of participants the outcome, such that th cannot be calculated. For model validation studinformation on the numb participants with the outcome.
		* For EPVs betwee frequency, overall 145 to 147.	en 10 and 20, the ite model performance	em should , and distr	be rated as either probab ribution of the predictors	ly yes or pr in the mode	obably no, depending on the ou el. For more guidance, see refer
2	Were continuous and categorica predictors handled appropriately	Yes: If continuous continuous or if co examined as linear restricted cubic spl polynomials. Probably yes: If cc converted into >2 / the model (i.e., dic using a prespecifie avoids sparse data/ improve statistical For model validati predictors are inclu- definitions or trans variables are categ points, ascompared study.	predictors are kept ntinuous predictors or non-linear using lines or fractional ontinuous predictors categories when incl hotomized or catego d method or in a wa 'would not intention significance. on studies, if continu- ided using the same formations, and cate orized using the sam l with the developm	as are are not luded in orized) ay that hally uous egorical ne cut hent	No: For model develop if continuous predictors converted into 2 catego included in the model. Probably no: If categor group definitions do no prespecified method or variables were split into but the decision of how variables is unclear. For model validation st continuous predictors a using different definitio transformations, or cate variables are categorize different cut points, as with the development s	ment studie s are ries when ical predict to use a continuous o >2 groups to split rudies, if rre included ons or egorical ed using compared tudy.	 No information on wheth continuous predictors are examined for nonlinearity no information on how categorical predictor grou defined. For model validation stud information on whether t same definitions or transformations and the s cut points are used, as co with the development stud
3	Were all enrolled participants included in the analysis?	If all participants e included in the dat	nrolled in the study a analysis.	are	If some or a subgroup of are inappropriately exc the analysis (because the missing data, unknown outliers)	of participar luded from ney were outcome,	nts No information on wheth enrolled participants are included in the analysis.
1.1		Yes: If there are no predictors or outco explicitly reports t	o missing values of omes and the study hat participants are i	not	No: If participants with are omitted from the an the method of handling	n missing da nalysis, or if missing da	ta If there is insufficient information to determine method of handling miss

Apj	, v						
	plicability						
risk "Pr spe	The answer to all signaling questions is c of bias can be considered low. If ≥ 1 obably no," the judgment could still l cific reasons should be provided why isidered low.	of the answers is "No" or of the answers is "No" or of "Low risk of bias" but the risk of bias can be	a potential for risk of bias ab	No" or "Probably no," there is bias, except if defined at low ove.	of the s this dor	f the signaling questions and mignaling questions and mignaling questions is judged to nain at high risk of bias.	
If +1	Low risk of	bias	If the answer	High risk of bias	If rolar	Unclear risk of bias	
				· · · · ·			
9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	redictors and their assigned ghts in the final model espond to the results from the orted multivariable analysis?		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. 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.	
3	underfitting, and optimism in model performance accounted for?					If it is unclear whether the	
	Ware model everfitting	Yes: If internal validation techni (bootstrapping and cross-validat all model development procedur to account for any optimism in r and subsequent adjustment of th	iques ion) including res, were used nodel fitting, e model liced	No: If no internal validation h been performed, or if internal validation consists only of a s random split-sample of partic data,	as ingle ipant	No information: No informa is provided on whether inter validation techniques, includ all model development procedures, have been appli	
7	Were relevant model performance measures evaluated appropriately?	Probably yes: if authors present a table of predicted probabilities with confidence intervals and corresponding outcome frequencies across subgroups		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.		survival outcomes are used references to relevant literat or specific mention of meth such as using Kaplan–Meie estimates), or no informatio thresholds for estimating classification measures is gi	
		Yes: If both calibration (via cali and discrimination (c-index) are appropriately (including relevan tailored for models predicting su outcomes).	bration plot) evaluated t measures urvival	If both calibration and discrimination are not evaluat if only goodness-of-fit tests (Hosmer-Lemeshow test), are to evaluate calibration or if fo	ed, or used r	Either calibration or discrimination are not report or no information is provide to whether appropriate performance measures for	
6	Were complexities in the data (censoring, competing risks, sampling of control participants) accounted for appropriately?	for appropriately, or if it is clear potential data complexities have identified appropriately as unimp	that any been portant.	affect model performance are ignore. For example, case-cor studies that do not estimate ba risk or studies with censoring competing risks that do not us survival analysis or other appropriate methods.	ntrol aseline or ae	whether complexities in the are present or accounted for appropriately if present.	
5	Was selection of predictors based on univariable analysis avoided?	of univariable analysis prior to n modeling.	nultivariable	basis of univariable analysis p multivariable modeling.	prior to	indicate that univariable selection is avoided.	
		provide comparison of included participants or if sensitivity anal imputation methods are convinc is low	vs. excluded ysis with ing that bias on the basis	Probably no: If authors provid comparison of included vs. ex participants or if sensitivity ar with imputation methods are reported, but the results are no convincing to rule out bias fro excluding missing data If the predictors are selected of	de acluded aalysis ot om on the	If there is no information to	
		with missing data were excluded	l and authors	of methods to handle missing	data.		

Page 57 of 62

<u>Participants</u> : 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.
<u>Predictors</u> : 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.

	BMJ	Open	
ticipants: do you have concern t the included participants or ing do not match the review	Included participants and clinical setting match the review question.	Included participants and clinical setting were different from the review question.	If relevant inf participants a reported.
dictors: does the definition, essment, or timing of predictors tch 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 inf predictors is r
teome: does the definition, ing, or determination of come match the review stion?	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 outcome, time determination

Supplemental File 6. Model outcome definitions

Study ID	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 /	Dooth	The outcome was mentality rick at 2 menths after admission
Horita / 2013	Death	The outcome was mortanty risk at 2 months after admission. '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 fo 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) transferpatient 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	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 has 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	Default was defined as failure to collect drugs for 2 months or more after registration

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	(interruption > 2)	
	months)	
	Default,	
	Abandon, or	
	LTF	Defaulter or cases were defined as patients on anti-tuberculosis treatment at the TBCU who failed to turn up for the
	(interruption >2	scheduled appointments despite usual attempts to recall them by phone or mail, as described below, and from whom
Chee / 2000	months)	least one home visit during the study was recorded
	Default,	
	Abandon, or	
	LTF	
	(interruption >2	
Cherkaoui / 2014	months)	Treatment default was defined as an interruption in TB treatment for >=2 consecutive months.
	Default,	
	Abandon, or	
	LIF (intermention >2	Intermetion of two two art for one according to a work that 2 months and completion of two two twithin 0 months who
Podrigo / 2012	(interruption >2 months)	Interruption of treatment for any reason for more instance of $\mathcal{L}^{(0)}(x)$ is already within 9 months when
Kalhori	Treatment	patient is placed on a 6 monun regiment of drug intake of ~80% the presented dose.
(predicting) /	success (cure +	
2009	completion)	For each patient dependent variable was recorded whether or not the patient finished the treatment course and get of
	Unfavorable	
	outcome (death	
Sauer / 2018	+ failure)	The primary outcome was treatment failure, which we defined as failure of therapy or death.
	Unfavorable	
	outcome (death.	
	failure, LTF,	Treatment interruption or default, treatment failure, transferred out cases and those lost to follow-up were grouped
Baussano / 2008	NE)	'unsuccessful outcomes
		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
		declared cured, including both negative smear microscopy at the end of treatment at least one previous follow-up to
	Unfavorable	in case of not providing sputum samples, cure is declared if treatment completed and absent of disease clinical evid
	outcome (death,	(categories 1 and 2). (ii) Unsuccessful outcome-if treatment of PTB patients resulted in failure (i.e. remaining smea
Costa-Veiga /	failure, LTF,	positive after 5 months of treatment, cat. 3), default (i.e. patients who interrupted their treatment for two consecutive
2017	NE)	months or more after registration, cat. 4), death (cat. 5) or were transferred-out (cat. 6)
	Unfavorable	
	outcome (death,	
	failure, LTF,	We label 'Cured' and 'Treatment Complete' to be favorable outcomes and 'Died', 'Treatment failed', and 'Lost to foll
Killian / 2019	NE)	up' to be unfavorable outcomes
	Unfavorable	
	outcome (death,	
M. J / 2019	failure, LTF,	Favourable treatment outcomes included cure and treatment completed. Unfavourable treatment outcomes included
Madan / 2018	INE)	death, loss to follow-up, treatment failure, transfer out, or a switch to MDK 1 B treatment.
	outcome (death	
Mburu / 2018 -	failure LTF	
Unfavorable	NE)	The primary analyses compared favorable versus unfavorable outcomes at end of treatment
	Other	The values of outcomes might be any values from 1 to 5 which means different outcomes. Value 1 means nation
Kalhori (fuzzv) /	composite	completed the treatment course in frame of DOTS. 2 means the nation thas been cured. 3 means nations has quitted
2009	outcome	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 -	2 + 2*(Age 41-65) + 5*(Age = 65) + 2*(Male gender) + 4*(MDR TB) + 3*(HIV) + 3*(Malnutrition) + 2*(Alcoholism) + 3*(Malnutrition) + 2*(Alcoholism) + 3*(Malnutrition) + 3*(Malnutrition
Death	2*(Male*diabetes) + 3*(HIV*pulmonary TB) - 1*(diabetes) - 1*(pulmonary TB)
Abdelbary / 2017 -	$9*(M_{1}, \dots, 1, \dots, 1, \dots, 1, \dots, 1, 1, 0) + 10*(MDD) + 10*(MDD, \dots, 1, 1) + 15*(MDD, \dots, 1, 1)$
Abdelbary / 2017 -	8° (No or low education) + 40° (MDR) + 10° (AFB sinear +2) + 13° (AFB sinear +3)
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)
	Don't report final model, but show the beta coefficients. The coefficients are written as predictor (beta-coefficient): age ³ 65 (2.497),
Aljohaney / 2018	congestive heart failure (1.231), bilateral disease on chest x-ray (1.192)
Bastos / 2016	$3^{(Hypoxemic respiratory failure) + 2^{(Age>=30) + 1^{(Bilateral involvement) + 1^{(At least one of: HIV, diabetes, liver failure/cirrnosis, concessive heart failure, chronic respiratory disease) + 1*(Hemoglobin<12)$
Dastos / 2010	Nomogram with: residency status (residential vs. homeless) sex. geographic origin (non-EU vs. EU) case definition (other than definite
Baussano / 2008	vs. definite), treatment setting (inpatient and unknown vs. outpatient), age (continuous)
	-3.2 + 0.8* (male gender) $+ 0.7*$ (unemployment) $+ 0.4*$ (retreatment case) $+ 1.1*$ (alcohol abuse) $+ 0.6*$ (no data about alcohol) $+ 0.4*$
Belilovsky / 2010	0.8*(severe TB form) - $0.3*$ (urban residence) + $0.4*$ (age 25-50) + $0.8*$ (pulmonary TB) + $0.5*$ (prison history)
C1 / 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:
Chang / 2004	Current smokers (3.44), ex-smokers (2.48), history of default (10./4), no history of default (0.80), The OP for each and interiment of the formation distance (OP). Non Chinese marks (0.80),
	friends (0.08) Treatment duration (1.85) Treatment duration is categorical as 6 months, 9 months, and >9 months, but only one OR is
Chee / 2000	presented.
	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
Charles	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
Cnerkaoui / 2014	Know now long your 1 B treatment is supposed to last (1 point for no) Have you ever smoked cigarettes (-1 point for no)
Costa-Veiga / 2017	isomogram with: r11v, previous treatment, age class (23-44, 13-24, 43-64, >64), 1V drug use, pathologies (other disease comorbidity: ves/no)
Costa Velga / 2017	$9^{*}(Male sex) + 7^{*}(nation t aged 55+) + 6^{*}(currently taking ART) + 7^{*}(unable to walk unaided) + 7^{*}(hemoglobin < 80, severe anemia) + 10^{*}(matrix)$
Gupta-Wright / 2019	6*(positive on urine TB-LAM)
Horrito / 2012	1*A as (years) + 10*(evycan requirement) 20*(elbumin) + 5*(comi dependent ADI) + 10*(total dependent ADI)
Horita / 2015	1 'Age (years) + 10'(oxygen requirement) - 20'(alounnin) + 5'(senn-dependent, ADL) + 10'(total dependent, ADL)
Hussain / 2019	None
	Learned parameters by training set for each predictor written as predictor (learned parameter): Case type (0.467), treatment category (-
Kalhori (fuzzy) /	(0.0/9), risky sex (-0.945), prison (0.992), sex (0.400), recent 1B intection (0.75), diabetes (2.445), low body Weight (1.515), 1B type (0.050) log (0.235) reprint (2.309) log (0.237) log (0.235) reprint (2.309) log (0.237) log (0
Kalhori (logistic) /	(0.550), rengen (-0.255), previous imprisonment (2.556), age (0.257), area (0.8655), fil v (0.751)
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))
	Unclear. No constant term provided. Here are the predictor (OR): Mediastinal shift (2.1), average smear score (1.5), extensive lesions
Keane / 1997	(3.6), any previous treatment (2.3), cavities (1.7) , weight (0.98)
	dense layer and 4 units for the negative layer
Killian / 2019	
	One point for each parameter: septic shock, HIV with CD4 < 200, creatinine > 140 (male) or >120 (female), P:F O2 ratio < 200. chest
Koegelenberg / 2015	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 methovynhanyl) provionic acid (1.3)
Luics / 201 /	Written as predictor (OR): New TB with 1+ smear grade (5.78) New TB with 2+ smear grade (2.60) New TB with 3+ smear grade (1.60)
	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
	(33.44), $111 v + 100 on AK1$ with CD4 200-550 (05.98), $111 v + 100 on AK1$ with CD4 50-200 (0.94), $111 v + 100 on AK1$ with CD4 $<50(49.20) HIV+ diagnosed after TB with CD4>500 (1.05) HIV+ diagnosed after TB with CD4 350-500 (2.40) HIV+ diagnosed after TB$
	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-
Madan / 2018	$34 (9.41)$, Female $35-44 (1.75)$, Female $\geq 45 (4.49)$, Male $15-24 (10.63)$, Male $25-34 (2.74)$, Male $35-44 (2.9)$, Male $\geq 45 (3.96)$
Mburu / 2018 -	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
Failure	hba1c, regimen, age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine, respectively
10010	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,
Mburu / 2018 -	62.47, 62.63, 61.65, 55.62, 59.21, 54.48 for hba1c, regimen, creatinine, BMI, BUN, weight, age, random blood glucose, HIV positive
Nguyen (general non)	$\frac{1}{100} = \frac{1}{100} = \frac{1}$
/ 2018	4 + 4*[Miliary TB] + 6*[TB-CXR] + 6*[HIV positive] + 6*[HIV unknown]
Nguyen (TB-DM) /	16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[IDU] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliarv TB] + 10*[Chronic kidney failure] + 10*[Chronic kidney fa
2019	6*[AFB positive smear] + 24*[Positive HIV]
Nguyen (TB-HIV) /	Prognostic score: 5*[Age >= 65] + 12*[Resident of LTCF] + 9*[Meningeal TB] + 6*[abnormal CXR] + 9*[diagnosis confirmed with
2019	
2018	positive culture or NAA + 10*[culture not converted or unknown]

Page 61 of 62

BMJ Open

	1.899108 [diagnosis commined with positive culture of NAA] + 2.186505 [culture not converted of 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 field 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 infection or space-occupying lesions] + 5* [Malignancies]

2
3
4
5
6
7
/
8
9
10
11
12
13
14
14
15
16
17
18
19
20
∠ I 22
22
23
24
25
26
27
27
20
29
30
31
32
33
34
25
22
36
37
38
39
40
41
יד גע
42
43
44
45
46
47
48
<u>10</u>
77
50
51
52
53
54
55
56
50
5/
58
59
60

Section/topic#Checklist itemReported on page #TITLETITLETitle1Identify the report as a systematic review, meta-analysis, or both.1ABSTRACTStructured summary2Provide 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.3INTRODUCTIONRationale3Describe the rationale for the review in the context of what is already known.5-6Objectives4Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).Supplemental File 2METHODSProtocol and registration5Indicate 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. 7Eligibility criteria orieria6Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., vears considered, language, publication status) used as criteria for eligibility, giving rationale.7-8Information sources7Describe 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				
TITLE Title 1 Identify the report as a systematic review, meta-analysis, or both. 1 ABSTRACT 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 1 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 1 Abstract and p. 7 Protocol and 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., vears 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	Section/topic	#	Checklist item	Reported on page #
Title 1 Identify the report as a systematic review, meta-analysis, or both. 1 ABSTRACT	TITLE			
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 INTRODUCTION 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 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., dtabases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. 8	Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
Structured summary2Provide 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.3INTRODUCTIONRationale3Describe the rationale for the review in the context of what is already known.5-6Objectives4Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).Supplemental File 2METHODSEndicate 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. 7Eligibility criteria6Specify 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.7-8Information sources7Describe 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	ABSTRACT		·	<u>.</u>
INTRODUCTIONRationale3Describe the rationale for the review in the context of what is already known.5-6Objectives4Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).Supplemental File 2METHODSProtocol and registration5Indicate 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. 7Eligibility criteria6Specify 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.7-8Information sources7Describe 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	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
Rationale3Describe the rationale for the review in the context of what is already known.5-6Objectives4Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).Supplemental File 2METHODS9Protocol and registration5Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, 	INTRODUCTION			
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	Rationale	3	Describe the rationale for the review in the context of what is already known.	5-6
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	Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Supplemental File 2
Protocol and registration5Indicate 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. 7Eligibility criteria6Specify 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.7-8Information sources7Describe 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	METHODS			
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	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
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	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
Search8Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.Supplemental file 3	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	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 process10Describe 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 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 items11List 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	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 studies12Describe 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 	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	Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results14Describe 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	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 studies15Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).N/A	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	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	RESULTS			
Study selection17Give 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 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 characteristics18For 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	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

1	
2	
3	
4	
5	
6	
7	
/ 0	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
10	
10	
19	
20	
21	
22	
23	
24	
25	
26	
20	
27	
20	
29	
30	
31	
32	
33	
34	
35	
36	
37	
20	
20	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
70 /0	
49 50	
50	
51	
52	
53	
54	
55	

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
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
21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta- regression [see Item 16]).	N/A
	·	
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
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
26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	19
		·
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
	19 20 21 22 23 24 25 26 27	19 Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). 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. 21 Present results of each meta-analysis done, including confidence intervals and measures of consistency. 22 Present results of any assessment of risk of bias across studies (see Item 15). 23 Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta- regression [see Item 16]). U One sensitivity of subgroup analyses, meta- regression [see Item 16]). 24 Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). 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). 2 27 Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

BMJ Open

A systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults

Journal:	BMJ Open
Manuscript ID	bmjopen-2020-044687.R2
Article Type:	Original research
Date Submitted by the Author:	09-Feb-2021
Complete List of Authors:	Peetluk, Lauren; Vanderbilt University School of Medicine, Epidemiology Ridolfi, Felipe; Instituto Nacional de Infectologia Evandro Chagas Rebeiro, Peter; Vanderbilt University School of Medicine, Epidemiology; Vanderbilt University School of Medicine, Division of Infectious Diseases Liu, Dandan; Vanderbilt University School of Medicine, Biostatistics Rolla, Valeria; Instituto Nacional de Infectologia Evandro Chagas Sterling, Timothy; Vanderbilt University School of Medicine, Division of Infectious Diseases
Primary Subject Heading :	Infectious diseases
Secondary Subject Heading:	Global health, Patient-centred medicine, Public health
Keywords:	Tuberculosis < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, STATISTICS & RESEARCH METHODS

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

reliez oni

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults

Lauren S. Peetluk, MPH,¹ Felipe M. Ridolfi, MD, MSc,² Peter F. Rebeiro, PhD, MHS,^{1,3} Dandan

Liu, PhD,⁴ Valeria C. Rolla, MD, PhD,² Timothy R. Sterling, MD³

¹Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine,

Nashville, Tennessee, USA

²Instituto Nacional de Infectologia Evandro Chagas (INI) – Fiocruz, Rio de Janeiro, Brazil

³Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of

Medicine, Nashville, TN, USA

⁴Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA

teliezon

Corresponding author: Lauren S. Peetluk, MPH A2209 Medical Center North 1161 21st Avenue South Nashville, TN 37203 E-mail: lauren.s.peetluk@vanderbilt.edu

Word count main text: 3617

BMJ Open

ABSTRACT

Objective: To systematically review and critically evaluate prediction models developed to predict tuberculosis (TB) treatment outcomes among adults with pulmonary tuberculosis. **Design:** Systematic review

Data sources: PubMed, Embase, Web of Science, and Google Scholar were searched for studies published January 1, 1995 - January 9, 2020.

Study selection and data extraction: Studies that developed a model to predict pulmonary TB treatment outcomes were included. Study screening, data extraction, and quality assessment were conducted independently by two reviewers. Study quality was evaluated using the Prediction model Risk Of Bias Assessment Tool (PROBAST). Data were synthesized with narrative review and in tables and figures.

Results: 14,739 articles were identified, 536 underwent full-text review, and 33 studies presenting 37 prediction models were included. Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=9, 25%). Most models (n=29, 78%) measured discrimination (median c-statistic=0.75; interquartile range: 0.68-0.84), and 17 (46%) reported calibration, often the Hosmer-Lemeshow test (n=13). Nineteen (51%) models were internally validated, and six (16%) were externally validated. Eighteen studies (54%) mentioned missing data, and of those, half (n=9) used complete case analysis. The most common predictors included age, sex, extrapulmonary TB, body mass index (BMI), chest x-ray results, previous TB, and HIV. Risk of bias varied across studies, but all studies had high risk of bias in their analysis.

Conclusions: TB outcome prediction models are heterogeneous with disparate outcome definitions, predictors, and methodology. We do not recommend applying any in clinical settings

without external validation, and encourage future researchers adhere to guidelines for developing and reporting of prediction models.

Registration: The study was registered on the international prospective register of systematic reviews PROSPERO (CRD42020155782)

to pretterien ont

ARTICLE SUMMARY:

Strengths and limitations

- Prediction models for tuberculosis treatment outcomes have the potential to inform interventions or treatment management protocols to promote cure among tuberculosis patients at the greatest risk of unsuccessful treatment outcomes, but the methods and clinical utility of existing models had not been formally evaluated.
- This was the first systematic review of prediction models for tuberculosis treatment outcomes.
- The review used a comprehensive search strategy, conducted thorough bias assessment with the Prediction Model Risk of Bias Assessment Tool (PROBAST) tool, and offers recommendations for future model development and validation studies for predicting tuberculosis treatment outcomes.
- Evidence synthesis and quality assessment were limited by incomplete reporting in primary studies, as well as heterogeneities in study populations, such as multi-drug resistance and age.
- External validation studies or studies written in languages other than English, Spanish,
 Portuguese, or French were excluded.

BACKGROUND

Tuberculosis (TB) is one of the top ten causes of death worldwide and a leading cause of death from an infectious disease. In 2018, 10 million people developed TB and 1.45 million people died from it globally, despite widespread availability of curative treatment.[1] Global treatment success was 85% for all new and relapse TB patients in 2018. For HIV-associated TB, it was 75%. These proportions are lower than the End TB Strategy target of \geq 90% treatment success.[2]

Heeding early recognition that *Mycobacterium tuberculosis* develops resistance rapidly in response to single-drug therapy, TB has been treated with combination regimens for more than 50 years.[3] Aside from weight-based dosing, the World Health Organization (WHO) and other TB guidelines authorities recommend a standardized approach for treatment of almost all TB patients.[4–6] The current recommendation for drug-susceptible TB includes 2 months of isoniazid, rifampin, pyrazinamide, and ethambutol, followed by 4 months of isoniazid and rifampin.

Due to the long duration of TB treatment, it would be beneficial to understand early predictors of unsuccessful TB treatment outcomes to identify patients needing tailored treatment approaches, such as directly observed therapy (DOT) or extended treatment course. Research suggests that individual characteristics, such as HIV, age, undernutrition, diabetes, TB disease severity, extrapulmonary TB, history of TB, adherence, alcohol use, and adverse drug reactions, are associated with unsuccessful TB treatment outcomes, but results vary by setting and patient population.[7–10]

Prediction models, defined as any combination or equation of two or more predictors to estimate an individualized probability of a specific endpoint within a defined period of time, are

increasingly common in TB research.[11] The large number of recent prediction models for TB outcomes highlights the common desire to identify TB patients at greatest risk of an unsuccessful treatment outcome. However, to date, there has not been a formal synthesis or quality assessment of existing prediction models for TB treatment outcomes, which is essential to determine whether they should be used to inform care and may help guide development of future models. Thus, we conducted a systematic review to identify, describe, compare, and synthesize clinical prediction models designed to predict TB treatment outcomes among persons with pulmonary

Topper terren on t

TB.

METHODS AND ANALYSIS

All steps of the systematic review were carried out according to guidelines set by Cochrane Prognosis Methods Group (PMG) and PROGnosis RESearch Strategy (PROGRESS).[12–14] Reporting adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (**Supplemental File 1**). This study was pre-registered on Open Science Framework (OSF) (https://osf.io/rz3wp) and the international prospective register of systematic reviews (PROSPERO; CRD42020155782).

Study eligibility criteria

The review question was defined according to the PICOTS (Population, Intervention, Comaparator, Outcomes, Timing, Setting) framework (**Supplemental File 2**). In brief, the goal was to identify prognostic models developed to predict TB treatment outcomes among pulmonary TB cases. The main endpoint was unsuccessful TB treatment outcome, defined by the WHO as the combination of death, treatment failure, loss to follow-up, and/or not evaluated, as compared to successful TB treatment outcome, defined as the combination of cure or treatment completion (**Table 1**) [15]. Loss to follow-up was sometimes referred to as default or treatment abandonment.

Inclusion criteria were: 1) prognostic model studies with or without external validation[16]; 2) study population included adult, drug-susceptible, pulmonary, TB cases; 3) written in English, Spanish, Portuguese, and French; 4) published between January 1, 1995 and January 9, 2020; 5) treatment outcome was one of the following: cure, treatment completion, death, treatment failure, loss to follow-up, or not evaluated.

Exclusion criteria were: 1) predictive value of more than one variable was evaluated but not combined in a prediction model; 2) study population was only multi-drug resistant (MDR) TB cases, only extrapulmonary TB cases, or only children (< 18 years-old); 3) outcome was

BMJ Open

evaluated during treatment such as: two-month smear/culture conversion, acquired resistance, adverse events, quality of life; 4) long-term outcomes, such as relapse, recurrence, or post-treatment mortality.

The decision to include only articles in English, Spanish, Portuguese, and French was based on study team capabilities. The dates reflect modern TB treatment practice; first-line TB treatment regimens were not available until the early 1990s.[17,18] Articles that included a combination of drug-susceptible and drug-resistant cases, or a combination of children and adults were included.

Search strategy and selection criteria

The following electronic databases were searched on January 9, 2020: PubMed, Embase, Web of Science, and the first 200 references from Google Scholar. This combination of databases achieved best overall recall for systematic reviews in a recent study.[19] Clinicaltrials.gov and retractiondatabase.org were also searched for unpublished research. Reference lists of retrieved articles were checked to identify eligible studies.

Search terms relating to the "prediction model" component of the search were adapted from a PubMed search strategy that captured prediction model studies with sensitivity of 98%.[20] That component was combined with terms relating to TB treatment outcomes. The search strategy, developed in PubMed, was adapted for all other databases with assistance from a reference librarian (**Supplemental File 3**).

Article selection was conducted in three stages. The first stage was automatic deduplication and title screening, carried out using *revtools* in RStudio (version 1.2).[21] Remaining articles were imported into Covidence, a web-based software platform that streamlines systematic reviews, where abstracts (Stage 2) and full text (Stage 3) were manually screened.[22] Stages 2 and 3 were carried out by two independent reviewers (LSP and FMR).

Page 10 of 64

Discordance was discussed between reviewers, and if consensus was not reached, a third party arbitrated (one of TRS, VCR, PFR, DL). In stage 3, reasons for exclusion were documented according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Data analysis

 Data from selected studies were recorded using a database designed in REDCap (Vanderbilt University).[23,24] Data extraction was informed by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the Prediction Model Risk of Bias Assessment Tool (PROBAST).[16,25,26] CHARMS checklist and PROBAST are in **Supplemental Files 4 and 5**, respectively.

Quality assessment and applicability of included studies was assessed using PROBAST by dual independent review.[16,26] PROBAST was specifically designed to assess risk of bias of prediction model studies, which included identifying deficiencies in study design, conduct, or analysis that led to inaccurate estimates of predictive performance. PROBAST has 4 domains: participants, predictors, outcome, and analysis with 20 total signaling questions. Each question was answered on the scale: yes, probably yes, no, probably no, no information. Domains were scored as low, high, and unclear risk of bias. PROBAST also guides assessment of applicability of participants, predictors, and outcomes from each included study to the review question.

Results were summarized narratively and in tables and figures. Meta-analysis was not possible due to lack of external validation and use of disparate predictors, outcome definitions, and modeling methods. For studies that presented multiple models with the same set of predictors and outcomes, but different methods, the best-performing method was included in data synthesis. For studies presenting multiple models with different sets of predictors (i.e. baseline data vs. longitudinal data), the model developed using only baseline data was included. If studies developed multiple models for different outcomes or with different populations, all models were

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml
BMJ Open

included. To further evaluate the impact of study population heterogeneities on prediction model performance, we additionally examined results after stratifying studies by inclusion/exclusion of MDR and younger age groups.

Patient and public involvement

Neither patients nor the public were involved in the design, conduct, or reporting of the research, as it was not feasible or appropriate for this systematic review. The study protocol is publicly available at https://osf.io/rz3wp.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Study selection

The search identified 14,739 unique studies. After excluding irrelevant titles, 6,426 abstracts were screened, 536 articles underwent full-text review, and 33 model development studies presenting 37 prediction models were included (**Figure 1**).

Study characteristics

Of the 33 studies, most were retrospective cohorts (n=25, 76%), three (9%) were prospective cohort studies, two (6%) were case-control studies, and three (9%) were nested case-control studies. Data from nearly half of studies (n=16, 48%) were collected from surveillance systems; eleven (33%) studies used a data collection form developed specifically for their study and six studies (18%) extracted data from medical records. Median sample size was 803 (interquartile range (IQR): 291-4167). Full details on included studies are in **Table 2**.

Thirteen studies (41%) took place in Asia, eight (25%) in Africa, six (19%) in Europe, four (12%) in North America, and one (3%) included sites in Europe and Argentina. Fewer than half (n=14, 45%) took place in high-burden TB settings.¹ One study did not report study location. (**Tables 2 and 3**).

Reporting of population characteristics varied by study (**Table 4**). Among 18 studies that reported a measure of central tendency (mean or median) for age, the median of those measures was 41 years (IQR: 37-49). Of 17 studies that reported the minimum age of participants, seven (41%) had a minimum age of 15, one (6%) had a minimum age of 16, one (6%) had a minimum age of 17, and the remainder had minimum age of 18. Eighteen studies reported including persons living with HIV (PLWH); 5 of these included only TB/HIV patients. Thirteen studies reported including persons with diabetes; one of which included only TB/DM. Eight studies

BMJ Open

reported including some participants with MDR, though prevalence of MDR was low in all studies. Ten studies included only hospitalized patients, and in 14 studies, all participants were on directly observed therapy (DOT).

Model characteristics

Model outcomes included death (n=16, 43%), treatment failure (n=6, 16%), default (n=6, 16%) or a composite outcome (n=8, 23%) (**Tables 2 and 5**). The complete outcome definition for all included studies is in **Supplemental File 6**.

Most models were developed using clinical/epidemiologic predictors (n=34, 92%), two (6%) used multiple biomarkers, and one (3%) used adherence data. The most common candidate predictors were age, sex, extrapulmonary TB, smear result, BMI, x-ray findings, and previous TB. The most common predictors retained in the final models were age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV (**Figure 2**).

Only three models (8%) used survival analysis; most models used logistic regression (n=29, 78%) and five (14%) used a machine learning approach. More than half of studies (n=19, 51%) considered variables for inclusion in the multivariable model based on unadjusted associations with the outcome. Model building methods varied widely between models (**Table 5**).

Only 19 (51%) models were internally validated, including ten (53%) split-sample validation, five (26%) bootstrap resampling, and four (21%) cross-validation. Six (16%) models were externally validated. Many models (n=30, 81%) reported discrimination with c-statistic (concordance statistic) or area under the receiver operating characteristic (AUROC), which are equivalent and quantify the ability of the model to distinguish between patients who do and do not develop an outcome. Only 17 (46%) reported calibration, the agreement between observed

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

and predicted outcomes. Most studies assessed calibration with Hosmer-Lemeshow tests (n=13, 77%); only two studies provided a calibration plot, the preferred reporting method for prediction model studies,[16,27,28] and one reported the calibration slope (**Table 2**). Models were presented a variety of ways, the most common of which was a weighted risk score (n=16, 43%); details on model presentation are in **Supplemental File 7**.

Quality assessment

Grading of PROBAST signaling questions is summarized in **Figure 3**, and the summary risk of bias for the participants, predictors, outcome, and analysis domains and assessment of applicability are shown in **Figure 4**. More than half of the studies were at low risk of bias for the population and outcomes domains, but all studies were at high risk of bias in the analysis domain.

Common sources of population bias included use of non-nested case-control design[29,30], nested case-control design without proper estimation of baseline risk,[31,32] or inappropriate inclusion/exclusion criteria.[33,34] Sources of predictor bias included lack of standardized assessment of key predictors (i.e. HIV, diabetes, chest x-ray scoring)[9,29,31,34–36] or timing of data collection/availability that would limit the intended use of the model.[9,29,37] Within the outcomes domain, sources of bias included subjective[35] or non-standard[32,38] outcome measures and inconsistent outcome ascertainment.[29]

Bias in the analysis domain was widespread. More than half of the models included were likely overfit due to low events per variable (EPV) ratios (**Table 5**). Only 6 studies handled continuous and categorical variables appropriately (i.e., didn't dichotomize continuous variables, considered non-linearity of continuous variables).[31,39–43] Most studies used complete caseanalysis or did not mention missing data; no study used multiple imputation in their main

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

analysis. One study with low amounts of missing data (<5%) conducted sensitivity analysis with multiple imputation.[44] A different study excluded only two people out of a total sample size of 1007 with missing data, which would have little impact on model performance.[45] Fewer than half (n=14) of studies avoided univariable predictor selection, and only three studies used survival analysis, appropriately accounting for censoring.[36,45,46] Performance measures were appropriately reported (i.e. calibration assessed with plot and discrimination assessed with c-statistic/AUROC) in three studies.[41,44,47] Only two studies estimated optimism (degree to which data are overfit) or accounted for potential overfitting with penalization of model parameters.[35,41] Ten studies appropriately presented their model with model coefficients or nomograms, which prevents bias from rounding or transforming model coefficients to generate a risk score.[30,33,35,37,38,45,47–55]

About half of the models (n=19, 51%) were applicable to the review question in all domains. However, unclear reporting of target population or predictor and outcome definitions limited assessment of applicability for several studies.[38,49,50,56,57] Additionally, studies that included only hospitalized patients with specific laboratory parameters may not be routinely available in the clinical setting.[39,40,42] Results from analyses stratified by inclusion of patients with MDR and minimum age <18 are presented in **Supplemental File 8**.

DISCUSSION

In this comprehensive, systematic review of prediction models for pulmonary TB treatment outcomes, we identified 33 model development studies presenting 37 prediction models. Although diagnostic prediction models for prevalent TB were previously systematically reviewed, this is the first review of TB treatment outcomes.[58] The included prediction models were developed for predicting death, treatment failure, default, or a composite unfavorable outcome during TB treatment. Most models reported good performance (c-statistic/AUROC>0.7), but all were evaluated to have high risk of bias due to poor reporting, exclusion of missing data, weak methodologic approaches, lack of calibration assessment, and limited validation. Population heterogeneities, such as differences in inclusion/exclusion of individuals with MDR and younger ages, and varying predictor and outcome definitions limited comparisons between models.

More than half of the models included in the review were developed in low burden TB settings, and none were developed specifically in South America. Prediction of TB treatment outcome is especially important in high burden TB settings, where resources may be limited, and risk assessment can guide resource allocation toward patients who need the most involved care.

Common risk factors included in the models were consistent with well-established risk factors for poor TB treatment outcomes, including age, sex, HIV, extrapulmonary TB, baseline smear results, and previous TB treatment. Among studies that included PLWH, only three considered factors related to management/severity of HIV, such as receipt of antiretroviral therapy, CD4 cell count, or viral load, which likely impacted TB treatment outcomes.[40,46,51] Laboratory values or metabolic biomarkers, such as hemoglobin, hemoglobin A1c or random blood glucose, may also be associated with treatment outcome and worth considering as candidate predictors. There is increasing evidence that diabetes impacts TB treatment outcomes,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 17 of 64

BMJ Open

but caution is warranted about how to best define diabetes in the context of a prediction model to ensure consistency and reproducibility across studies.[59] Behavioral characteristics, such as tobacco use, alcohol use, and drug use were rarely included in final prediction models and are difficult to collect objectively, suggesting their role in prediction models for TB treatment outcomes may be limited.

Additionally, several studies excluded participants with HIV, diabetes, extrapulmonary TB, or MDR TB, because these factors negatively influence treatment outcomes. However, careful consideration should be given to inclusion/exclusion criteria in prediction model studies, given that information should be available at the time of intended model use, which may not always hold for these aforementioned factors.[60] This is especially questionable for MDR, given that conventional drug-susceptibility testing results are not available for several weeks after TB diagnosis; though more recent advances in rapid molecular methods such as GeneXpert or line-probe assays offer rapid screening.[61]

TB researchers should thoughtfully consider how to appropriately handle complexities of censoring and competing risks in TB outcomes research. Only three studies in this review used survival analysis, despite the long duration of TB treatment outcome assessment and relatively high rates of losses to follow-up across studies, and no studies considered competing risks, such as death due to other causes.[62] Losses to follow-up were frequently excluded, which can lead to selection bias.

Though all included studies were at high risk of bias in the analysis domain, we want to highlight two studies with some exemplary characteristics.[41,44] Pefura-Yone et al.[41] provide clear explanations of study design, inclusion/exclusion criteria, and data collection procedures; TB diagnosis and treatment outcome definitions were standard.[63] Non-linearity of continuous

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

variables was considered with restricted cubic splines, and no continuous variables were categorized or dichotomized; the final model includes four predictors that are easy to collect and routinely assessed in most TB control programs, especially those in high burden settings. The performance of the model was internally validated with bootstrap validation, and the discrimination (c-statistic=0.808) was corrected for optimism. Model calibration was presented graphically with calibration plots. The final model was presented as a nomogram with instructions for use, which facilitates use in external validation studies. Gupta-Wright and colleagues developed and externally validated a clinical risk score to predict mortality in highburden, low-resource settings.⁴³ They used clinical trial data with very low amounts of missing data for model development, and externally validated the clinical risk score with data collected independently from two other studies (a clinical trial and a prospective cohort). Given high amounts (42%) of missing data in the validation cohort, they conducted sensitivity analysis using multiple imputation for missing data; the c-statistic differed slightly between complete case and multiple-imputation analyses in the validation cohort (0.68 vs. 0.64). Candidate predictors were based on *a priori* clinical knowledge, previous literature, and required variables were objective, reproducible, and available in low-resource settings, consistent with recommended approaches. [26,60,64] Additionally, they reported model performance with the c-statistics and calibration plots for development and validation cohorts, and reported results according to TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidance.[27,28] Regardless, each of these models requires external validation prior to use in clinical practice.

There are several limitations of this study. Data extraction was subject to reporting in the primary study, which varied widely and was often incomplete, leading to challenges evaluating

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 19 of 64

BMJ Open

differences in model performance due to heterogeneities in study populations. Additionally, though most studies reported discrimination, few presented a calibration curve, arguably the most important measure of model performance, further inhibiting assessment and comparison of model performance.[28,65] We did not include external validation studies, which is an essential step for translation to clinical practice. However, several studies in the review did not include the full model equation, which impedes their ability to be externally validated. Upon searching for studies that externally validated prediction models in this review, we found three studies[66–68] that evaluated the same model (TBscore).[36] Briefly, these studies evaluated the ability of TBscore to monitor treatment response in a new setting[66], refined the instrument (TBccoreII) using exploratory factor analysis[67], and then evaluated TBscoreII for use in patients with TB/HIV.[68] To our knowledge, no other studies included in the review were externally validated by other sources. Finally, we excluded 10 studies that were not available in English, Spanish, Portuguese, or French; all abstracts were available in English, and none reported model performance metrics, so they likely would have been excluded for different reasons regardless.

The findings of this review not only serve as a comprehensive overview of existing TB outcome prediction models but can act as a resource for future model development and validation of prediction models for TB treatment outcomes. We encourage researchers to focus future TB outcome prediction models on easily collected and readily available predictors that are widely generalizable. We highlight age, sex, extrapulmonary TB, BMI, chest x-ray results, previous TB, and HIV as common predictors of TB treatment outcomes. Additionally, when building a new prediction model, it is recommended to first prune the set of considered predictors based on expert opinion and previous literature, rather than univariable analysis or variable selection processes[26,60,64] Future model development or validation studies should

BMJ Open

adhere to the TRIPOD guidelines, which provide a 22-item checklist and aims to improve the reporting of prediction model development studies.[27,28] We also encourage researchers consider PROBAST criteria to limit bias in design and conduct of prognostic studies.

Prediction models are an important tool in TB management. They can lay the foundation for future impact studies by providing risk estimation to target novel treatment approaches, resource allocation, or intensive case management towards patients who are least likely to achieve cure and most likely to benefit from intervention, especially in high-burden and lowresources areas. Use of prediction models can potentially help guide tuberculosis treatment practices to achieve the End TB Strategy goal of >90% treatment success, but methodologic rigor and detailed reporting must be improved. Though our findings suggest that none of the existing models are ready for clinical application without extensive external validation, we hope of guu. they direct future researchers to make use of guidelines for development and reporting of prediction models.

BMJ Open

FOOTNOTES

Ethics approval: Not required.

Transparency statement: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported. No important aspects of the study have been omitted, and any discrepancies from the study as planned were explained.

Contributorship: LSP conceptualized the research question, designed the protocol, and drafted the manuscript. LSP and FMR screened studies. FMR, PFR, DL, VCR and TRS provided feedback on the research design, original protocol, and revised successive drafts of the manuscript. All authors approved the final version of the manuscript.

Funding: This work was supported by the National Center for Advancing Translational Sciences [CTSA Award No. TL1TR000447 to L.S.P.] and the National Institutes of Allergy and Infectious Diseases [F31AI152614-01A1 to L.S.P],. Its contents are solely the responsibility of the authors and do not necessarily represent the official views the National Center for Advancing Translational Sciences or the National Institutes of Health.

Competing interests: None declared.

Data sharing: The study protocol is available online at https://osf.io/rz3wp. Most included studies are publicly available. Additional data and code are available upon request.

Exclusive license: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.

REFERENCES

- 1 Global Tuberculosis Report 2019. Geneva: : World Health Organization 2019. Licence: CC BY-NC-SA 3.0 IGO.
- 2 The END TB Strategy. Geneva: : World Health Organization 2015.
- 3 Kerantzas CA, Jacobs WR. Origins of Combination Therapy for Tuberculosis: Lessons for Future Antimicrobial Development and Application. *mBio* 2017;**8**:e01586-16. doi:10.1128/MBIO.01586-16
- 4 Nahid P, Dorman SE, Alipanah N, *et al.* Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clinical Infectious Diseases* 2016;**63**:e147–95. doi:10.1093/cid/ciw376
- 5 Guildelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. Geneva: World Health Organization 2017. Licence: CC BY-NC-SA 3.0 IGO. doi:WHO/HTM/TB/2017.05
- 6 WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: : World Health Organization 2019. Licence: CC BY-NC-SA 3.0 IGO.
- 7 Vasankari T, Holmström P, Ollgren J, et al. Risk factors for poor tuberculosis treatment outcome in Finland: A cohort study. BMC Public Health 2007;7:1–9. doi:10.1186/1471-2458-7-291
- 8 Ramachandran G, Agibothu K. Factors Influencing Tuberculosis Treatment Outcome in Adult Patients Treated with Thrice-Weekly Regimens. 2017;**61**:1–12.
- 9 Abdelbary BE, Garcia-Viveros M, Ramirez-Oropesa H, et al. Predicting treatment failure, death and drug resistance using a computed risk score among newly diagnosed TB patients in Tamaulipas, Mexico. Epidemiology and Infection 2017;145:3020–34. doi:10.1017/S0950268817001911
- 10 Torres NMC, Rodríguez JJQ, Andrade PSP, *et al.* Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis. *PLoS ONE* 2019;14:1–24. doi:10.1371/journal.pone.0226507
- 11 Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Medicine 2013;10:e1001381. doi:10.1371/journal.pmed.1001381
- 12 Riley R, Ridley G, Williams K, *et al.* Prognosis research: towards evidence-based results and a Cochrane methods group. 2014;**60**:863–5.
- 13 Moons KG, Hooft L, Williams K, *et al.* Implementing systematic reviews of prognosis studies in Cochrane. *The Cochrane database of systematic reviews* 2018;**10**:ED000129. doi:10.1002/14651858.ED000129

2	
3	
4	
5	
6	
0	
7	
8	
9	
10	
10	
11	
12	
13	
1/	
14	
15	
16	
17	
18	
10	
19	
20	
21	
22	
22	
25	
24	
25	
26	
27	
27	
28	
29	
30	
31	
27	
32	
33	
34	
35	
26	
50	
37	
38	
39	
10	
40	
41	
42	
43	
44	
1	
45	
46	
47	
48	
10	
+7	
50	
51	
52	
53	
)) 	
54	
55	
56	
57	
57	
28	
59	
60	

- 14 Debray TPA, Damen JAAG, Snell KIE, *et al.* A guide to systematic review and metaanalysis of prediction model performance. *BMJ (Online)* 2017;**356**. doi:10.1136/bmj.i6460
- 15 Definitions and reporting framework for tuberculosis 2013 revision. In: *Annex 2, TB case and treatment outcome definitions*. Geneva: : World Health Organization 2014.
- 16 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**:51–8. doi:10.7326/M18-1376
- 17 Iseman MD. Tuberculosis therapy: past, present and future. *Eur Resp J* 2002;**20**:87s–94s. doi:10.1183/09031936.02.00309102
- 18 Council STSMR. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis: the results up to 30 months. *Tubercle* 1981;:95– 102.
- 19 Bramer WM, Rethlefsen ML, Kleijnen J, *et al.* Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews* 2017;6:1–12. doi:10.1186/s13643-017-0644-y
- 20 Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS ONE 2012;7:3–8. doi:10.1371/journal.pone.0032844
- 21 Westgate MJ. revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods* 2019;**10**:606–14. doi:10.1002/jrsm.1374
- 22 Veritas Health Innovation, Melbourne A. Covidence systematic review sofware. Covidence. 2016. doi:10.1016/j.carbon.2012.08.062
- 23 Harris PA, Taylor R, Minor BL, *et al.* The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019;**95**:103208. doi:10.1016/j.jbi.2019.103208
- 24 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009;**42**:377–81. doi:10.1016/j.jbi.2008.08.010
- 25 Moons KGM, de Groot JAH, Bouwmeester W, *et al.* Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine* 2014;**11**. doi:10.1371/journal.pmed.1001744
- 26 Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Annals of Internal Medicine* 2019;**170**:W1–33. doi:10.7326/M18-1377

27 Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine* 2015;**162**:W1–73. doi:10.7326/M14-0698

- 28 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. 2015;**162**. doi:10.7326/M14-0697
- 29 Cherkaoui I, Sabouni R, Ghali I, *et al.* Treatment default amongst patients with tuberculosis in urban Morocco: Predicting and explaining default and post-default sputum smear and drug susceptibility results. *PLoS ONE* 2014;**9**. doi:10.1371/journal.pone.0093574
- 30 Keane VP, De Klerk N, Krieng T, et al. Risk factors for the development of non-response to first-line treatment for tuberculosis in Southern Vietnam. International Journal of Epidemiology 1997;26:1115–20. doi:10.1093/ije/26.5.1115
- 31 Chang KC, Leung CC, Tam CM. Risk factors for defaulting from anti-tuberculosis treatment under directly observed treatment in Hong Kong. *International Journal of Tuberculosis and Lung Disease* 2004;**8**:1492–8.
- 32 Chee CBE, Boudville IC, Chan SP, *et al.* Patient and disease characteristics, and outcome of treatment defaulters from the Singapore TB control unit A one-year retrospective survey. *International Journal of Tuberculosis and Lung Disease* 2000;**4**:496–503. doi:NA
- 33 Luies L, Reenen M Van, Ronacher K, *et al.* Predicting tuberculosis treatment outcome using metabolomics. *Biomarkers in Medicine* 2017;**11**:1057–67. doi:10.2217/bmm-2017-0133
- 34 Killian JA, Wilder B, Sharma A, *et al.* Learning to Prescribe Interventions for Tuberculosis Patients Using Digital Adherence Data. *KNOWLEDGE DISCOVERY AND DATA MINING* 2019;**NA**:2430–8. doi:10.1145/3292500.3330777
- 35 Belilovsky EM, Borisov SE, Cook EF, *et al.* Treatment interruptions among patients with tuberculosis in Russian TB hospitals. *International Journal of Infectious Diseases* 2010;**14**:e698–703. doi:10.1016/j.ijid.2010.03.001
- 36 Wejse C, Gustafson P, Nielsen J, *et al.* TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. *Scandinavian Journal of Infectious Diseases* 2008;**40**:111–20. doi:10.1080/00365540701558698
- 37 Nguyen DT, Graviss EA. Development and validation of a risk score to predict mortality during TB treatment in patients with TB-diabetes comorbidity. *BMC Infectious Diseases* 2019;**19**:10. doi:10.1186/s12879-018-3632-5
- 38 Kalhori SRN, Zeng X. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. *Lecture Notes in Engineering and Computer Science* 2009;**2179**:774–8.

1	
2	
3	
4	
5	
6	
0	
/	
8	
9	
10	
11	
12	
13	
1/	
15	
10	
16	
17	
18	
19	
20	
21	
22	
23	
24	
27	
25	
20	
2/	
28	
29	
30	
31	
32	
33	
34	
35	
26	
30	
3/	
38	
39	
40	
41	
42	
43	
44	
45	
46	
-10 /17	
4/	
48	
49	
50	
51	
52	
53	
54	
55	
56	
50	
5/	
58	
59	
60	

- 39 Horita N, Miyazawa N, Yoshiyama T, et al. Poor performance status is a strong predictor for death in patients with smear-positive pulmonary TB admitted to two Japanese hospitals. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2013;107:451–6. doi:10.1093/trstmh/trt037
- 40 Koegelenberg CFN, Balkema CA, Jooste Y, *et al.* Validation of a severity-of-illness score in patients with tuberculosis requiring intensive care unit admission. *South African Medical Journal* 2015;**105**:389–92. doi:10.7196/SAMJ.9148
- 41 Pefura-Yone EW, Kuaban C, Assamba-Mpom SA, *et al.* Derivation, validation and comparative performance of a simplified chest X-ray score for assessing the severity and outcome of pulmonary tuberculosis. *Clinical Respiratory Journal* 2015;**9**:157–64. doi:10.1111/crj.12112
- 42 Valade S, Raskine L, Aout M, *et al.* Tuberculosis in the intensive care unit: A retrospective descriptive cohort study with determination of a predictive fatality score. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2012;**23**:173–8. doi:10.1155/2012/361292
- 43 Wang Q, Han W, Niu J, *et al.* Prognostic value of serum macrophage migration inhibitory factor levels in pulmonary tuberculosis. *Respiratory Research* 2019;**20**:50. doi:10.1186/s12931-019-1004-3
- 44 Gupta-Wright A, Corbett EL, Wilson D, *et al.* Risk score for predicting mortality including urine lipoarabinomannan detection in hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation and external validation cohort study. *PLoS Medicine* 2019;**16**:1–20. doi:10.1371/journal.pmed.1002776
- 45 Zhang Z, Xu L, Pang X, *et al.* A Clinical scoring model to predict mortality in HIV/TB coinfected patients at end stage of AIDS in China: An observational cohort study. *BioScience Trends* 2019;**13**:136–44. doi:10.5582/bst.2018.01309
- 46 Podlekareva DN, Grint D, Post FA, *et al.* Health care index score and risk of death following tuberculosis diagnosis in HIV-positive patients. *International Journal of Tuberculosis and Lung Disease* 2013;**17**:198-206+i. doi:10.5588/ijtld.12.0224
- 47 Baussano I, Pivetta E, Vizzini L, *et al.* Predicting tuberculosis treatment outcome in a lowincidence area. *International Journal of Tuberculosis and Lung Disease* 2008;**12**:1441–8.
- 48 Costa-Veiga A, Briz T, Nunes C. Unsuccessful treatment in pulmonary tuberculosis: Factors and a consequent predictive model. *European Journal of Public Health* 2018;**28**:252–8. doi:10.1093/eurpub/ckx136
- 49 Niakan Kalhori SR, Nasehi M, Zeng XJ. A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion. *IAENG International Journal of Applied Mathematics* 2010;**40**:1–6.
- 50 Kalhori SRN, Zeng X-J. PREDICTING THE OUTCOME OF TUBERCULOSIS TREATMENT COURSE IN FRAME OF DOTS - From Demographic Data to Logistic

Regression Model. In: *Proceedings of the International Conference on Health Informatics*. SciTePress - Science and and Technology Publications 2009. 129–34. doi:10.5220/0001431401290134

- 51 Madan C, Chopra KK, Satyanarayana S, *et al.* Developing a model to predict unfavourable treatment outcomes in patients with tuberculosis and human immunodeficiency virus co-infection in Delhi, India. *PLoS ONE* 2018;**13**:e0204982. doi:10.1371/journal.pone.0204982
- 52 Nguyen DT, Jenkins HE, Graviss EA. Prognostic score to predict mortality during TB treatment in TB / HIV co-infected patients. *PLoS ONE* 2018;**13**:1–12. doi:10.1371/journal.pone.0196022
- 53 Nguyen DT, Graviss EA. Development and validation of a prognostic score to predict tuberculosis mortality. *Journal of Infection* 2018;77:283–90. doi:10.1016/j.jinf.2018.02.009
- 54 Pefura-Yone EW, Balkissou AD, Poka-Mayap V, *et al.* Development and validation of a prognostic score during tuberculosis treatment. *BMC Infectious Diseases* 2017;**17**:1–9. doi:10.1186/s12879-017-2309-9
- 55 Rodrigo T, Caylà JA, Casals M, *et al.* A predictive scoring instrument for tuberculosis lost to follow-up outcome. *Respiratory Research* 2012;**13**:1–9. doi:10.1186/1465-9921-13-75
- 56 Hussain OA, Junejo KN. Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Informatics for Health and Social Care* 2019;**44**:135–51. doi:10.1080/17538157.2018.1433676
- 57 Sauer CM, Sasson D, Paik KE, *et al.* Feature selection and prediction of treatment failure in tuberculosis. *PLoS ONE* 2018;**13**:1–14. doi:10.1371/journal.pone.0207491
- 58 Wyk SSV, Lin HH, Claassens MM. A systematic review of prediction models for prevalent pulmonary tuberculosis in adults. *Int J Tuberc Lung Dis*;**21**.
- 59 Huangfu P, Ugarte-Gil C, Golub J, *et al.* The effects of diabetes on tuberculosis treatment outcomes: an updated systematic review and meta-analysis. *The International Journal of Tuberculosis and Lung Disease* 2019;**23**:783–96. doi:10.5588/ijtld.18.0433
- 60 Steyerberg EW. *Clinical Prediction Models*. New York, NY: : Springer New York 2009. doi:10.1007/978-0-387-77244-8
- 61 Sharma SK, Dheda K. What is new in the WHO consolidated guidelines on drug-resistant tuberculosis treatment? The Indian journal of medical research. 2019;**149**:309–12. doi:10.4103/ijmr.IJMR_579_19
- 62 Wolbers M, Koller MT, Witteman JCM, *et al.* Prognostic models with competing risks methods and application to coronary risk prediction. *Epidemiology* 2009;**20**:555–61. doi:10.1097/EDE.0b013e3181a39056
- 63 National Tuberculosis Control Program. Manual for health personnel. Yaounde: 2012.

2
2
3
4
5
6
U -
7
8
9
10
10
11
12
13
11
14
15
16
17
18
10
19
20
21
22
~~ >>
23
24
25
26
20
27
28
29
30
21
31
32
33
34
25
55
36
37
38
20
72
40
41
42
 4٦
44
45
46
47
10
4ð
49
50
51
57
52
53
54
55
56
50
57
58
59
60
00

- 64 Royston P, Moons KGM, Altman DG, *et al.* Prognosis and prognostic research: Developing a prognostic model. *BMJ (Online)* 2009;**338**:1373–7. doi:10.1136/bmj.b604
- 65 Calster BV. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology* 2016;:10.
- 66 Janols H, Abate E, Idh J, *et al.* Early treatment response evaluated by a clinical scoring system correlates with the prognosis of pulmonary tuberculosis patients in Ethiopia: A prospective follow-up study. *Scandinavian Journal of Infectious Diseases* 2012;44:828–34. doi:10.3109/00365548.2012.694468
- 67 Rudolf F, Lemvik G, Abate E, *et al.* TBscore II: Refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. *Scandinavian Journal of Infectious Diseases* 2013;**45**:825–36. doi:10.3109/00365548.2013.826876
- 68 Wejse C, Patsche CB, Kühle A, *et al.* Impact of HIV-1, HIV-2, and HIV-1+2 dual infection on the outcome of tuberculosis. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2015;**32**:128–34. doi:10.1016/j.ijid.2014.12.015
- 69 Aljohaney AA. Mortality of patients hospitalized for active tuberculosis in King Abdulaziz University Hospital, Jeddah, Saudi Arabia. *Saudi Medical Journal* 2018;**39**:267–72. doi:10.15537/smj.2018.3.22280
- 70 Bastos HN, Osório NS, Castro AG, *et al.* A prediction rule to stratify mortality risk of patients with pulmonary tuberculosis. *PLoS ONE* 2016;**11**:1–14. doi:10.1371/journal.pone.0162797
- 71 Gupta-Wright A, Corbett EL, Wilson D, *et al.* Risk score for predicting mortality including urine lipoarabinomannan detection in hospital inpatients with HIV-associated tuberculosis in sub-Saharan Africa: Derivation and external validation cohort study. *PLoS Medicine* 2019;**16**:1–20. doi:10.1371/journal.pmed.1002776
- 72 Horita N, Miyazawa N, Yoshiyama T, *et al.* Development and validation of a tuberculosis prognostic score for smear-positive in-patients in Japan. *International Journal of Tuberculosis and Lung Disease* 2013;**17**:54–60. doi:10.5588/ijtld.12.0476
- 73 Podlekareva DN, Grint D, Post FA, *et al.* Health care index score and risk of death following tuberculosis diagnosis in HIV-positive patients. *The International Journal of Tuberculosis and Lung Disease* 2013;**17**:198–206. doi:10.5588/ijtld.12.0224
- 74 Wang Q, Han W, Niu J, *et al.* Prognostic value of serum macrophage migration inhibitory factor levels in pulmonary tuberculosis. *Respiratory Research* 2019;**20**:50. doi:10.1186/s12931-019-1004-3
- 75 Wejse C, Gustafson P, Nielsen J, *et al.* TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical

course. *Scandinavian Journal of Infectious Diseases* 2008;**40**:111–20. doi:10.1080/00365540701558698

- 76 Mburu JW, Kingwara L, Ester M, *et al.* Use of classification and regression tree (CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor tuberculosis treatment outcomes and associated risk factors. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* 2018;11:10–6. doi:10.1016/j.jctube.2018.01.002
- 77 Thompson EG, Du Y, Malherbe ST, *et al.* Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis* 2017;**107**:48–58. doi:10.1016/j.tube.2017.08.004
- 78 Chee CBE, Boudville IC, Chan SP, *et al.* Patient and disease characteristics, and outcome of treatment defaulters from the Singapore TB control unit A one-year retrospective survey. *International Journal of Tuberculosis and Lung Disease* 2000;4:496–503.
- 79 Rodrigo T, Caylà JA, Casals M, *et al.* A predictive scoring instrument for tuberculosis lost to follow-up outcome. *Respiratory Research* 2012;**13**:1–9. doi:10.1186/1465-9921-13-75
- 80 Kalhori SRN, Zeng X-J. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. In: *Lecture Notes in Engineering and Computer Science*. NA 2009. 774–8. doi:NA

1	
2	
3	
4	
5	
5 C	
0	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
27	
20	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
30	
10	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
52	
22	
54	
55	
56	
57	
58	
59	

Table 1. World Health Organization definition of treatment outcomes for TB pa	atients
---	---------

Outcome	Definition
	Completion of treatment without evidence of failure, but without
Treatment	documentation of a negative sputum smear or culture in the last month of
completion	treatment and/or on at least one previous occasion, either because tests
	were not done or because results are unavailable
Cure	Bacteriologic confirmation of a negative smear or culture at the end of TB
Cure	treatment and on at least one previous occasion
Treatment success	Composite of cured and treatment completed
Treatment failure	Sputum smear or culture is positive at month 5 or later during treatment
Death	TB patient who dies for any reason before starting or during the course of
	treatment
Loss to follow-up	TB patient who did not start treatment or whose treatment was interrupted
Loss to tonow-up	for 2 consecutive months or more
Not evaluated	TB patient for whom no treatment outcome was assigned, which includes
	cases who "transferred out" to another treatment unit as well as cases for
(transfer out)	whom the treatment outcome is unknown to the reporting unit

Figure 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow chart of inclusion process



tor ocer texicular

45 46 47

Table 2. Study characteristics

É						No with				Risk of
B 4 First author, year 5	Population	Study years	Study design	Location	Validation	outcome / sample size (%)	Predictors in final model	Performance measures	Model presentation	bias (population, predictor, outcome, analysis)
6 <u>Death</u>										
7 8 Abdelbary[9] / 9 2017	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 261/4216 (6%) Validation: 260/4215 (6%)	Age (<41, 41-65, ≥65), sex, MDR, HIV, malnutrition, alcoholism, diabetes, pulmonary TB	c-statistic = 0.70 Sensitivity = 60% Specificity = 71%	Risk score	Low, High, Low, High
10 Abdelbary[9] / 11 2017 (TB-DM) 12	TB-DM cases	2006 - 2013	Retrospective cohort	Mexico	None	88/2121 (4%)	Sex, malnutrition, BCG vaccinated, AFB smear (positive vs. negative)	c-statistic = 0.68	Risk score	Unclear, High, Low, High
13 _{Aljohaney[69]} / 14 2018	Hospitalized TB patients	Dec 2011 – Dec 2016	Retrospective cohort	Saudi Arabia	None	41/291 (14%)	<u>Clinical model:</u> Age, congestive heart failure <u>Clinical + lab model:*</u> Age > 65, congestive heart failure, bilateral disease on chest xray	<u>Clinical model:</u> Accuracy = 86% <u>Clinical & lab model:*</u> Accuracy = 90%	Odds ratios	Unclear, Unclear, Unclear, High
16 17 ^{Bastos[70] / 2016} 18	Inpatient and outpatient TB cases on DOT	2007 - 2013	Retrospective cohort	Portugal	External (setting)	Development: 121/681 (18%) Validation: 24/103 (23%)	Hypoxemic respiratory failure, age (≥50 vs. <50), bilateral involvement, comorbidities (at least one of HIV, diabetes, liver at least one of: HIV, diabetes, liver failure/cirrhosis, congestive heart failure, chronic respiratory disease), hemoglobin (<12 vs. ≥12)	AUROC = 0.84 (95% CI: 0.76-0.93) Sensitivity = 41.8% Specificity = 92.1%	Risk score	Low, Unclear, Low, High
19 Gupta-Wright[71] / 20 2019 21	Hospitalized TB- HIV patients	Oct 2015 – Sept 2017	Retrospective cohort	Malawi and South Africa	External (setting)	Development: 94/315 (30%) Validation: 147/644 (23%)	Sex, age 55+, currently taking ART, ability to walk unaided, severe anemia, positive TB-LAM	c-statistic = 0.68 (95% CI: 0.61-0.74) HL test: p=0.13 Calibration plot	Risk score	Low, Low, Low, High
22 Horita[72] / 2013 23	Hospitalized TB patients	Jan 2008 – Jul 2011	Retrospective cohort	Japan	External (setting)	Development: 36/179 (20%) Validation: 48/244 (20%)	Age, oxygen requirement, albumin, activities of daily living	AUROC = 0.893 Sensitivity = 0.92 Specificity = 0.73	Risk score	Low, Low, Low, High
24 25 _{koegelenberg[40]} / 26 ²⁰¹⁵ 27	Hospitalized TB patients	Jan 2012 – May 2013	Retrospective cohort	South Africa	None	38/83 (46%)	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	Mean score in survivors: 2.27 (SD=1.47) Mean score in non- survivors: 3.58 (SD=1.08)	Risk score	Low, Low, Low, High
28 Nguyen[53] 29 (general pop) / 2018	TB cases	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (split-sample)	Development: 253/3378 (7%) Validation: 270/3377 (8%)	Age group (15-44, 44-64, >64), US born, homeless, resident of long term care facility, chronic kidney failure, meningeal TB, miliary TB, HIV positive, HIV unknown	AUROC = 0.80 (95% CI: 0.77-0.82) HL test:X ² =6.3, p=0.613	Risk score	Low, Unclear, Unclear, High
BU B1Nguyen[37] (TB- DM) / 2019 B2	TB-DM patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	112/1227 (9%)	Age ≥65, US-born, homeless, IDU, chronic kidney failure, TB meningitis, Miliary TB, AFB positive smear, HIV positive	AUROC = 0.82 (95% CI: 0.78-0.87) HL test: X ² =4.54, p=0.81 Brier score=0.07	Risk score	Unclear, Unclear, Unclear, High
B3 B4 ^{Nguyen[52] (TB- HIV) / 2018 B5}	TB-HIV patients	Jan 2010 – Dec 2016	Retrospective cohort	Texas	Internal (bootstrap)	57/450 (13%)	Age ≥ 45, resident of LTCF, meningeal TB, abnormal CXR, diagnosis confirmed by positive culture of NAA, culture not converted or unknown	AUROC = 0.79 (95% CI 0.70-0.87) HL test: X ² =4.25, p=0.51 Brier score: 0.09	Risk score	Low, High, Unclear, High
36 37 ^{Pefura-Yone[54] /} 2017 38	TB patients	Jan 2012 – Dec 2013	Retrospective cohort	Cameroon	Internal (bootstrap)	213/2250 (9%)	Age, adjusted BMI, clinical form (PTB+, PTB-, EPTB), HIV	C-statistic: 0.808 HL test: X ² =6.44, p=0.60 Sensitivity = 80.7% Specificity = 68.2% Calibration plot	Model coefficients	Low, Low, Low, High
89 40 ^{Podlekareva[73] / 2013 41}	TB/HIV patients	Jan 2004 – Dec 2006	Retrospective cohort	52 cities in Europe and Argentina	None	995†	DST performed, treatment with RHZ, and cART at/near TB diagnosis	Crude RH = 0.62 (95% CI: 0.64-0.84)	Risk score	Low, Unclear, Low, High
42 43 44	Hospitalized TB patients	Mar 2000 – Jul 2009	Retrospective cohort	France	Internal (bootstrap)	20/53 (38%)	Miliary TB, catecholamine infusion, mechanical ventilation on admission	AUROC = 0.92 (95% CI: 0.85-0.98) Brier score = 0.13	Risk score	Unclear, Low,

BMJ Open

1 2								Optimism = 0.03 Accuracy = 85% Sensitivity - 75% Specificity = 91%		Low, High
3 4 Wang[74] / 2019 5	HIV-negative, culture-confirmed, pulmonary TB cases	Jan 2014 – Dec 2016	Prospective cohort	China	External (setting)	Development: 36/287 (13%) Validation: 15/104 (14%)	Age, cavitary lesion, pleural effusion, drug resistance, disseminated, albumin, c-reactive protein, white blood cell count, IL-6, MIF	AUROC = 0.85 ± 0.028	Odds ratios	Low, Low, Low, High
5 Wejse[75] / 2008	Pulmonary TB patients on DOT	1996 - 2001	Retrospective cohort	Guinea Bissau	None	100/698 (14%)	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	AUROC = 0.65 (95% CI: 0.6-0.7) Sensitivity = 0.45 Specificity = 0.75	Risk score	Low, High, Low, High
3 9 Zhang[45] / 2019 10	TB/HIV patients at end stage of AIDS	Aug 2009 – Jan 2018	Retrospective cohort	China	Internal (split-sample)	Development: 157/807 (19%) Validation: 40/200 (20%)	Anemia, TB meningitis, severe pneumonia, hypoalbuminemia, unexplained infection or space-occupying lesions, malignancy	AUROC = 0.867 (95% CI: 0.832-0.902) Sensitivity = 79.6% Specificity = 82.9%	Risk score	Low, Low, Low, High
11 _{Treatment failure}										
12 13 Abdelbary[9] / 14 ²⁰¹⁷	TB cases	2006 - 2013	Retrospective cohort	Mexico	Internal (split-sample)	Development: 2109† Validation: 6322†	Education (no or low vs. higher than primary school), MDR, AFB smear (>+2, +1, negative)	c-statistic = 0.65 Sensitivity = 52% Specificity = 66%	Risk score	Low, High, Low, High
15 Kalhori[49] 16 _(logistic) / 2010 17	TB cases at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 828/4836 (17%) Validation: 2418†	Gender, age, weight nationality, prison, case type	AUROC = 0.70 Accuracy = 81.64% HL test: X ² =11.935, df=8, p=0.154	Model coefficients	Unclear, Unclear, Unclear, High
18 19 _{Keane[30] / 1997} 20 21	Smear-positive TB patients on standard first-line regimen with DOT	1990 - 1995	Non-nested case control	Vietnam	None	130/803 (16%)	3 month model: Extensive lesions, mediastinal shift, average smear score 3rd month, weight, progressive x-ray, any previous treatment Baseline model: Mediastinal shift, average smear score, extensive lesions, any previous treatment, cavities, weight	$\frac{3 \text{ month:}}{\text{Sensitivity} = 80\%}$ $\text{Specificity} = 80\%$ $\frac{\text{Baseline:}}{\text{Sensitivity} = 70\%}$ $\text{Specificity} = 80\%$	Model coefficients	High, Unclear, Unclear, High
22 23 ^{Luies[33] / 2017}	Smear-positive pulmonary TB cases on DOT	May 1999 – Jul 2002	Nested case- control	South Africa	Internal (cross-validation)	10/31 (32%)	3,5,-Dihydroxybenzoic acid, (3-(4-Hydroxy-3-methoxyphenyl) propionic acid	AUROC = 0.89 (95% CI: 0.7-1.00)	Model coefficients	High, Unclear, Unclear, High
25 Mburu[76] / 2018 26	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	13/321 (4%)	HbA1c, regimen (retreatment), age, weight, random blood glucose, BMI, BUN, HIV positive result, ever smoker, creatinine	AUROC = 0.56 ± 0.07	Relative score	Low, Low, Low, High
27 Default						·				Ĭ
28 29 Thompson[77] / 30 ²⁰¹⁷	HIV uninfected adults with newly diagnosed pulmonary TB	Apr 2010 – Apr 2013	Retrospective cohort	South Africa	Internal (cross-validation) and external (setting)	6/99 (6%)	18 splice junctions and 13 genes	AUROC (internal) = 0.87 AUROC (external) = 0.63	Heatmap of differentially expressed genes	Low, Low, Low, High
31 32 Abdelbary[9] / 2017 (TB-DM) 33	TB cases	2006 - 2013	Retrospective cohort	Mexico	None	93/2121 (4%)	Age (<40 vs. ≥40), sex, HIV	c-statistic= 0.62	Risk score	Unclear, High, Unclear, High
34 35 _{Belilovsky[35]} / 36 2010 37	Hospitalized TB patients	1993 - 2002	Retrospective cohort	Russia	External (geographical)	Development: 1326/3904 (34%) Validation: 4662/12803 (36%)	Sex, unemployment, retreatment case, alcohol abuse (yes, no, no data), severe TB form, residence (urban vs. rural), age (25-50 vs. other), pulmonary TB (vs extrapulmonary), prison history	Belgrood: AUROC = 0.75 Orel: AUROC = 0.75 Pskov: AUROC = 0.78 Yaroslavi: AUROC = 0.75 Calibration table	Model coefficients	Unclear, High, High, High
38 39 40 _{Chang[31] / 2004} 41 42	All tuberculosis patients	Jan 1999 – Mar 1999	Nested case- control	China	None	102/408 (25%)	Baseline:* Ever smoker (current, former, never), retreatment (history of default, no history of default, not) Longitudinal: Smoking status (current, former, never), retreatment (with history of default, without history of default, never), unsatisfactory adherence in first two months (good, poor, fair, unknown), subsequent hospitalization, treatment side effects in last month of treatment	$\frac{\text{Baseline:}^{*}}{\text{AUROC} = 0.70 (95\% \text{ CI:} 0.63-0.76)}$ HL test: X ² = 1.448, df=5, p=0.919 <u>Longitudinal:</u> AUROC = 0.85 (95% CI: 0.80-0.90)	Odds ratios	High, High, Low, High
14										31

BMJ Open

1								HL test: $X^2 = 5.887$, df=6, p=0.436		
2 Chee[78] / 2000	TB cases	1996	Nested case- control	Singapore	None	38/71 (54%)	Chinese race, extent of family support, treatment duration	Accuracy = 74.6%	Model coefficients	High, Unclear, High, High
4 5 6 Cherkaoui[29] / 6 2014 7	TB patients with definite or probable pulmonary or extrapulmonary TB	Jun 2010 – Oct 2011	Non-nested case-control	Morocco	None	91/277 (33%)	Age <50, work interfering with ability to take TB treatment, retreatment regimen, daily DOT, moderate or severe side effects, told friends about TB, current smoker, never smoker, symptom resolution in <2 months, knowledge of TB treatment duration	AUROC = 0.85 (95% CI: 0.80-0.90) Sensitivity = 82.4% Specificity = 87.6% HL test: X ² =0.77, p- value=1.00	Survey tool	High, High, High, High
3 9 _{Rodrigo[79] / 2012 10}	New TB cases	Jan 2006 – Dec 2009	Prospective cohort	Spain	Internal (split-sample)	Development: 92/1490 (6%) Validation: 103/1589 (6%)	Immigrant, living alone, living in an institution, previous TB treatment, linguistic barriers (poor understanding), IV drug use, unknown IV drug use	AUROC = 0.67 (95% CI: 0.65-0.70) Sensitivity = 65.05% Specificity = 67.36%	Risk score	Low, Low, Low, High
11 <u>Unfavorable</u> outcome										
12 <u>Entremi</u> 13 Kalhori[50] 14 ^{predicting) / 2009[†]}	TB patients at DOT registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 6920† Validation: 2966†	Age, gender, nationality, prison, area, weight	Classification rate = 89.8% R2 = 0.45	Model coefficients	Unclear, Unclear, Unclear, High
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29Sauer[57] / 2018 [†] 30 31 32 33 34 35 36 37 38 39 40 41 42 43	TB cases	Data available through March 2018	Retrospective cohort	Azerbaijan, Belarus, Georgia, Moldova, Romania	Internal (split-sample)	Development: 103/411 (25%) Validation: 44/176 (25%)	<u>Forward selection (FS):*</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Backwards elimination (BE):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Stepwise selection (SS):</u> Drug sensitivity, employment status, smear microscopy, dissemination <u>Lasso:</u> Country, employment, extrapulmonary, cavity size, decrease in lung capacity, smear microscopy, drug sensitivity, chest imaging <u>Random forest (RF):</u> Top 5 by mean decrease accuracy: lung cavity size, type of resistance, employment status, country, total cavities Top 5 by mean decrease Gini index: Age of onset, drug regimen, lung cavity size, number of daily contacts, culture	$\frac{FS:*}{AUROC = 0.74}$ (95% CI: 0.66-0.82) Sensitivity = 0.36 Specificity = 0.89 Misclassification = 0.24 <u>BE:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.3 Specificity = 0.88 Misclassification = 0.27 <u>SS:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>Lasso:</u> AUROC = 0.72 (95% CI: 0.64-0.80) Sensitivity = 0.21 Specificity = 0.96 Misclassification = 0.23 <u>RF:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.96 Misclassification = 0.23 <u>RF:</u> AUROC = 0.73 (95% CI: 0.65-0.81) Sensitivity = 0.30 Specificity = 0.88 Misclassification = 0.27 <u>SVM linear:</u> AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0.21 Specificity = 0.94 Misclassification = 0.24 <u>SVM polynomial:</u> AUROC = 0.69 (95% CI: 0.60-0.77) Sensitivity = 0 Specificity = 1 Misclassification = 0.25	List	Unclear, Unclear, Unclear, High

BMJ Open

1 Baussano[47] / 2008 [§]	Pulmonary TB cases	2001 - 2005	Retrospective cohort	Italy	Internal (bootstrap)	576/1242 (46%)	Residency (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)	AUROC= 0.75 Calibration slope = 0.98 $R^2 = 0.24$	Nomogram	Low, Unclear, Low, High
B 4 Costa-Veiga[48] / 2017 [§]	Pulmonary TB cases	2000 - 2012	Retrospective cohort	Portugal	External (temporal)	Development: 1152/10766 (11%) Validation: 4714†	HIV, previous treatment, age class (25-44, 15-24, 45-64, >64), IV drug use, pathologies (other disease comorbidity)	AUROC= 75.9% (95% CI: 74.1-77.7) Sensitivity = 71% Specificity = 73%	Nomogram	Low, Low, Low, High
6 7 8 9 Killian[34] / 2019 ⁸ 10 11	TB patients (99DOTS program)	Feb 2017 – Sep 2018	Retrospective cohort	India	None	433/4167 (10%)	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 <u>w-misses</u> : missed doses in last week t-misses: total missed doses in 35 days units and a dense layer with 48 units for the dense layer and 4 units for the penultimate layer <u>Random forest</u> : 150 trees and no max depth based on DAT from first 35 day	LEAP* AUROC = 0.743 <u>lw-misses:</u> AUROC = 0.607 <u>t-misses:</u> AUROC = 0.630 <u>Random forest:</u> AUROC = 0.722	None	High, High, Unclear, High
12 13 14 ^{Madan[51] / 2018§} 15	TB-HIV patients on DOT with first-line TB treatment	2015	Retrospective cohort	India	None	78/448 (17%)	Sputum smear grade, previous TB,; disease classification, HIV status, ART status, CD4 cell count, sex and age group (with interaction terms between age group and sex; sputum smear status and type of TB; HIV status at TB diagnosis and CD4 cell category).	AUROC = 0.783 HL test p-value = 0.149	Model coefficients	Low, Low, Low, High
16 17 ^{Mburu[76] / 2018§}	Smear-positive TB patients	Feb 2014 – Aug 2015	Prospective cohort	Kenya	Internal (cross-validation)	32/340 (9%)	HbA1c, treatment regimen (retreatment), creatinine, BMI, BUN, weight, age, random blood glucose, HIV positive result, male gender	AUROC = 0.65 ± 0.06	Relative score	Low, Low, Low, High
10 Other outcome										
20 2	TB patients at DOTS registration	2005	Retrospective cohort	Iran	Internal (split-sample)	Development: 7254† Validation: 2418†	Case type, treatment category, risky sex, prison, sex, recent TB infection, diabetes, low body weight, TB type, length, previous imprisonment, age, area, HIV	Mean absolute percentage error = 1.24	Learned parameters	Unclear, Unclear, High, High
23 24. Hussain[56] / 2019 25 26	Pulmonary and extrapulmonary TB patients (TB Reach)	2011 - 2014	Retrospective cohort	Unknown	Internal (split-sample)	Development: 3371† Validation: 842†	Random forest*, artificial neural networks, and SVM	<u>Random forest:*</u> Accuracy = 76.32%	None	Unclear, Unclear, Unclear, High
 Abbreviations 27 *Indicates best 28 validation, initi 29 *Outcome is c 30 *Outcome is c *Outcome is a 31 *Outcome is tr 32 33 34 35 36 37 38 39 40 41 42 	: AUROC=Area under re I-performing/most relevan ernal validation, no valida nber unknown omposite of death and tre omposite of death, treatm value from 1 to 5 (1= pat eatment completion	ceiver operating nt model, which ation). If internal atment failure (le ent failure, loss t ient completed t	characteristic; c-s is included throug and external valid osses to follow-up to follow-up, and n he treatment cours	tatistic=concord hout the manusc lation were perf and not evaluat not evaluated ie in frame of D	ance statistic; DOTS= rript (see methods secti formed, both are reporte ed (unknown) outcome OTS, 2=cured, 3= quit	Directly Observed Th on for details). Perfo ed. es were excluded) treatment, 4=failed t	herapy, DM=Diabetes; HL=Hosmer-Lemeshow; TB=Tuberculosis; rmance measures are reported for highest level of validation performe reatment and 5=death)	d (ranked from strongest to wea	ıkest: external	

Table 3. Characteristics of patient populations in the 33 included studies with prediction models for tuberculosis treatment outcomes

Characteristic	Studies	Categories	N(%) or
	reporting	8	Median [IOR]
	characteristic.		
	n (% of total)		
Sample size	33 (11)	_	803 [291, 4167]
Study duration	32 (97)	-	4 [2 7]
years			. [-,,]
Study design	33 (100)	Prospective cohort	3 (9)
		Retrospective cohort	25 (76)
		Nested case-control	3 (9)
		Non-nested case-control	2 (6)
Data source	33 (100)	Medical record	6 (18)
		National registry or surveillance	13 (39)
		system	
		Local registry or surveillance	1 (3)
		system	
		Regional registry or surveillance	2 (6)
		system	
		Data collect form for study	11 (33)
		purposes	
Study region	32 (97)	Africa	8 (25)
		Asia	13 (41)
		Europe	6 (19)
		North America	4 (12)
		South America	0 (0)
		Global	1 (3)
High burden TB	31 (94)	All	143(42)
setting*			
		Some	1 (3)
		None	17 (55)
Missing data	18 (54)	Complete case-analysis	9 (50)
		Missing indicator method	4 (22)
		Heckman's method	1 (6)
		Simple imputation	2 (12)
		Sensitivity analysis with	1 (6)
		imputation	
		Other	1 (5)
Number of models developed	33 (100)	1	25 (76)
		2	4 (12)
		3	1 (3)

		4	2 (6)
		7	1 (3)
Reasons for multiple models developed	8 (24)	Different outcomes	1 (12)
		Different predictors considered	4 (50)
		Different methods	2 (25)
		Different outcomes	1 (12)
		Different populations and	1 (12)
		outcomes	

*Determined based on study location and WHO list of 30 high-burden TB countries in the 2019 Global Tuberculosis Report (1).

2	
3	
4	
5	
ر د	
6	
7	
8	
9	
10	
11	
10	
12	
13	
14	
15	
16	
17	
10	
10	
19	
20	
21	
22	
23	
24	
25	
25	
26	
27	
28	
29	
30	
31	
32	
22	
22	
34	
35	
36	
37	
38	
39	
40	
41	
40 40	
42	
43	
44	
45	
46	
47	
48	
<u>4</u> 9	
79 50	
50	
51	
52	
53	
54	

60

Included? Characteristic Yes No Unknown Median [IQR][‡], n 15 41 [37-49], n=18 Age* _ -18 HIV 7 8 23% [10-100], n=17 19 Diabetes 13 1 12% [5-21], n=11 MDR 8 7 18 1% [1-3], n=8 12 20 6% [4-12], n=10 Other drug 1 resistance Extrapulmonary 22 7 11% [4-17], n=16 4 TB† Previous TB 19% [9-30], n=17 20 1 12 DOT 14 0 19 100% [100-100], n=14 19 Hospitalized 13 1 100% [100-100], n=10 patients

Abbreviations: DOT=directly observed therapy; IQR=interquartile range; MDR=multi-drug resistance; TB=tuberculosis

*Based on the measure of central tendency reported in the study (mean: n=11; median: n=7) [†]Forms of extrapulmonary TB differ by study but included some of the following: Miliary, meningeal, pleural, peritoneal, disseminated, blood/bone, abdominal

[‡]Other than age (which is reported in years), this is the percentage of the population that has the characteristic among studies that include patients with the characteristic. For example, among the 18 studies that include persons with HIV, 17 report how many people had HIV and among those, the median percentage of the population with HIV is 23%.

Table 5. Methods reported for the 37 models of the 33 included studies with prediction models for tuberculosis treatment outcomes

Characteristic	Studies reporting characteristic, n (%)	Categories	N(%) or median [IQR]
Type of outcome	37 (100)	Single	29 (78)
		Composite	8 (22)
Outcome	37 (100)	Death	16 (43)
		Treatment failure	6 (16)
		Default, Loss to follow-up,	6 (16)
		or treatment interruption	
		Unfavorable outcome	6 (16)
		Treatment success	2 (6)
		Other [‡]	1 (3)
Number - prevalence of outcome*	32 (87)	-	94 [38-171] 15% [9-26]
Events per candidate variable [†]	30 (81)	-	6 [3-11]
Events per variable (in final model)	29 (78)	-	14 [9-26]
Predictor types		Clinical/epidemiologic	34 (92)
		Adherence	1 (3)
		Biomarker	2 (5)
Analysis	37 (100)	Logistic regression	29 (78)
		Survival analysis	3 (8)
		Machine learning	5 (14)
Method for considering predictors in multivariable models	36 (97)	All candidate predictors	12 (32)
		Based on unadjusted association with outcome	19 (51)
		Based on clinical relevance	1 (3)
		Other§	4 (14)
Selection of predictors during modeling	31 (84)	Full model approach	2 (6)
		Forward selection	7 (23)
		Backwards elimination	5 (16)
		Stepwise selection	8 (26)
		Random Forest	1 (3)
		Hosmer-Lemeshow model building criteria	4 (13)
		Bayesian model averaging	3 (10)
		Pairwise selection	1 (3)

BMJ Open

1
2
2
J ⊿
4
5
6
7
8
9
10
11
12
12
14
14
15
16
17
18
19
20
21
21
22
23
24
25
26
27
28
29
30
31
27
5Z
33
34
35
36
37
38
39
40
/1
40
42
43
44
45
46
47
48
49
50
51
51
52
53
54
55
56
57
58

59

60

P-value for consideration in model	17 (46)	0.01	2 (12)
		0.02	3 (18)
		0.11	1 (6)
		0.2	6 (35)
		0.25	5 (29)
P-value for retention in MV model	20 (54)	0.02	9 (45)
		0.1	9 (45)
		0.15	1 (5)
		0.2	1 (5)
Internal validation	19 (51)	Split-sample	10 (53)
		Bootstrap	5 (26)
		Cross-validation	4 (21)
External validation	6 (16)	Temporal	1 (17)
	6	Geographic	1 (4)
		Setting	4 (67)
Calibration	17 (46)	Calibration plot [¶]	2 (12)
		Calibration slope [¶]	1 (6)
		Hosmer-Lemeshow	13 (77)
		goodness of fit p-value [¶]	0.51 [0.20, 0.79]
		Calibration table [¶]	2 (12)
		Mean absolute error [¶]	1 (6)
Discrimination	30 (81)	C-statistic (AUROC) [¶]	30 (100)
			0.75 [0.68-0.84]
		Log rank test [¶]	2 (5)
Classification	18 (49)	Sensitivity	14 (78)
	- (-)		70 [54, 78]
		Specificity	13 (72)
			75 [71, 88]
		Accuracy	2 (11)
		Other**	2 (11)
Model presentation	34 (92)	Risk score	16 (43)
		Model coefficient	8 (22)
		Nomogram	2 (6)
		Odds ratios/relative scores	4 (12)
		Survey tool	1 (2)

Abbreviations: AUROC=area under receiver operating characteristic; c-statistic=concordance statistic

*Prevalence of outcome in the population used to develop the prediction model (i.e.

derivation/development subset if split-sample technique was used or full sample if the model was not validated or if bootstrap/cross-validation was used)

[†]Only 5 studies report the exact number of predictors considered. Otherwise, the number of candidate predictors was estimated from the provided tables or lists of candidate predictors in the source paper.

[‡]Outcome is a value from 1 to 5 (1= patient completed the treatment course in frame of DOTS, 2=cured, 3= quit treatment, 4=failed treatment and 5=death)

[§]Other methods of determining which variables to consider for prediction model include: principal components analysis (n=1), screening for multi-collinearity via correlation coefficient (n=1), one study used a combination of a priori and selection via univariable association, and the other used machine learning pre-processing (n=1)

[¶]Sums to more than 100%, because some studies report multiple measures of calibration or discrimination

Based on the following cut-off methods: Youden (n=4) concordance probability (n=1), estimated at nearest 0,1 for studies that present a range of sensitivity and specificity in a table or figure (n=4), or unknown (n=5)

**Other includes one study that reports false positive rate and one study that includes a graph of sensitivity vs. specificity.

Figure 2. Most common predictors considered and included

[See Figure 2]

Figure 2 legend:

Considered: the predictor as evaluated as a candidate predictor prior to multivariable modeling Included: the predictor was considered and subsequently included in the final multivariable model

Figure 3. Heatmap of signaling questions from risk of bias assessment with PROBAST

[See Figure 3]

Figure 3 legend:

PROBAST questions (additional details in Supplemental File 5)

- Participants 1: What study design was used and was it appropriate?
- Participants 2: Were all inclusion and exclusion criteria appropriate?
- Predictors 1: Were predictors defined as assessed the same way for all participants?
 - Predictors 2: Were predictor assessments made without knowledge of data outcome?
 - Predictors 3: Are all predictors available at the time the model was intended to be used?
- Outcome 1: Was the outcome determined appropriately?
- Outcome 2: Was the outcome pre-specified or standard?
- Outcome 3: Were predictors excluded from outcome definition?
- Outcome 4: Was the outcome defined and determined in a similar way for all participants?
- Outcome 5: Was the outcome determined without predictor information?
- Outcome 6: Was the time interval between predictor assessment and outcome determination appropriate?
- Analysis 1: Were there a reasonable number of participants with the outcome?
- Analysis 2: Were continuous and categorical variables handled appropriately?
- Analysis 3: Were all enrolled participants included in the analysis?
- Analysis 4: Were participants with missing data handled appropriately?
 - Analysis 5: Was selection of predictors based on univariable analysis avoided?
 - Analysis 6: Were complexities in data (censoring, competing risks, sampling of control participants) accounted for appropriately?
 - Analysis 7: Were relevant model performance measures evaluated appropriately?

Analysis 8: Were model overfitting, underfitting, and optimism in the model performance accounted for?

Analysis 9: Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?



[See Figure 4]

to beet terien only





203x203mm (300 x 300 DPI)

7	
8	



203x203mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml





203x203mm (300 x 300 DPI)
BMJ Open

Supplemental File 1. PRISMA Checklist

Section/topic #		Checklist item	Reported o #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT	<u> </u>		<u></u>
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			1
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).	Supplement 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
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.	Supplement 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; Supplem Files 4 and
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; Supplem 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., l 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			<u>+</u>
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; Tabl 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; Figu and 4

		-	
Results of individual studies	sults of ividual studies20For 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	thesis 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 studie			
	external validation		
Comparator	Models will be compared to each other, as there is no other relevant comparator for this systematic review		
	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		
Outcome	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		
Timing	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		
Setting	support for individuals at greatest risk for unsuccessful TB treatment outcomes.		

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Supplemental File 3. Search Strategy

PubMed	 ((validat*[tiab] OR predict*[tii] 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 finding*[tiab] OR factor*[tiab] OR model*[tiab])) (stratification[tiab] OR "ROC Curve"[Mesh] OR discrimination[tiab] OR discriminate[tiab] OR "c-statistic"[tiab] OR "c statistic"[tiab]) (stratification[tiab] OR TROC Curve"[Mesh] OR discrimination[tiab] OR indices[tiab] OR "c-statistic"[tiab] OR "c statistic"[tiab]) (stratification[tiab] OR nodel*[tiab]) (utberculosis[Mesh] OR tuberculosis[tiab]) (tuberculosis[Mesh] OR tuberculosis[tiab]) (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 dies[tiab])) 1 OR 2 3 AND 4 5 AND 6 AND (humans[Filter]) AND ("1995"[Date - Publication] : "3000"[Date - Publication]))
Embase	 (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. (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"/ tuberculosis/ or tuberculosis.ti,ab (outcome\$ or mortality\$ or death\$ or fail\$ or recur\$ or relapse\$ or default\$ or abandon\$ or loss\$ or cure\$ or success\$ or unsuccess\$ or di or dies).ti,ab. 1 or 2 3 and 4 5 and 6 limit 7 to (human and yr="1995 -Current")
Web of Science	 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 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*)) TS=(stratification or discrimination or discriminate or c-statistic or "c statistic" or "area under the curve" or AUC or calibration or indices algorithm or multivariable or "receiver operating characteristic") TS=(tuberculosis) TS=(outcome* or mortality* or death* or fail* or recur* or relapse* or default* or abandon* or loss* or cure* or success* or unsuccess* die or died or dies) 1 or 2 3 and 4 5 and 6; IC Timespan=1995-2019
Google	tuberculosis treatment outcome prediction prognostic model development validation

Supplemental File 4. CHARMS Checklist

Domain	Key items	on pag
SOURCE OF DATA	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	
	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting,	
	inclusion and exclusion criteria)	
PARTICIPANTS	Participant description	
	Details of treatments received, if relevant	
	Study dates	
	Definition and method for measurement of outcome	
	Was the same outcome definition (and method for measurement) used in all patients?	
OUTCOME(S) TO BE	Type of outcome (e.g., single or combined endpoints)	
PREDICTED	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	
	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease	
	characteristics)	
C A NDID A TE	Definition and method for measurement of candidate predictors	
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)?	
OR INDEX TESTS)	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)	
	Number of participants with any missing value (include predictors and outcomes)	
MISSING DATA	Number of participants with missing data for each predictor	
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	
	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)	
MODEL DEVELOPMENT	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)	
	Calibration (adjustion plat adjustion glong Hagman Lamaghaw test) and Disarimination (C statistic D	
MODEL	statistic, log-rank) measures with confidence intervals	
PERFORMANCE	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a- priori cut points were used	
	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)	
MODEL EVALUATION	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	
	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)	
RESULTS	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
NTERPRETATION 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.
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

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/probabl	<u>y yes</u>	No/probably no	No information
1	What study design was used and was it appropriate?	Yes: If a cohort design (or proper registry data) you have confidence in participant enrollment is described Probably yes: a nested of case-cohort design (wit adjustment of the baseli the analysis) has been u design was used but par	(including RCT was used and data quality and s clearly case–control or h proper ne risk/hazard in sed or a cohort ticipant	No: If a non-nested case–control de has been used Probably no: a nested case-control s was used without proper adjustment baseline risk/hazard	sign If the method of participant sampling is unclear. study t of
2	Were all inclusion and exclusion criteria appropriate?	design was used but participant enrollment was data quality is unclear 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 having the outcome and so are no lo at risk of developing outcome, or if specific subgroups are excluded that may have altered the performance o prediction model for the intended ta population. Probably no: inclusion and exclusio criteria are unclear and it seems pos that there was bias in selection of participants that could lead to the m 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. t f the rrget sible odel
				· L .	Γ
	Low risk of t	pias		High risk of bias	Unclear risk of bias
If the "Prob ≥ 1 of	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		If the answer to a "No" or "Probab except if defined	any of the signaling questions is oly no," there is a potential for bias, I 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 2: Predictors

could still be "Low risk of bias" but specific reasons should

be provided why the risk of bias can be considered low.

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.

domain at high risk of bias.

		Artobably yes: If it is l outcome information during predictor asses entirely clear (retrosp collection/surveillance	was not used sment, but not ective data e data)	Predictors would not be available	able at	No information on whether
3	Are all predictors available a the time the model was inten to be used?	t ded intended to be used for	he model is or prediction	the time the model is intended used for prediction.	to be	redictors would be available a time the model is intended to b used for prediction.
	Low risk of	bias	Н	ligh risk of bias		Unclear risk of bias
If th "Pr ≥1 c cou be p	ne answer to all signaling ques obably yes," then risk of bias c of the answers is "No" or "Pro Id still be "Low risk of bias" b provided why the risk of bias c	tions is "Yes" or an be considered low. If bably no," the judgment ut specific reasons should an be considered low.	If the answer to a "No" or "Probabl bias, except if def	ny of the signaling questions is ly no," there is a potential for fined at low risk of bias above.	If relevan the signal signaling at high ris	It information is missing for some ling questions and none of the questions is judged to put this do sk of bias.
Don Bia outo met dete	main 3: Outcome s in model performance can oc come determination can result thods are inconsistently applied ermination can also result in bi	cur when methods used to de from use of suboptimal meth l across participants, or when as.	etermine outcomes ods, tests, or criter l knowledge of pre	incorrectly classify participants w ia that lead to unacceptably high l dictors influence outcome determ	vith or with evels of err ination. Inc	out the outcome. Bias in methods ors in outcome determination, wl orrect timing of outcome
	Signaling question	Yes/probably	yes	No/probably no		No information
1	Was the outcome determined appropriately?	If a method of outcome det been used which is conside acceptable by guidelines or publications on the topic Note: This is about level of error within the method of the outcome (see concerns applicability about whether of the outcome method is a	determination has sidered optimal or s or previous cIf a clearly suboptimal meth used that causes unacceptal determining outcome status participantsel of measurement of determining rms for ther the definitionof determining rms for ther the definition		is been or in	No information on how outcom determined
2	Was the outcome pre- specified or standard?	Yes: If the method of outcodetermination is objective, outcome definition is used, prespecified categories are outcomes. (i.e. outcome as: based on previously publis published study protocol, of guidelines) Probably yes: The outcome is not clearly based on guide previous research, but outca assessment is objective and inadvertently alter study re	ome or if a standard or if used to group sessment is hed studies, or clinical e determination lelines or ome t would not sults	No: If the outcome definition was standard and not prespecified Probably no: a non-standard or no prespecified outcome was used, a unclear whether the outcome defi- could introduce bias. *Caution with composite outcom favor a better model by excluding outcome components or including atypical events	s not on- ind it is inition es that g typical g	No information on whether the outcome definition was prespec 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		is is e included arly stated) If ≥1 of the predictors forms part of the outcome definition dictors are ition If		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 (asymptot)		If outcomes were clearly defined determined in a different way for participants	and some	No information on whether out were defined or determined in a similar way for all participants
5	Was the outcome determined without predictor information	Yes: If predictor informatic known when determining t status, or outcome status de clearly reported as determin knowledge of predictor infor have been available at time assessment, but outcome d objective and knowing infor predictors would not influe	on was not he outcome etermination is ned without ormation. ormation might of outcome efinition is ormation about nce outcome	No: If it is clear that predictor inf was used when determining the o status Probably no: it is likely predictor information was available at the t outcome assessment, and outcom definition is subjective and know predictors could influence outcom determination.	ormation utcome ime of e ledge of ne	No information on whether out was determined without knowle of predictor information

6	Was the time interval between predictor determination appropriate	sased on culture results, assed on culture results, f the time interval betwee issessment and outcome was appropriate to enable and representative numb poutcomes to be recorded information on the time is equired to allow a repre- of the relevant outcome of predictor assessment and letermination were from aken within an approprise	etc) een predictor determination e the correct type er of relevant , or if no interval is sentative number occur or if loutcome information ate time interval.	If the tin assessme too short type and relevant	ne interval between predi ent and outcome determin t or too long to enable the representative number o outcomes to be recorded	ctor nation is correct f	If no information was provide the time interval between pre- assessment and outcome determination.
	Low risk of bia	15	Н	ligh risk	of bias		Unclear risk of bias
If th "Pr ≥1 cou sho low	he answer to all signaling question obably yes," then risk of bias can of the answers is "No" or "Probal and still be "Low risk of bias" but build be provided why the risk of b 7.	ns is "Yes" or be considered low. If oly no," the judgment specific reasons ias can be considered	If the answer to an "No" or "Probably bias, except if def	ny of the s y no," the ined at lo	signaling questions is re is a potential for w risk of bias above.	If relevan the signal signaling at high ris	t information is missing for son ing questions and none of the questions is judged to put this sk of bias.
Do Sta bia see	main 4: Analysis tistical analysis is a critical part o s in reported model performance k statistical advice when complet	f prediction model devel measures. Model develo	lopment and validati pment studies includ	ion. The u de many s	use of inappropriate statis tteps where flawed metho	tical analys ods can disto	is methods increases the potent ort results. We recommend revi
	Signaling question	For model develor	probably yes	number	No/probably	no t studies if	No information
1	Were there a reasonable numbe of participants with the outcome	of participants with the number of cana is ≥20 (EPV ≥20). r ? For model validati participants with th	the outcome relative didate predictor para * on studies, if the nur ne outcome is ≥100.	we to imeters mber of	the number of participa outcome relative to the candidate predictor part <10 (EPV <10).* For model validation st number of participants outcome is <100.	nts with the number of ameters is udies, if the with the	 no information on the nu candidate predictor parar or number of participants the outcome, such that th cannot be calculated. For model validation studinformation on the numb participants with the outcome
		* For EPVs betwee frequency, overall 145 to 147.	en 10 and 20, the iter model performance,	m should , and distr	be rated as either probab ibution of the predictors	ly yes or pr in the mode	obably no, depending on the ou el. For more guidance, see refer
2	Were continuous and categorica predictors handled appropriately	Yes: If continuous continuous or if co examined as linear restricted cubic spl polynomials. Probably yes: If cc converted into >2 of the model (i.e., dic using a prespecifie avoids sparse data/ improve statistical For model validati predictors are included definitions or trans variables are categ points, ascompared study.	predictors are kept a ntinuous predictors or non-linear using lines or fractional ontinuous predictors categories when incl hotomized or catego d method or in a way 'would not intentional significance. on studies, if continu- ided using the same formations, and cate orized using the same d with the development	as are luded in prized) yy that ally uous egorical ne cut ent	No: For model develop if continuous predictors converted into 2 catego included in the model. Probably no: If categor group definitions do no prespecified method or variables were split into but the decision of how variables is unclear. For model validation st continuous predictors a using different definitio transformations, or cate variables are categorize different cut points, as a with the development s	ment studie s are ries when ical predict t use a continuous o >2 groups t o split udies, if re included ons or egorical ad using compared tudy.	 No information on wheth continuous predictors are examined for nonlinearity no information on how categorical predictor grou defined. For model validation stud information on whether t same definitions or transformations and the s cut points are used, as co with the development stud
3	Were all enrolled participants included in the analysis?	If all participants e included in the dat	nrolled in the study a analysis.	are	If some or a subgroup of are inappropriately exc the analysis (because the missing data, unknown outliers)	of participar luded from ney were outcome,	ats No information on wheth enrolled participants are included in the analysis.
		Yes: If there are no predictors or outco	o missing values of omes and the study	not	No: If participants with are omitted from the an the method of handling	missing da alysis, or if missing da	ttaIf there is insufficientinformation to determinettamethod of handling missi

Ap		1	1				
	plicability						
risk "Pr spe	The answer to all signaling questions is to f bias can be considered low. If ≥ 1 obably no," the judgment could still li- cific reasons should be provided why sidered low.	s ies or Probably yes," then of the answers is "No" or be "Low risk of bias" but y the risk of bias can be	a potential for risk of bias ab	No" or "Probably no," there is bias, except if defined at low bove.	of the s this dor	ant information is missing for f the signaling questions and n ignaling questions is judged to nain at high risk of bias.	
If +1	Low risk of	bias	If the answer	High risk of bias	If rolar	Unclear risk of bias	
				. /			
9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	coefficients reported sis.	If the predictors and regression coefficients in the final model correspond to reported results multivariable analysis. (i.e. ro of model coefficients to create "risk score" are inappropriate determined).	n do not from unding e a ly	If it is unclear whether the regression coefficients in the final model correspond to reported results from multivariable analysis.		
8	were model overfitting, underfitting, and optimism in model performance accounted for?	Probably yes: If internal validati and optimism was estimated as y then optimism-corrected perform measures were not appropriately (accounting for all model develop procedures)	ion was used very low, and nance v calculated opment	Probably no: Internal validation bootstrapping or cross-validation was conducted but did not ince all model development proceet including any variable selection were not used to correct mode performance measures.	on with ion lude lures on or el		
		Yes: If internal validation techni (bootstrapping and cross-validat all model development procedur to account for any optimism in r and subsequent adjustment of th	iques ion) including res, were used nodel fitting, e model	No: If no internal validation h been performed, or if internal validation consists only of a s random split-sample of partic data,	as ingle ipant	No information: No informa is provided on whether inter validation techniques, inclue all model development procedures, have been appli	
7	Were relevant model performance measures evaluated appropriately? Probably yes: if authors present predicted probabilities with con intervals and corresponding out frequencies across subgroups		a table of fidence come	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.		references to relevant litera or specific mention of meth such as using Kaplan–Meie estimates), or no informatic thresholds for estimating classification measures is g	
		Yes: If both calibration (via cali and discrimination (c-index) are appropriately (including relevan tailored for models predicting su outcomes).	bration plot) evaluated t measures urvival	If both calibration and discrimination are not evaluat if only goodness-of-fit tests (Hosmer-Lemeshow test), are to evaluate calibration or if fo	ed, or used r	Either calibration or discrimination are not report or no information is provide to whether appropriate performance measures for	
6	Were complexities in the data (censoring, competing risks, sampling of control participants) accounted for appropriately?	for appropriately, or if it is clear potential data complexities have identified appropriately as unim	that any been portant.	affect model performance are ignore. For example, case-cor studies that do not estimate ba risk or studies with censoring competing risks that do not us survival analysis or other appropriate methods.	ntrol useline or se	whether complexities in the are present or accounted for appropriately if present.	
5	on univariable analysis avoided?	modeling. If any complexities in the data a	re accounted	multivariable modeling.	t could	selection is avoided.	
	Was selection of predictors based	imputation methods are convinc is low	on the basis	comparison of included vs. ex participants or if sensitivity ar with imputation methods are reported, but the results are no convincing to rule out bias fro excluding missing data If the predictors are selected of horis of mutariable analysis of	cluded nalysis ot om on the	If there is no information to	
		with missing data were excluded provide comparison of included participants or if sensitivity anal	d and authors vs. excluded vsis with	of methods to handle missing Probably no: If authors provid	data. le		

Page 57 of 64

BMJ Open

<u>Participants</u> : 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.
<u>Predictors</u> : 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 /	2	
2019	Death	The outcome was mortality risk at 2 months after admission. '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	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) transferpatient 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	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 /	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
<u>ci</u> (2004	Default, Abandon, or	Default was defined as failure to collect drugs for 2 months or more after registration

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	(interruption >?	
	(interruption > 2 months)	
	Default.	
	Abandon, or	
	LTF	Defaulter or cases were defined as patients on anti-tuberculosis treatment at the TBCU who failed to turn up for the
	(interruption >2	scheduled appointments despite usual attempts to recall them by phone or mail, as described below, and from whon
Chee / 2000	months)	least one home visit during the study was recorded
	Default,	
	Abandon, or	
	LTF	
	(interruption >2	
Cherkaoui / 2014	months)	Treatment default was defined as an interruption in TB treatment for $>=2$ consecutive months.
	Default,	
	Abandon, or	
	LTF	
	(interruption >2	Interruption of treatment for any reason for more than 2 months, non-completion of treatment within 9 months whe
Rodrigo / 2012	months)	patient is placed on a 6 month regimen. or drug intake of <80% the prescribed dose.
Kalhori	Treatment	
(predicting) /	success (cure +	
2009	completion)	For each patient dependent variable was recorded whether or not the patient finished the treatment course and get c
	Unfavorable	
G / 2019	outcome (death	
Sauer / 2018	+ failure)	i ne primary outcome was treatment failure, which we defined as failure of therapy or death.
	Untavorable	
	outcome (death,	
D (2000	Tailure, LTF,	I reatment interruption or default, treatment failure, transferred out cases and those lost to follow-up were grouped
Baussano / 2008	NE)	unsuccessful outcomes
		In line with WHO criteria, SVIG-1B categorized a six possible and mutually exclusive categories for treatment
		outcomes, grouped in this study into a binary outcome: (1) Successful outcome-if PTB patients were treated before
	Unfovorshis	in each of not providing control negative sinear microscopy at the end of treatment at least one previous follow-up to
	outcome (death	in case of not providing sputin samples, cure is declared in treatment completed and absent of disease clinical evid
Costa Veica /	failure I TE	(valcgories 1 and 2). (ii) Unsuccessful outcome-if treatment of P I B patients resulted in failure (i.e. remaining smea
2017	NF)	months or more after registration cat 4) death (cat 5) or were transferred out (cat 6)
2011	Unfavorable	months or more after registration, cat. +), death (cat. 5) or were transferred-out (cat. 0)
	outcome (death	
	failure LTF	We label 'Cured' and 'Treatment Complete' to be favorable outcomes and 'Died' 'Treatment failed' and 'Lost to foll
Killian / 2019	NE)	up' to be unfavorable outcomes
	Unfavorable	
	outcome (death	
	failure, LTF.	Favourable treatment outcomes included cure and treatment completed. Unfavourable treatment outcomes included
Madan / 2018	NE)	death, loss to follow-up, treatment failure, transfer out, or a switch to MDR TB treatment.
	Unfavorable	
	outcome (death,	
Mburu / 2018 -	failure, LTF,	
Unfavorable	NE)	The primary analyses compared favorable versus unfavorable outcomes at end of treatment
	Other	The values of outcomes might be any values from 1 to 5 which means different outcomes. Value 1 means patient
Kalhori (fuzzy) /	composite	completed the treatment course in frame of DOTS, 2 means the patient has been cured, 3 means patients has quitted
2009	outcome	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

Abdelbary / 2017 - Death Abdelbary / 2017 - Failure	
Death Abdelbary / 2017 - Failure	2 + 2*(Age 41-65) + 5*(Age >= 65) + 2*(Male gender) + 4*(MDR TB) + 3*(HIV) + 3*(Malnutrition) + 2*(Alcoholism) + 2*(Alcohol
Abdelbary / 2017 - Failure	2*(Male*diabetes) + 3*(HIV*pulmonary TB) - 1*(diabetes) - 1*(pulmonary TB)
Failure	$9*(N_1, \dots, 1, \dots, 1, \dots, 1, 1, 0*(MDD) + 10*(MDD) + 10*(MDD, \dots, 1, 2) + 15*(MDD, \dots, 1, 2))$
Abdelbary / 2017 -	$8^{(100 \text{ of 10W education}) + 40^{(MDR) + 10^{(AFB sinear + 2) + 13^{(AFB sinear + 5)}}$
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)
	Don't report final model, but show the beta coefficients. The coefficients are written as predictor (beta-coefficient): age ³ 65 (2.497),
Aljohaney / 2018	congestive heart failure (1.231), bilateral disease on chest x-ray (1.192)
Bastos / 2016	$3^{(Hypoxemic respiratory failure) + 2^{(Age>=50) + 1^{(Bilateral involvement) + 1^{(At least one of: HIV, diabetes, liver failure/cirrnosis, concestive heart failure, chronic respiratory disease) + 1*(Hemoglobin<12)$
Bastos / 2010	Nomogram with: residency status (residential vs. homeless) sex. geographic origin (non-EU vs. EU) case definition (other than definite
Baussano / 2008	vs. definite), treatment setting (inpatient and unknown vs. outpatient), age (continuous)
	-3.2 + 0.8*(male gender) + 0.7*(unemployment) + 0.4*(retreatment case) + 1.1*(alcohol abuse) + 0.6*(no data about alcohol) +
Belilovsky / 2010	0.8*(severe TB form) - $0.3*$ (urban residence) + $0.4*$ (age 25-50) + $0.8*$ (pulmonary TB) + $0.5*$ (prison history)
C1 (2001	Dont report final model. Just show odds ratios of predictors but don't report intercept term, which are written as predictor (OR) as follows:
Chang / 2004	Current smokers (3.44) , ex-smokers (2.48) , history of default (10.74) , no history of default (0.80) ,
	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 finands (0.08). Treatment duration (1.85). Treatment duration is actegorized as 6 menths, 0 menths, and >0 menths, but only one OP is
Chee / 2000	presented
0100/2000	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
C1 1 1 1 0 0 1 1	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
Cherkaoui / 2014	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: ves/no)
Custa- v ciga / 201 /	9*(Male sex) + 7*(nation aged 55+) + 6*(currently taking ART) + 7*(unable to walk unaided) + 7*(hemoglobin < 80, severe anomia) +
Gupta-Wright / 2019	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
	Learned parameters by training set for each predictor written as predictor (learned parameter): Case type (0.467), treatment category (-
Kalhori (fuzzy) /	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
2009 Kalhari (lacistic) /	(0.950), length (-0.255), previous imprisonment (2.398), age (0.257), area (0.8895), HIV (0.731)
2010	exp(-0.93 - 0.71*(gender) + 0.02*(age) - 0.02*(weight) + 0.5*(nationality) + 0.99*(nrison) + 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))
	Unclear. No constant term provided. Here are the predictor (OR): Mediastinal shift (2.1), average smear score (1.5), extensive lesions
Keane / 1997	(3.6), any previous treatment (2.3), cavities (1.7), weight (0.98)
	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
Killian / 2010	dense layer and 4 units for the penultimate layer
Kiiliali / 2019	One point for each parameter: sentic shock HIV with CD4 < 200 creatinine > 140 (male) or >120 (female). D: FO2 ratio < 200 chect
Koegelenberg / 2015	radiograph showing miliary pattern/parenchymal infiltrates, absence of TB treatment at admission
L (2017	
Lutes / 2017	Written as predictor (OR): 3,5,-Dihydroxybenzoic acid (25.6), 3-(4-Hydroxy-3-methoxyphenyl) propionic acid (1.3)
	within as predictor (UK): New 1B with 1+ smear grade (5.78), New 1B with 2+ smear grade (2.69), New 1B with 5+ smear grade (1.69), New TB without smear (1.67). New TB with smear positive unknown grade (1.00). Previously treated smear positive TP (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 1B with CD4>300 (1.05), HIV+ diagnosed after 1B with CD4 350-500 (2.49), HIV+ diagnosed after TB
	with CD4 200 350 (8 88) HIV+ diagnosed after TP with CD4 50 200 (6 70) HIV+ diagnosed after TP with CD4 <50 (12 00) E1-25
Madan / 2018	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 ≥ 45 (3.96).
Madan / 2018 Mburu / 2018 -	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) 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
Madan / 2018 Mburu / 2018 - Failure	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) 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
Madan / 2018 Mburu / 2018 - Failure	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) 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 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,
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 -	 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) 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 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
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable	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) 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 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
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable Nguyen (general pop) (2018	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) 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 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 $6^{*}[Age 45-64] + 12^{*}[Age>65] + 2^{*}[US born] + 2^{*}[Homeless] + 4^{*}[Resident of LTCF] + 8^{*}[Chronic kidney failure] + 10^{*}[Meningeal TB]$
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable Nguyen (general pop) / 2018	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) 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 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 $6^{*}(Age 45-64) + 12^{*}(Age>65) + 2^{*}(US born] + 2^{*}(Homeless) + 4^{*}(Resident of LTCF] + 8^{*}(Chronic kidney failure] + 10^{*}(Meningeal TB) + 4^{*}(Miliary TB) + 6^{*}(HIV positive] + 6^{*}(HIV unknown]$
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable Nguyen (general pop) / 2018 Nguyen (TB-DM) / 2019	 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) 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 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 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] 16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliary TB] + 6*[AFB positive smear] + 24*[Positive HIV]
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable Nguyen (general pop) / 2018 Nguyen (TB-DM) / 2019 Nguyen (TB-HIV) /	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) 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 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 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] 16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliary TB] + 6*[AFB positive smear] + 24*[Positive HIV] Prognostic score: 5*[Age >= 65] + 12*[Resident of LTCF] + 9*[Meningeal TB] + 6*[AFB positive smear] + 24*[Resident of LTCF] + 9*[Meningeal TB] + 6*[AFB positive scores] + 12*[Resident of LTCF] + 9*[Meningeal TB] + 6*[AFB positive smear] + 24*[Rositive HIV]
Madan / 2018 Mburu / 2018 - Failure Mburu / 2018 - Unfavorable Nguyen (general pop) / 2018 Nguyen (TB-DM) / 2019 Nguyen (TB-HIV) / 2018	 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) 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 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 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] 16*[Age >= 65] + 5*[US-born] + 11*[Homeless] + 20*[Chronic kidney failure] + 20*[TB meningitis] + 13*[Miliary TB] + 6*[AFB positive smear] + 24*[Positive HIV] 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]

Page 61 of 64

BMJ Open

	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]
Sauer / 2018	Negatively correlated: drug sensitivity (sensitive), employment status (employed), microscopy: 1 to 99 acid-resistant bacteria in 100 field of view when stained by Ziehl-Nielsen, dissemination (diffuse pulmonary nodules detected)
Гhompson / 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
Weise / 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<200 MUAC<200
Zhang / 2019	2*[Anemia (HGB < 90g/L)] + 2*[Tuberculous meningitis] + 5*[Severe pneumonia] + 2*[Hypoalbuminemia] + 7* [Unexplained infection or space-occurving lesions] + 5* [Malignancies]
shang / 2019	of space occupying testonol + 5 [[nunghaneles]

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	Excluded	Unknown
	(N=11)	(N=7)	(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)	4		
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),	Nationality (3), Age (2),	Age (12), previous TB (9),
	extrapulmonary TB (4),	HIV (2), living situation	BMI (8), extrapulmonary
	HIV (4), other	(2), previous TB (2), sex	TB (6), sex (6)
	comorbidities (4), smear	(2), treatment regimen (2)	
	result (4)		

) 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)

³ [^]Witten as predictor (number of models included in). Top 5 unless there was a tie, in which case more predictors were listed.

<u>Summary:</u> 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	Excluded	Unknown
	(N=10)	(N=11)	(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),	Age (7), sex (5),	Age (7), nationality (5),
	extrapulmonary TB (4),	extrapulmonary TB (4),	previous TB (5), BMI (4),
	previous TB (4)	hemoglobin (3), HIV (3),	sex (4), treatment regimen
		MDR (3), other lab values	(4), x-ray findings (4)
		(3), x-ray findings (3)	

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.</p>

2	
3	
4	
5	
5	
6	
7	
8	
٥ ٥	
20	
10	
11	
12	
12	
1.4	
14	
15	
16	
17	
10	
18	
19	
20	
21	
22	
22	
23	
24	
25	
25	
26	
27	
28	
20	
29	
30	
31	
32	
22	
24	
34	
35	
36	
37	
20	
38	
39	
40	
41	
12	
42	
43	
44	
45	
10	
46	
47	
48	
49	
50	
51	
52	
53	
55	
54	
55	
56	
57	
50	
20	
59	
60	

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

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
12	
17	
14	
15	
10	
1/	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
30	
38	
30	
10	
40 1	
41	
42	
43 44	
44 4	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	

 Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). 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. Present results of each meta-analysis done, including confidence intervals and measures of consistency. Present results of any assessment of risk of bias across studies (see Item 15). Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Summarize the main findings including the strength of evidence for each main outcome; 	13-14; Figures 3 and 4 11-14; Table 2 N/A N/A N/A
 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. Present results of each meta-analysis done, including confidence intervals and measures of consistency. Present results of any assessment of risk of bias across studies (see Item 15). Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Summarize the main findings including the strength of evidence for each main outcome; 	11-14; Table 2 N/A N/A N/A
Image: Present results of each meta-analysis done, including confidence intervals and measures of consistency. Image: Present results of any assessment of risk of bias across studies (see Item 15). Image: Present results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Image: Present results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Image: Present results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Image: Present results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Image: Present results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A N/A N/A
 Present results of any assessment of risk of bias across studies (see Item 15). Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Summarize the main findings including the strength of evidence for each main outcome; 	N/A N/A
 Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). Summarize the main findings including the strength of evidence for each main outcome; 	N/A
Summarize the main findings including the strength of evidence for each main outcome;	
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
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
6 Provide a general interpretation of the results in the context of other evidence, and implications for future research.	19
7 Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	20
	 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). Provide a general interpretation of the results in the context of other evidence, and implications for future research. Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.