A clinical prediction model for unsuccessful pulmonary tuberculosis treatment outcomes

Lauren S. Peetluk[1], Peter F. Rebeiro[1,2], Felipe M. Ridolfi[3], Bruno B. Andrade[4,5,6,7,8,9], Marcelo Cordeiro-Santos[10,11], Afranio Kritski[12], Betina Durovni[3], Solange Calvacante[3,12], Marina C. Figueiredo[4], David W. Haas[4,13], Dandan Liu[2*], Valeria C. Rolla[3*], Timothy R. Sterling[4*], and the Regional Prospective Observational Research in Tuberculosis (RePORT)-Brazil network[^]

[1]Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA
[2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA
[3]Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil
[4]Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA
[5]Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Instituto Brasileiro para Investigação da Tuberculose, Fundação José Silveira, Salvador, Bahia, Brazil
[6]Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil
[7]Universidade Salvador (UNIFACS), Laureate Universities, Salvador, Bahia, Brazil
[8]Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador, Bahia, Brazil
[9]Faculdade de Tecnologia e Ciências (FTC), Salvador, Bahia, Brazil
[10]Fundação Medicina Tropical Dr. Heitor Vieira Dourado, Manaus, Brazil
[11]Universidade do Estado do Amazonas (UEA), Manaus, Brazil
[12]Universidade Federal do Rio de Janeiro, Faculdade de Medicina, Rio de Janeiro, Brazil
[13]Department of Internal Medicine, Meharry Medical College, Nashville, TN, USA

*These authors contributed equally to this work

^Membership of the RePORT-Brazil Network is in the Acknowledgements

Corresponding author: Lauren S. Peetluk, MPH; A2209 Medical Center North, 1161 21st Avenue South, Nashville, TN, USA, 37203, E-mail: l.peetluk@vumc.org

Summary: We detail the rigorous development and internal validation of a prognostic model, including seven easily collected variables that accurately predict unsuccessful pulmonary tuberculosis treatment outcome. The model can be applied at point-of care with a nomogram or web application.

Abstract

Background: Despite widespread availability of curative therapy, tuberculosis treatment outcomes remain suboptimal. Clinical prediction models can inform treatment strategies to improve outcomes. Using baseline clinical data, we developed a prediction model for unsuccessful TB treatment outcome and evaluated the incremental value of HIV-related severity and isoniazid acetylator status.

Methods: Data originated from the Regional Prospective Observational Research for Tuberculosis Brazil cohort, which enrolled newly-diagnosed tuberculosis patients in Brazil from 2015-2019. This analysis included participants with culture-confirmed, drug-susceptible pulmonary tuberculosis who started first-line anti-tuberculosis therapy and had ≥12 months of follow-up. The endpoint was unsuccessful tuberculosis treatment: composite of death, treatment failure, regimen switch, incomplete treatment, or not evaluated. Missing predictors were imputed. Predictors were chosen via bootstrapped backward selection. Discrimination and calibration were evaluated with c-statistics and calibration plots, respectively. Bootstrap internal validation estimated overfitting, and a shrinkage factor was applied to improve out-of-sample prediction. Incremental value was evaluated with likelihood ratio-based measures.

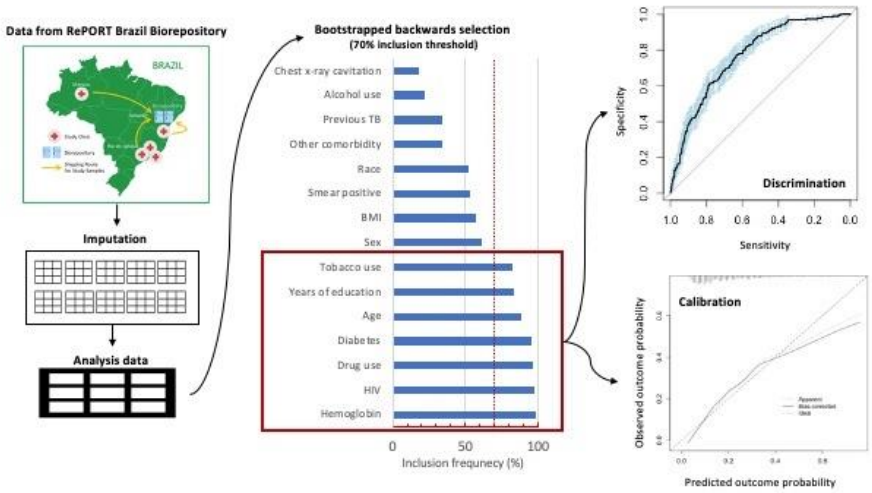
Results: Of 944 participants, 191 (20%) had unsuccessful treatment outcomes. The final model included seven baseline predictors: hemoglobin, HIV-infection, drug use, diabetes, age, education, and tobacco use. The model demonstrated good discrimination (c-statistic=0.77; 95% confidence interval: 0.73-0.80) and was well-calibrated (optimism-corrected intercept and slope: -0.12 and 0.89,

respectively). HIV-related factors and isoniazid acetylation status did not improve prediction of the final model.

Conclusions: The prediction model, using information readily available at treatment initiation, performed well in this population. The findings may guide future work to allocate resources or inform targeted interventions for high-risk patients.

Keywords: pulmonary tuberculosis; prognosis; prediction model; epidemiologic research; HIV coinfection

Graphical_abstract



Introduction

Tuberculosis (TB) is a leading cause of death worldwide.[1] Despite widespread availability of effective drugs, TB treatment outcomes remain suboptimal. The World Health Organization (WHO) estimated the global TB treatment success rate was 85% in 2018. [1] A recent study in Brazil, a high-burden TB country, found that from 2015-2019 only 67% of TB patients reported to the Brazilian National TB Program Notifiable Disease Information System were successfully treated, whereas 20% were lost to follow-up or transferred, 9% died, and 4% failed treatment or relapsed.[2] This is well below the End TB Strategy goal of 90% treatment success by 2025.[1,3]

To improve overall treatment outcomes, clinicians and researchers should swiftly identify patients most likely to have unsuccessful treatment, then augment treatment or intervention strategies to support them. One approach to identify high-risk patients is clinical prediction modeling, which estimates an individual's risk of a specific endpoint within a defined time period.[4] In a recent systematic review, 33 studies presenting 37 prediction models for end-of-treatment TB outcomes were identified.[5] All models suffered bias, due to poor reporting of study population and data collection, exclusion of missing data, univariate analysis-based model selection procedures, lack of validation, or limited generalizability.

With data available at treatment initiation, we developed and internally validated a prediction model for unsuccessful pulmonary TB treatment outcomes among patients with culture-confirmed, drug-susceptible, pulmonary TB who were treated with standard anti-TB therapy. Additionally, given the strong effect of HIV-infection on TB treatment outcomes,[6] and the importance of isoniazid metabolism for safety and efficacy of TB treatment,[7] we evaluated the incremental value of HIV-related severity measures and isoniazid acetylator status to provide insight about the importance of collecting these data in routine care.

Materials and methods

Study design and population

This study used data from The Regional Prospective Observational Research for Tuberculosis (RePORT) Brazil cohort, a prospective study of TB patients at five sites across three regions in Brazil: three in Rio de Janeiro (Instituto Nacional de Infectologia Evandro Chagas, Clínica de Saúde Rinaldo Delmare, Secretaria de Saúde de Duque de Caxias), one in Salvador (Instituto Brasileiro para Investigação da Tuberculose), and one in Manaus (Fundação Medicina Tropical Dr. Heitor Vieira Dourado).[2] Participants were consecutively enrolled at each site from June 2015 – June 2019 with active follow-up through June 2020. The population of RePORT-Brazil is broadly representative of TB cases in Brazil.[2]

RePORT-Brazil participants with newly diagnosed, culture-confirmed, drug-susceptible pulmonary TB, who were ≥18 years-old and started a standard first-line anti-TB therapy regimen within the last 7 days were included. Standard anti-TB therapy included isoniazid, rifampin or rifabutin, pyrazinamide and ethambutol for two months, followed by isoniazid and rifampin for four months.[8] Participants who received anti-TB therapy for ≥7 days within 30 days of enrollment, received >7 days of fluoroquinolone therapy within 30 days of enrollment, were pregnant or breastfeeding, or did not plan to remain in the region during follow-up were excluded.

Standardized clinical, demographic, and outcome information were collected longitudinally at three clinical visits (TB treatment initiation (baseline), two months after initiating treatment, and end of TB treatment) and via telephone follow-up every six months until 24 months.[9,10] Methods and results are reported according to the Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (**Supplementary File 1**).[11]

Outcomes and predictors

Participants were assigned one of six mutually exclusive TB treatment outcomes, based on new WHO guidelines (**Table 1**)[12]. The outcomes were cure, treatment completion, treatment failure, death due to any cause, treatment incomplete, and not evaluated. Cure and treatment completion were collectively considered successful TB treatment and were the referent outcome. Treatment failure, death due to any cause, treatment incomplete, and not evaluated were collectively considered unsuccessful TB treatment, which was the outcome of interest. Standard TB treatment typically lasts 6 months, but we allowed follow-up though one year to ensure completion of treatment. For a 20% outcome rate and 12-21 candidate predictors, we estimated between 710 and 1242 participants were needed for sufficient precision of the prediction; thus, developing the model using available data was well justified.[13]

Candidate predictors were selected *a priori* using previous TB prediction models [5] and clinical input from co-authors. Fifteen baseline candidate predictors were considered: age, sex, self-reported race, years of formal education, body mass index (BMI), previous TB, cavitation on chest radiograph, smear positive, HIV-infection, diabetes (self-reported history of diabetes or glycated hemoglobin ≥ 6.5%),[14] hemoglobin, any other chronic disease comorbidity, tobacco use, drug use, and alcohol use. Full definitions in **Supplementary File 2**.

Model development and validation

Model development and validation are detailed in **Figure 1**. Missing values for predictors were multiply imputed over 10 iterations. Imputed values were averaged by taking the median of continuous variables and mode of categorical variables across all 10 imputed datasets, and summarized in one complete dataset for primary analysis.[15] Sensitivity analyses were performed replicating all model selection and validation steps in complete case analysis and each imputed dataset (**Supplementary File 3**).

Candidate predictors were examined for collinearity with redundancy analysis based on Hoeffding D statistic.[16] Non-linearity was evaluated with a chunk test using restricted cubic splines with four knots for continuous predictors (age, BMI, education, and hemoglobin).[16] Age was the only non-linear variable and was therefore categorized (18-24.9, 25-34.9, 35-44.9 45-54.9, and ≥55 years) for primary analysis, and modeled using a restricted cubic spline in sensitivity analyses (**Supplementary File 4**).

The primary model building process used logistic regression for a binary endpoint. Bootstrapped backward selection was used to identify the most important set of predictors for unsuccessful TB treatment.[17] Variables selected in ≥70% of the 500 bootstraps were included in the final model.[15,18] Sensitivity analyses evaluated a 50% threshold.

Performance was evaluated with discrimination and calibration. Discrimination was quantified with the c-statistic.[19] Calibration was assessed using calibration plot, calibration intercept, calibration slope, and a Hosmer-Lemeshow goodness-of-fit test.[20] Overall model fit was evaluated with the Brier score. Internal validation with bootstrap resampling was used to estimate optimism-corrected performance measures.[16] Predictions from the final model account for shrinkage according to the heuristic shrinkage factor, estimated as $\chi^2_{model} - df / \chi^2_{model}$, where "$\chi^2_{model}$" is the chi-square model and "df" is the degrees of freedom.[21] Coefficients and model performance

for the main model building process were compared to model approximation[16] and least absolute shrinkage and selection operator (LASSO).[22] (**Supplementary File 5)**

We conducted decision curve analysis to evaluate the net-benefit of using the prediction model to inform care decisions across a range of threshold probabilities[23,24] (additional methods in **Supplementary File 6)**. We suggest cut-offs, defined *a priori*, for three risk groups based on clinical relevance and potential utility in future studies, including sensitivity, specificity, positive predictive value, and negative predictive value.

Comparison with existing TB outcome prediction models

Of 37 TB outcome prediction models identified in a recent systematic review,[5] only one could be externally validated. Definitions of predictors were matched to the original study, (**Supplementary File 7**) and missing data were imputed and summarized as described above. Two methods of external validation were evaluated: 1) original model coefficients were applied to our study cohort, and 2) coefficients were updated by fitting a new model with each of the included predictors.[25] Discrimination and calibration were estimated with c-statistics and calibration plots, respectively.

Added value analysis

An *a priori* decision was made to evaluate the added value of HIV-related disease severity and *NAT2* acetylator status, given these data are not routinely collected but may be important for TB treatment outcome. *NAT2* acetylator status (fast, intermediate, and slow)[26] and three HIV-related disease severity characteristics – CD4 T-cell count <200 cells/mm^3, plasma HIV-1 RNA ≥200 copies/mL, and ART experience (experienced vs. naïve) – were added to the final model (**Supplementary File 2**). Added value was quantified using likelihood-based measures, net-reclassification index, and integrated discrimination index.[27,28]

R version 4.9.1 was used for analysis. Code is available online: https://github.com/lspeetluk/report-tb-pm.

Results

The study included 944 culture-confirmed, drug-susceptible pulmonary TB patients who started standard anti-TB therapy. Median age was 35 years (interquartile range (IQR): 25-49), 34% were female, 80% were non-white, and 19% were persons living with HIV (PLWH) (**Table 2**).

Of the 944 participants, 191 (20%) had an unsuccessful outcome. Median time to unsuccessful outcome was 107 days (IQR: 41-176 days), compared to median time to successful outcome of 186 days (IQR: 179-205 days). Overall, 914 (97%) of 944 participants had complete data for every predictor.

After bootstrapped backward selection, the most important predictors of unsuccessful treatment were hemoglobin, HIV-positivity, drug use, diabetes, age, education, and tobacco use (**Table 3**). The model demonstrated good discrimination with a c-statistic of 0.77 (95% confidence interval (CI): 0.73-0.80) (**Figure 2A**), and good calibration, with a near-diagonal calibration curve at predicted risks below 0.40 (**Figure 2B**); departure from diagonal above 0.40 is likely due to lack of very high-risk patients. The Brier score was 0.14, and Hosmer-Lemeshow goodness-of-fit test p-value was 0.1. The model showed good internal validation with optimism-corrected c-statistic of 0.75 (95% CI: 0.71-0.78)**;** optimism-corrected calibration intercept and slope were -0.12 and 0.89, respectively.

Variable selection and model performance results were consistent in complete case analysis within each imputed dataset (**Supplementary File 3**) and sensitivity analyses using a spline for age (**Supplementary File 4**). Model performance was similar using a 50% inclusion threshold, model approximation, and LASSO. (**Supplementary File 5**).

After applying the heuristic shrinkage factor of 0.91, predicted risks from the final model can be applied to new populations using a nomogram (**Figure 3**), web-based application (https://lauren-peetluk.shinyapps.io/tb_outcome_risk_calculator), or the following formula:

Probability(unsuccessful=1) = 1/(1+exp(-**X**β*0.91)), where **X**β =

0.66 – 0.18*[hemoglobin] + 0.71*[HIV] + 0.50*[former drug use]

+ 1.19*[current drug use] + 0.65*[diabetes] – 0.48*[age 25-35] – 0.71*[age 35-45]

– 1.99*[age 45-55] – 0.46*[age 55+] – 0.06*[years of education]

+ 0.63*[former smoker] + 0.56*[current smoker]

Application of the model to inform risk-based interventions is recommended when the cost-benefit ratio of the intervention under consideration is between 1:9 and 2:3 or when the risk threshold at which intervention is considered is between 11 and 41% (**Figure 4**; **Supplementary File 6)**. Sensitivity, specificity, positive predictive value, and negative predictive value across three potential risk groups are in **Table 4**.

The prediction model developed in this study performed well compared to external validation of the Costa-Veiga model, which had, at best, a c-statistic of 0.68 (95% CI: 0.64-0.71).[29] (**Supplementary File 7**) Of HIV-related and *NAT2* acetylator status variables, only ART experience added notable value to the final prediction model with a net reclassification index of 0.24 and ~ 3% new information gained (**Table 5**).

Discussion

This analysis describes the development and internal validation of a prediction model for unsuccessful outcome in a prospective cohort of Brazilian patients with culture-confirmed, drug-susceptible, pulmonary TB. The prediction model combined seven variables, including demographics, clinical characteristics, and a single laboratory parameter, all of which are widely available in clinical settings at the time of TB treatment initiation. Results were robust to different methods of handling non-linearity of age, missing data, and were consistent across several variable selection techniques. Individual risk from the final model can be easily calculated in clinical settings with the provided nomogram, risk formula, or online calculator. The model performed well compared to an existing

prediction model,[29] with improved discrimination and calibration. Internal bootstrap validation indicated slight overfitting of the model, but external validation is necessary prior to model implementation in any new setting.[16]

In the model, hemoglobin, HIV-infection, drug use, and diabetes were the strongest predictors of unsuccessful outcome. These factors have been consistently reported as associated with unsuccessful TB outcome. Anemia is linked to worse prognosis and elevated mortality following TB diagnosis.[30] It has also been suggested as an alternative to CD4 T-cell counts in TB/HIV co-infected populations, given its low cost, wide availability, and possibly similar predictive value.[31]

There are clear clinical implications of TB-HIV co-infection, including the effect of HIV-infection on TB outcomes. In Brazil, 11% of TB cases were co-infected with HIV, yet PLWH accounted for 28% of TB deaths.[1] Factors such as poverty, undernutrition, poor access to healthcare services, and crowded living conditions additionally interact with both TB and HIV-infection to fuel worse treatment outcomes for both, and concomitant treatment of TB and HIV is difficult due to drug-drug interactions and increased pill burdens.[32] Despite these complexities, the only factor related to HIV-infection that added minor predictive value in this cohort was ART experience, but the power to detect differences in the models may have been limited by sample size.

Research has additionally suggested a synergistic relationship between TB and diabetes: diabetes increases risk of TB and impacts TB treatment outcomes.[33] The definition of diabetes used in this study was based on self-reported history of diabetes and baseline hyperglycemia (glycated hemoglobin ≥6.5), which may have influenced observed results. Several recent studies have suggested that TB causes dysregulation in blood glucose levels, which can manifest as transient pre-diabetes or frank diabetes. Although this resolves with treatment in some patients, it may still confer increased risk of unsuccessful TB outcomes.[34]

The impact of behavioral factors on TB treatment outcomes have long been explored, such as drug, alcohol, and tobacco use. Most studies suggest that drug use and alcohol use contribute to poor adherence, treatment discontinuation, and loss to follow-up.[35] However, mechanistic effects, such as compromising the immune system, may also play a role in treatment outcomes.[36] Because a composite endpoint was used, we could not assess whether drug use affects TB treatment outcomes by way of poor retention in care or due to biologic reasons, but regardless, it was an important predictor of TB treatment outcome.

We found no association between *NAT2* acetylator status and TB treatment outcome. Previous research suggests that slow acetylators have increased risk of drug-induced hepatotoxicity, whereas fast acetylators may not achieve target bactericidal activity, increasing risk for therapeutic failure or relapse.[7,26] However, most studies of acetylator status have primarily focused on toxicity endpoints, rather than end of treatment outcomes, and more research in this area is needed. Participants with treatment-associated toxicities may have undergone dose alterations without complete regimen modification or treatment failure and would not have been captured as unsuccessful outcomes in this study.

There were limitations to this study. First, the model used a composite endpoint. This reflects the commonly used TB treatment outcome definition recommended by the WHO and provides a general risk estimate for unsuccessful outcome. However, there may be heterogeneous predictors for biologic (death, treatment failure, regimen switch) and behavioral outcomes (incomplete treatment, not evaluated), and analytic methods that account for competing risks or multiple outcomes should be considered in future studies.[37] Second, rather than using Rubin's rules for multiple imputation, data were averaged across imputations into a single dataset. There exists debate about how to develop and validate prediction models with multiply imputed data; some recommended approaches suggest using stacked datasets or weighting methods, which were not explored in this study due to complexities in using them together with bootstrapped backward

selection.[21,38] However, given low amounts of missing data and consistency of the current approach with each individually imputed dataset and in complete case analysis, there is confidence in the current approach. Third, some of the variables included in the model may not be available in all settings, such as surveillance data, which do not typically include laboratory parameters. Fourth, only one existing prediction model could be externally validated with our data, though several others may apply to this setting. Fifth, external validation of the developed prediction model was not performed due to lack of data for a validation population, but we conducted comprehensive internal validation with the bootstrap.[16] Regardless, it is not yet clear how generalizable the model will be in other settings. Notably, RePORT-Brazil belongs to the RePORT international consortia (reportinternational.org), including sites in India, South Africa, China, Indonesia, and the Philippines. All sites are expected to begin data harmonization efforts in the coming years, which will provide opportunities for future external validation.[9]

Strengths of the study include that RePORT Brazil is a moderately-sized prospective cohort study of drug-susceptible TB patients in Brazil, who are representative of cases throughout Brazil.[2] All variables included are easily collected and should be readily known at time of TB treatment initiation. The steps of prediction model development adhered to best practice guidelines for developing prognostic models, and sensitivity analyses confirmed the robustness of results.[4,11] The bootstrapped backwards selection strategy is a preferred alternative to univariable stepwise modeling approaches.[17,18] The proportion of times a variable was retained in the model provides information about the importance of that predictor; variables included more frequently likely have a stronger independent association with the outcome than variables included less often. Performance measures were corrected for overfitting using bootstrapped internal validation, and final model coefficients were multiplied by a heuristic shrinkage factor, which both aimed to correct for overfitting and improve performance in future datasets.[16,25] Additionally, the model performed favorably compared to an existing prediction model with good discrimination, calibration, and performance.

Regarding implementation, this model may be useful to target interventions or modified treatment regimens, such as those of shorter duration but higher cost, to individuals at high-risk of unsuccessful outcome. The practicality and acceptability of risk thresholds is context-specific and should be considered relative to the expected cost-benefit of the intervention/treatment strategy. Several recent systematic reviews and meta-analyses have evaluated the impact of interventions on TB treatment outcomes, including DOT, novel digital health technologies, patient education and counseling, and financial support (e.g. cash transfer programs), all of which show potential to improve treatment outcomes and could be considered in future risk-based intervention studies.[39]

Conclusion

In conclusion, this paper details the development and internal validation of a straightforward and easy-to-use clinical prediction model that combines seven routinely available predictors to estimate individual risk of an unsuccessful TB treatment outcome among drug-susceptible, pulmonary TB patients on standard therapy. The model can be implemented with pen and paper, a nomogram, or using a web application. Though the model requires external validation prior to widespread implementation in any new setting, the individual risks derived from the model may be useful in future studies to allocate resources or target interventions to patients at the highest risk of unsuccessful outcomes.

NOTES

Author contributions

LSP conceptualized the research question, conducted the analysis, and drafted the initial manuscript. DL, VR, and TRS provided thorough feedback on the research design and analysis interpretation, supervised the analysis, and revised successive drafts of the manuscript. PFR assisted with methodology conceptualization, analysis interpretation and revised successive manuscript drafts. DH conducted genetic analysis and revised successive manuscript drafts. FR, BA, MCS, AK, BD, SC, and MCF played pivotal roles in the conceptualization of the RePORT Brazil cohort, project administration, data and funding acquisition, and revised successive drafts of the manuscript. All authors approved of the final version of the manuscript.

Acknowledgements

The authors are grateful for the support and dedication of study staff at each enrollment site who made the study possible, including all members of the RePORT-Brazil consortium: Renata Spener-Gomes, Alexandra Brito de Souza, Jaquelane Silva Jesus, Aline Benjamin, Flavia Marinho Sant'Anna, Francine Peixoto Ignácio, Maria Cristina Lourenço, Adriano Gomes-Silva, Jamile G. de Oliveira, Adriana S. R. Moreira, Anna Cristina Calçada Carvalho, Elisangela C. Silva, Mayla Mello, Michael S. Rocha, Betania Nogueira, Vanessa Nascimento, Saulo Nery, Alice M. S. Andrade, Hayna Malta-Santos, Jéssica Rebouças-Silva, André M.C. Ramos, Sayonara Melo, Juan M. Cubillos-Angulo, and Laise de Moraes.

Funding support

This work was supported by by the Departamento de Ciência e Tecnologia (DECIT) - Secretaria de Ciência e Tecnologia (SCTIE) – Ministério da Saúde (MS), Brazil [25029.000507/2013-07 to V.C.R.], National Institutes of Allergy and Infectious Diseases at the National Institutes of Health (NIH/NIAID): [U01 AI069923, R01 A1120790, K01 AI131895 to P.F.R, F31 AI152614 to L.S.P, AI077505, AI110527, AI120790, and TR002243 to D.W.H.], and the National Center for Advancing Translational Sciences [CTSA Award No. TL1 TR002244 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. V.C.R., T.S., S.C., M.C.S., M.C.F., D.W.H., F.M.R., A.K., B.D., and B.B.A. report grants U01 AI069923 (CCASAnet; RePORT-Brazil supplement)/R01 AI20790 from NIAID during the conduct of the study. L.P. reports CTSA Award No. TL1 TR002244 - training grant from NCATS and U01 AI069923 (CCASAnet; RePORT-Brazil supplement), R01 AI20790, and F31 AI152614 – fellowship award from NIAID, during the conduct of the study. P.F.R. reports K01 AI131895 (career development award) during the conduct of the study and reports grant R21 AI145686 for registry linkage to improve mortality ascertainment among HIV cohorts in Latin America, and single honorarium for participation in expert panel on HIV care in 2020 from Gilead, outside the submitted work. T.R.S. reports royalties from UpToDate for Chapters on management of TB, outside the submitted work. V.C.R. reports individual payment for conference presentation from Brazilian Society of Infectious Diseases and serving on DSMB for Division of AIDS, outside the submitted work.

Conflicts of interest: None

References

1. Global Tuberculosis Report 2020. Geneva: World Health Organization, 2020.

2. Arriaga M, Amorim G, Queiroz A, et al. Novel Stepwise Approach to Assess Representativeness of a Large Multicenter Observational Cohort of Tuberculosis Patients: The Example of RePORT Brazil. International Journal of Infectious Diseases **2020**; 11:140.

3. The End TB Strategy. Geneva: World Health Organization, 2015.

4. Steyerberg EW, Moons KGM, Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Medicine **2013**; 10:e1001381.

5. Peetluk LS, Ridolfi FM, Rebeiro PF, Liu D, Rolla VC, Sterling TR. Systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults. BMJ Open **2021**; 11:e044687.

6. Mayer KH, Dukes Hamilton C. Synergistic Pandemics: Confronting the Global HIV and Tuberculosis Epidemics. Clinical Infectious Diseases **2010**; 50:S67–S70.

7. Ramachandran G, Swaminathan S. Role of pharmacogenomics in the treatment of tuberculosis: a review. Pharmacogenomics and Personalized Medicine **2012**; :10.

8. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis. Manual de Recomendações para o Controle da Tuberculose no Brasil. 2019; Available at: http://www.aids.gov.br/pt-br/pub/2019/manual-de-recomendacoes-para-o-controle-da-tuberculose-no-brasil.

9. Hamilton CD, Swaminathan S, Christopher DJ, et al. RePORT International: Advancing Tuberculosis Biomarker Research Through Global Collaboration. Clin Infect Dis **2015**; 61:S155–S159.

10. 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.

11. 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–W73.

12. Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis, 27-29 October 2020. 2021.

13. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ **2020**; :m441.

14. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care **2014**; 37:81–90.

15. Heymans MW, van Buuren S, Knol DL, van Mechelen W, de Vet HC. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Med Res Methodol **2007**; 7:33.

16. Jr FEH. Regression Modeling Strategies. 2nd ed. Basel, Switzerland: Springer International Publishing, 2015.

17. Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. Biometrical Journal **2018**; 60:431–449.

18. Austin PC, Tu JV. Bootstrap Methods for Developing Predictive Models. null **2004**; 58:131–137.

19. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology **2010**; 21:128–138.

20. Calster BV. A calibration hierarchy for risk models was defined: from utopia to empirical data. Journal of Clinical Epidemiology **2016**; :10.

21. Vergouwe Y, Royston P, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. Journal of Clinical Epidemiology **2010**; 63:205–214.

22. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **2011**; 73:273–282.

23. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res **2019**; 3:18.

24. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. JCO **2016**; 34:2534–2540.

25. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Statist Med **2004**; 23:2567–2586.

26. Haas DW, Podany AT, Swindells S, et al. Pharmacogenetic interactions of rifapentine plus isoniazid with efavirenz or nevirapine. Pharmacogenet genomics **2021**; 31:11.

27. Harrell Jr. FE. Statistically Efficient Ways to Quantify Added Predictive Value of New Measurements. 2019. Available at: https://www.fharrell.com/post/addvalue/. Accessed 23 February 2019.

28. Cook NR. Quantifying the added value of new biomarkers: how and how not. Diagnostic and Prognostic Research **2018**; 2:14.

29. Costa-Veiga A, Briz T, Nunes C. Unsuccessful treatment in pulmonary tuberculosis: Factors and a consequent predictive model. European Journal of Public Health **2018**; 28:252–258.

30. Isanaka S, Mugusi F, Urassa W, et al. Iron Deficiency and Anemia Predict Mortality in Patients with Tuberculosis. The Journal of Nutrition **2012**; 142:350–357.

31. Kerkhoff AD, Wood R, Cobelens FG, Gupta-Wright A, Bekker L-G, Lawn SD. The predictive value of current haemoglobin levels for incident tuberculosis and/or mortality during long-term antiretroviral therapy in South Africa: a cohort study. BMC Med **2015**; 13:70.

32. Panel on Opportunistic Infections in Adults and Adolescents with HIV. Guidelines for the prevention and treatment of opportunistic infections in adults and adolescents with HIV: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. : V.9-V.24. Available at: https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/Adult_OI.pdf.

33. Baker MA, Harries AD, Jeon CY, et al. The impact of diabetes on tuberculosis treatment outcomes: A systematic review. **2011**; :15.

34. Dooley KE, Chaisson RE. Tuberculosis and diabetes mellitus: convergence of two epidemics. The Lancet Infectious Diseases **2009**; 9:737–746.

35. Rehm J, Samokhvalov A, Neuman MG, et al. The association between alcohol use, alcohol use disorders and tuberculosis (TB). A systematic review. BMC Public Health **2009**; 9:1–12.

36. Deiss RG, Rodwell TC, Garfein RS. Tuberculosis and Illicit Drug Use: Review and Update. Clin Infect Dis **2009**; 48:72–82.

37. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks: Competing Risks and Risk Scores. Statist Med **2016**; 35:4056–4072.

38. Wahl S, Boulesteix A-L, Zierer A, Thorand B, van de Wiel MA. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. BMC Med Res Methodol **2016**; 16:144.

39. Richterman A, Steer-Massaro J, Jarolimova J, Luong Nguyen LB, Werdenberg J, Ivers LC. Cash interventions to improve clinical outcomes for pulmonary tuberculosis: systematic review and meta-analysis. Bull World Health Organ **2018**; 96:471–483.

Table 1. World Health Organization definitions of TB treatment outcomes for drug-susceptible patients on first-line drug regimen.

Outcome	Definition
Cure	A patient with bacteriologically confirmed TB at the beginning of treatment, who completed treatment as recommended by the national policy, and who had evidence of bacteriologic response (a negative smear or culture at the end of TB treatment and on at least one previous occasion more than 7 days apart)
Treatment completion	A patient who completed treatment as recommended by national policy but whose outcome does not meet definition for cure or failure, either because tests were not done or because results are unavailable
Treatment success	Composite of cured and treatment completed
Treatment failure	A patient whose regimen needed to be terminated due to lack of clinical response (sputum smear or culture is positive at month 5 or later during treatment) or whose regimen was permanently changed to new regimen or treatment strategy
Death	A patient who died for any reason during the course of TB treatment
Treatment incomplete	A patient whose TB treatment was interrupted for 2 consecutive months or more
Not evaluated	A patient for whom no TB treatment outcome was assigned, which includes cases who "transferred out" to another treatment unit as well as cases for whom the treatment outcome is unknown to the reporting unit

Table 2. Study population characteristics

Characteristic	N	N (%) or Median [IQR]
Age, years	944	35 [25, 49]
Female sex	944	317 (34)
Non-white race	943	751 (80)
Education, years	943	9.0 (6.0, 12.0)
BMI, kg/m^2	944	20.2 [18.3-22.5]
Previous TB diagnosis	935	142 (15)
Cavitation on chest x-ray	937	468 (50)
Smear positive	943	768 (81)
Diabetes	945	235 (25)
HIV-infection	938	182 (19)
ART experienced	182	71 (39)
CD4 count, cells/mm^3	175	126 [54-274]
CD4 count <200	175	116 (66)
Viral load, copies/mL	169	32,183 [309-205,226]
Viral load ≥ 200	169	130 (77%)
Hemoglobin	939	12.10 [10.7-13.3]
Any disease comorbidity*	944	143 (15)
Alcohol use	944	
Current		426 (45)
Former		366 (39)
Never		152 (16)
Drug use	943	
Current		117 (12)
Former		203 (22)
Never		623 (66)
Tobacco use	944	
Current		216 (23)
Former		273 (29)
Never		455 (48)
***NAT2* acetylator status**	944	

Rapid		78 (8%)
Intermediate		382 (40%)
Slow		484 (51%)
Treatment outcome category	944	
Cure		386 (41)
Treatment completion		367 (39)
Death		29 (3)
Treatment failure		43 (4)
Treatment incomplete		56 (6)
Not evaluated		63 (7)
Unsuccessful outcome		191 (20)

Italics indicate non-mutually exclusive groups, describing characteristics only among persons living with HIV

*Excluding diabetes and HIV-infection

Table 3. Results from bootstrapped backwards selection (x500) and final model (70% inclusion frequency) after imputation for missing values (N=944)

	Bootstrap inclusion frequency (%)	Coefficient	Standard error
Intercept	100.00	0.66	0.63
Hemoglobin	98.40	-0.18	0.05
HIV-infection	97.00	0.71	0.22
Former drug use	96.80	0.50	0.24
Current drug use	96.80	1.19	0.28
Diabetes	95.60	0.65	0.21
Age group 25-34.9	88.60	-0.48	0.25
Age group 35-44.9	88.60	-0.71	0.27
Age group 45-54.9	88.60	-1.09	0.33
Age group ≥55	88.60	-0.46	0.32
Years of education	83.00	-0.06	0.02
Former smoker	82.40	0.63	0.23
Current smoker	82.40	0.56	0.25
Female sex	60.80	..	..
BMI	57.40	..	..
Smear positive	53.00	..	..
Non-white race	52.00	..	..
Other comorbidity	34.40	..	..
Previous TB	34.20	..	..
Former alcohol use	22.40	..	..
Current alcohol use	22.40	..	..
Chest x-ray cavitation	18.00	..	..

Table 4. Classification measures for low, medium, and high risk groups.

Risk group	Predicted risk thresholds	N observed unsuccessful outcomes / N in group (%)	Sensitivity	Specificity	PPV	NPV
Low	<10%	13/307 (4%)	1.00	-	1.00	0.00
Medium	10-20%	47/277 (17%)	0.93	0.39	0.28	0.96
High	≥20%	131/360 (36%)	0.69	0.69	0.36	0.90

Abbreviations: NPV, negative predictive value; PPV, positive predictive value

Footnote: Classification measures are calculated considering true positives as individuals who experienced an unsuccessful outcome at or above the selected risk group, true negatives are individuals who experienced a successful outcome below the selected risk group, false positives are individuals who experienced a successful outcome but are at or above the selected risk group, and false negatives are individuals who experienced an unsuccessful outcome but are below the selected risk group. The intention of this table is to provide a range of estimates that can serve as clinical cut-points in future studies or clinical practice. For example, we could target novel adherence technologies to individuals in the high-risk group (n=360), which comprises 36% of the population but 69% of all unsuccessful outcomes, or the medium and high risk groups, which comprise 53% of the population, but 93% of all unsuccessful outcomes. More details on how to consider various risk thresholds from decision curve analysis are detailed in E6 Figure.

Table 5. Added value of HIV-related disease severity and *NAT2* acetylator status

Metric	ART experienced	CD4 < 200	VL ≥ 200	*NAT2*
LR test p-value	0.03	0.81	0.47	0.88
NRI	0.24	<0.01	0.16	0.13
IDI	0.01	<0.01	<0.01	<0.01
FNI_{LR}	0.03	<0.01	<0.01	<0.01
FNI_{Var}	0.03	<0.01	<0.01	<0.01
FNI_{Risk}	0.03	<0.01	<0.01	<0.01

Abbreviations: FNI_{LR}, fraction of new information based on comparison of model likelihood ratios; FNI_{Var}, fraction of new information based on ratio of the variances explained by the models; FNI_{Risk}, fraction new information based on variance of predicted risk to the sum of the variance of predicted risk and the average risk; IDI, integrated discrimination improvement; NRI, net reclassification index; VL, viral load

Figure Legends:

[Figure 1]

Figure 1 legend: Schematic representation of each model development step and assessment of model performance. First, missing data were imputed across 10 datasets and summarized into an analysis dataset. Next, redundancy analysis and linearity assessment were carried out to identify highly correlated sets of variables and variables with evidence of non-linearity. Following that, 500 repetitions of bootstrapped backwards selection were used to identify the most important predictors of unsuccessful TB treatment outcome, based on those included in at least 70% of bootstrap samples. Finally, model performance was evaluated in the original sample (apparent performance) and averaged over 2000 bootstrap samples (internal bootstrap validation).

*Sensitivity analyses were conducted repeating all steps following imputation the original data with missing information (complete case analysis) and in each of the imputed datasets.

^Model performance measures included discrimination (evaluated with the c-statistic) and calibration (evaluated by a calibration plot and the calibration slope and intercept).

[Figure 2a; Figure 2b]

Figure 2 legend**:** A) The receiver operating characteristic (ROC) curve measures discrimination of the model – how well the model can differentiate between those with and without an outcome. The blue error bars represent the 95% confidence intervals for across the ROC curve, using 2000 stratified bootstrap samples. The area under the ROC curve, which is equivalent to the c-statistic, is 0.77 (95% CI: 0.73-0.80). B) The calibration plot displays agreement between observed and predicted outcome

probabilities across deciles of outcome risk. An ideal calibration curve has an intercept of 0 and a slope of 1 (dashed line). The apparent calibration (dotted line) is calibration of the model in the original data, and the bias-corrected line is corrected for overfitting using 200 bootstrap samples. The bias-corrected calibration intercept and slope were -0.12 and 0.89, respectively. The top of the plot displays a histogram of the distribution of predicted probabilities of unsuccessful outcome for the 944 culture-confirmed, drug-susceptible, pulmonary TB patients included in the study.

[Figure 3]

Figure 3 legend: The nomogram can be used in clinical settings to estimate individual risk of an unsuccessful TB outcome. For example, for an individual who is 50-years old with diabetes as their only comorbidity, hemoglobin of 13 g/dL, 12 years of education, never drug use, and current tobacco use, their risk is calculated as: hemoglobin=78 points, HIV-infection=0 points, drug use=0 points, diabetes=12 points, age=0 points, education=7 points, and tobacco use=10 points. Total points = 107, which equates to approximately 11% risk of an unsuccessful outcome. This is equivalent to what one would get when using the formula provided in the text: $1/(1+\exp(-X\beta*0.91))$, where $X\beta = 0.66 – 0.18*[13] + 0.71*[0] + 0.50*[0] + 1.19*[0] + 0.65*[1] – 0.48*[0] – 0.71*[0] – 1.09*[1] –0.46*[0 – 0.06*[12] + 0.63*[0] + 0.56*[1] = -2.28$. Then risk = $1/(1+\exp(-(-2.28)*0.91) = 11\%$. This is also consistent with the calculation from the web app.

[Figure 4]

Figure 4 legend: The decision curve plots the standardized net benefit (y-axis) across a variety of risk thresholds (x-axis) for three scenarios: intervene on all (All), intervene on none (None), or intervene based on predicted risk from the risk model (Risk Model). Standardized net benefit quantifies the total benefit (true positive rate) minus the total harm (false positive rate), assuming a population prevalence of unsuccessful outcome of 20% and standardized to a maximum benefit of 1.[24] The lowest y-axis indicates the cost-benefit of intervention across risk thresholds. When an intervention has low perceived cost relative to high benefit, lower risk thresholds should be considered, because the harms of unnecessary intervention are minimal compared to benefit of necessary intervention. Alternatively, as the cost-benefit of the intervention approaches 1:1, the risk threshold at which intervention should be considered increases, because the costs or harms of unnecessary intervention start to balance out the benefit of necessary intervention. The two vertical lines bound the range (risk threshold: 11%-41%) where the lower 95% confidence interval estimate of using the risk model to inform treatment/intervention decisions has a higher standardized net benefit than treating/intervening on "All" patients and treating/intervening on "None", but the exact choice of the risk threshold should be selected based on cost-benefit considerations, which are intervention-specific. Use of the risk model to inform a novel treatment or intervention strategy is expected to have net benefit (true positive rate outweighs false positive rate, assuming outcome rate of 20%) when the cost-benefit ratio of the intervention is between 1:9 and 2:3.

Figure 1

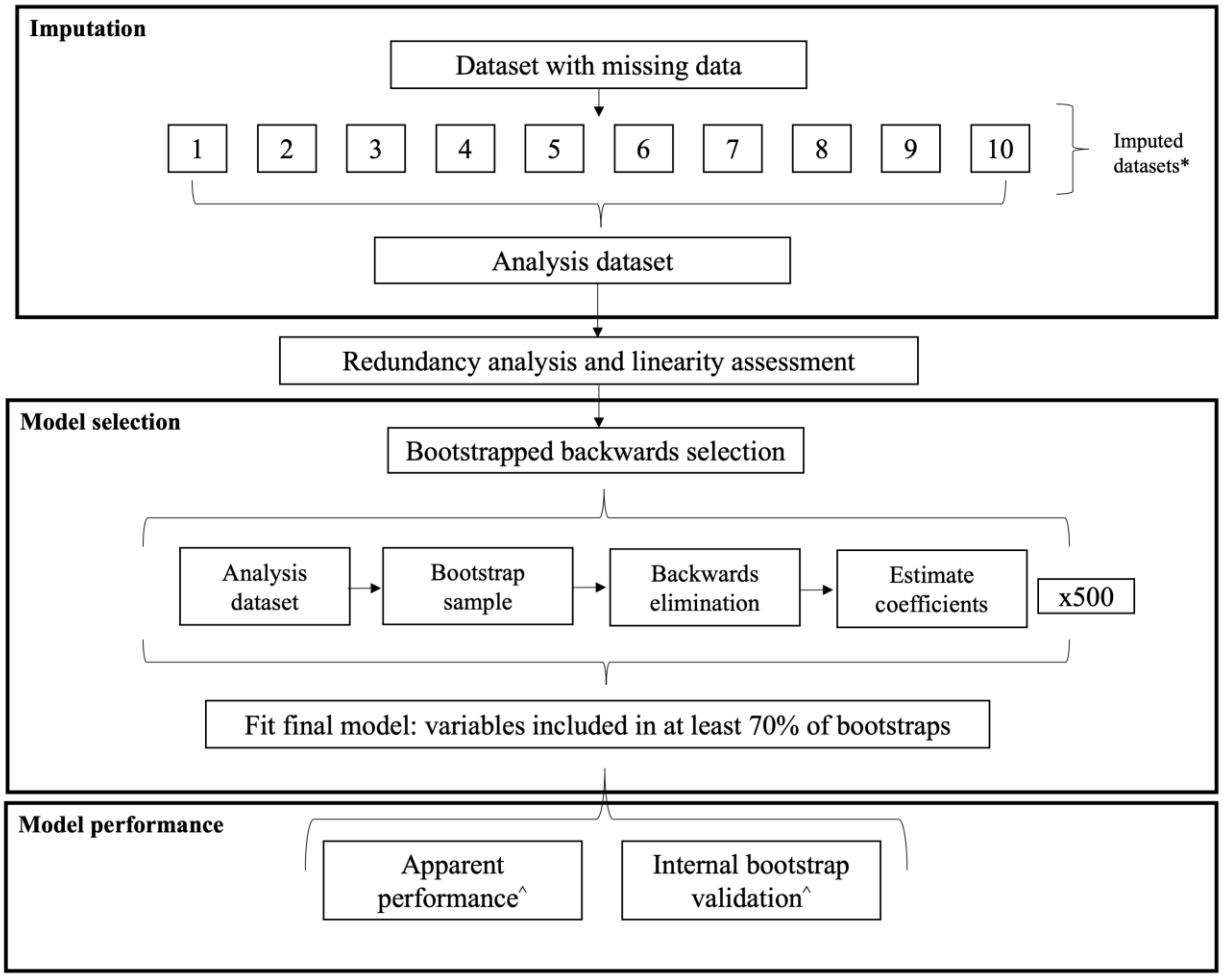


Figure 2

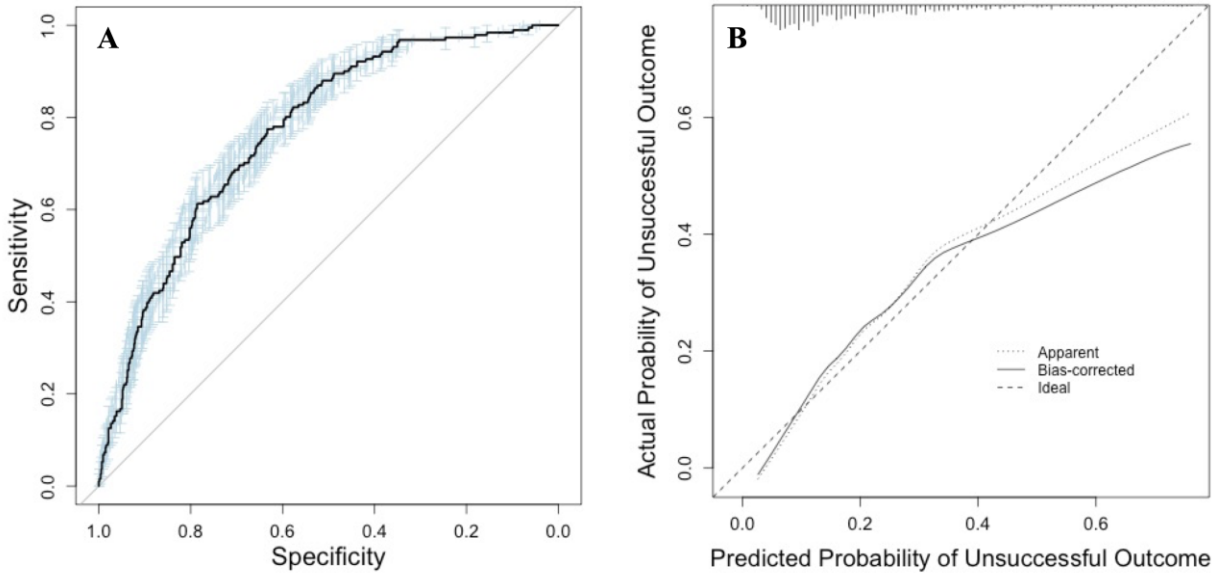


Figure 3

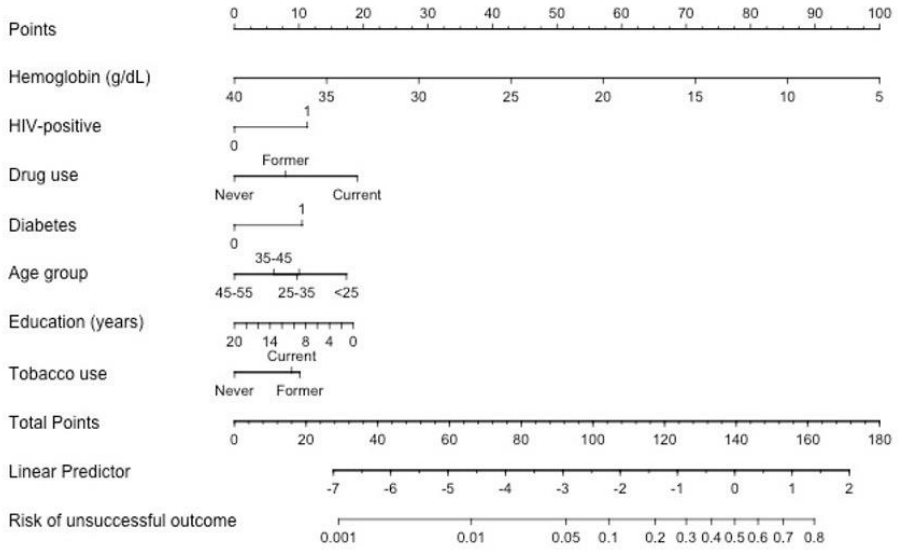


Figure 4