

Title page

Title

Discovery and validation of a personalized risk predictor for incident tuberculosis in low transmission settings

Authors

Rishi K. Gupta¹, Claire J. Calderwood¹, Alexei Yavlinsky², Maria Krutikov¹, Matteo Quartagno³, Maximilian C. Aichelburg⁴, Neus Altet^{5,6}, Roland Diel^{7,8}, Claudia C. Dobler^{9,10}, Jose Dominguez^{11,12,13}, Joseph S. Doyle^{14,15}, Connie Erkens¹⁶, Steffen Geis¹⁷, Pranabashis Haldar¹⁸, Anja M. Hauri¹⁹, Thomas Hermansen²⁰, James C. Johnston²¹, Christoph Lange^{22,23,24,25}, Berit Lange²⁶, Frank van Leth^{24,27,28}, Laura Muñoz²⁹, Christine Roder^{14,15}, Kamila Romanowski²¹, David Roth²¹, Martina Sester^{24,30}, Rosa Sloot³¹, Giovanni Sotgiu^{24,32}, Gerrit Woltmann¹⁸, Takashi Yoshiyama³³, Jean-Pierre Zellweger^{24,34}, Dominik Zenner¹, Robert W. Aldridge², Andrew Copas^{1,3}, Molebogeng X. Rangaka^{1,3,35,36}, Marc Lipman^{*37,38}, Mahdad Noursadeghi^{*39}, Ibrahim Abubakar^{*1}

**These authors contributed equally*

Affiliations

1. Institute for Global Health, University College London, London, UK
2. Institute of Health Informatics, University College London, London, UK
3. MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK
4. Department of Dermatology, Sozialmedizinisches Zentrum Ost-Donauspital, Vienna, Austria
5. Unitat de Tuberculosis, Hospital Universitari Vall d'Hebron-Drassanes, Barcelona, Spain
6. Unitat de TDO de la Tuberculosis "Servicios Clínicos", Barcelona, Spain
7. Institute for Epidemiology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
8. Lung Clinic Grosshansdorf, Airway Research Center North (ARCN), Großhansdorf, Germany
9. Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
10. Department of Respiratory Medicine, Liverpool Hospital, Sydney, Australia
11. Institut d'Investigació Germans Trias i Pujol, Badalona, Barcelona, Spain
12. CIBER Enfermedades Respiratorias, Badalona, Barcelona, Spain
13. Universitat Autònoma de Barcelona, Badalona, Barcelona, Spain
14. Department of Infectious Diseases, The Alfred and Monash University, Melbourne, Australia
15. Disease Elimination Program, Burnet Institute, Melbourne, Australia
16. KNCV Tuberculosis Foundation, The Hague, The Netherlands

17. Institute for Medical Microbiology and Hospital Hygiene, Philipps University of Marburg, Marburg, Germany
18. Respiratory Biomedical Research Centre, Institute for Lung Health, Department of Respiratory Sciences, University of Leicester, Leicester, UK
19. Hesse State Health Office, Dillenburg, Germany
20. International Reference Laboratory of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark
21. British Columbia Centre for Disease Control, Vancouver, BC, Canada
22. Division of Clinical Infectious Diseases, Research Center Borstel, Germany
23. German Center for Infection Research (DZIF), Clinical Tuberculosis Center, Borstel, Germany
24. Tuberculosis Network European Trials Group (TBnet)
25. Department of Medicine, Karolinska Institute, Stockholm, Sweden
26. Helmholtz Centre for Infection Research, Department of Epidemiology, Germany
27. Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands
28. Amsterdam University Medical Centres, location AMC, Department of Global health, Amsterdam, The Netherlands
29. Department of Clinical Sciences, University of Barcelona, Spain
30. Department of Transplant and Infection Immunology, Saarland University, Homburg, Germany
31. Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
32. Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Italy
33. Research Institute of Tuberculosis, Japan
34. Swiss Lung Association, Berne, Switzerland
35. Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa
36. Division of Epidemiology and Biostatistics, School of Public Health, University of Cape Town, Cape Town, South Africa
37. UCL-TB and UCL Respiratory, University College London, London, UK
38. Royal Free London NHS Foundation Trust, London, UK
39. Division of Infection & Immunity, University College London, UK

Equal contributions

M.L., M.N. and I.A. contributed equally to this work.

Correspondence

Professor Ibrahim Abubakar, Institute for Global Health, University College London, 3rd Floor, 30 Guilford St, London, WC1N 1EH, i.abubakar@ucl.ac.uk. ORCID: 0000-0002-0370-1430.

Abstract

The risk of tuberculosis (TB) is variable among individuals with latent *Mycobacterium tuberculosis* infection (LTBI), but validated estimates of personalized risk are lacking. In pooled data from 18 systematically-identified cohort studies from 20 countries, including 80,468 individuals tested for LTBI, 5-year cumulative incident TB risk among people with untreated LTBI was 15.6% (95% CI 8.0-29.2) among child contacts, 4.8% (3.0-7.7) among adult contacts, 5.0% (1.6-14.5) among migrants, and 4.8% (1.5-14.3) among immunocompromised groups. We confirmed highly variable estimates within risk groups, necessitating an individualized approach to risk-stratification. We thus developed a personalised risk predictor for incident TB (PERISKOPE-TB) that combines a quantitative measure of T-cell sensitization and clinical covariates. Internal-external cross-validation of the model demonstrated a random-effects meta-analysis C-statistic of 0.88 (0.82-0.93) for incident TB. In decision curve analysis, the model demonstrated clinical utility for targeting preventative treatment, compared to treating all, or no, people with LTBI. We challenge the crude current approach to TB risk estimation among people with LTBI, in favour of our evidence-based and patient-centered method, in settings aiming towards pre-elimination worldwide.

Introduction

Tuberculosis (TB) accounts for the greatest number of deaths from a single pathogen globally, with an estimated 1.5 million deaths and 10 million incident cases in 2018¹. The World Health Organization (WHO) End TB strategy ambitiously aims for a 95% reduction in TB mortality and 90% reduction in TB incidence by 2035². As part of this strategy, the priority for low transmission settings is to achieve pre-elimination (annual incidence of <1/100,000) by 2035². Preventative antimicrobial treatment for latent tuberculosis infection (LTBI) is considered critical for achieving this objective^{2,3}. In the absence of an assay to detect viable *Mycobacterium tuberculosis* (*M.tuberculosis*) bacteria, LTBI is currently clinically defined as evidence of T cell memory to *M.tuberculosis*, in the absence of concurrent disease and any prior treatment^{4,5}. Individuals with LTBI are generally considered to have a lifetime TB risk ranging from 5-10%⁴, which is reduced by 65-80% with preventative treatment⁶.

The positive predictive value (PPV) for TB using the current definition of LTBI is <5% over a two year period among risk groups such as adult TB contacts⁷⁻⁹. This may lead to a large burden of unnecessary preventative treatment, with associated risks of drug toxicity to patients, and excess economic costs to health services. The low PPV may also undermine the cascade of care, including uptake of preventative treatment among target groups, who perceive their individual risk of developing TB to be low^{10,11}. In fact, the risk of TB among individuals with LTBI is highly variable between study populations, with incidence rates ranging from 0.3-84.5 per 1,000 person-years of follow-up^{7,12}. Thus, quoting the 5-10% lifetime estimate is likely to be inaccurate for many people. Improved risk stratification is therefore essential to enable precise delivery of preventative treatment to those most likely to benefit^{5,13}. Multiple studies have shown that the magnitude of the T cell response to *M.tuberculosis* is associated with incident TB risk, raising hope that quantitative tuberculin skin test (TST) or interferon gamma release assay (IGRA) results may improve predictive ability^{14,15}. However, implementing higher diagnostic thresholds alone does not improve prediction on a population level due to a marked loss of sensitivity with this approach¹⁶.

In this study, we first sought to characterise the population risk of TB among people tested for LTBI using an individual participant data meta-analysis (IPD-MA). In order to study progression from LTBI to TB disease more accurately, we focused on settings with low transmission (defined as annual incidence $\leq 20/100,000$ persons), where there is a minimal risk of reinfection during follow-up. We confirmed highly

variable estimates of risk, necessitating an individual level approach to risk estimation. Here we develop and validate a directly data-driven personalized risk predictor for incident TB (PERISKOPE-TB) that combines a quantitative T cell response measure with key clinical covariates.

Results

Systematic review

Our systematic review identified 26 studies that aimed to assess the risk of progression to TB disease among individuals tested for LTBI in low TB transmission settings; corresponding authors of these studies were invited to contribute individual level data (Extended Data Figure 1). Of these, we received 18 individual level datasets, including participants recruited in 20 countries. The pooled dataset included a total of 82,360 individual records, of whom 51,697 had evidence of LTBI and 826 were diagnosed with TB. Of the received data, 80,468 participants (including 803 TB cases) had sufficient data for inclusion in the primary analysis (Extended Data Figure 2). The characteristics of the included study datasets are summarised in Table 1 and Supplementary Table 1. Characteristics of the eight eligible studies for which IPD were not obtained were similar to those included in the analysis (Supplementary Table 2). Eight studies recruited adults only; the remainder recruited both adults and children. The target population was recent TB contacts in nine studies^{17–25}, people living with HIV in two studies^{26,27}, mixed immunocompromised groups in two studies^{28,29}, transplant recipients in one study³⁰, mixed population screening in two studies^{31,32}, recent migrants in one study³³, and a combination of recent contacts and migrants in one study⁹. Median follow-up of all participants was 3.7 years (interquartile range (IQR) 2.1–5.3). All contributing studies reported baseline assessments for prevalent TB through routine clinical evaluations, and all included culture-confirmed and clinically diagnosed TB cases in their case definitions. Four studies had a proportion of participants lost to follow up >5%^{18,24,27,28}; baseline characteristics of those lost to follow-up were similar to those followed-up in each of these studies (Supplementary Table 3). All contributing studies achieved quality assessment scores of 6 or 7/7 (Supplementary Table 4).

Population-level analysis

In the pooled dataset, the 2-year cumulative risk of incident TB was estimated as 4.0% (2.6–6.3) among people with LTBI who did not receive preventative therapy, 0.7% (0.4–1.3) in people with LTBI who commenced preventative therapy and 0.2% (0.1–0.4) in people without LTBI (Figure 1; Supplementary Table 5). The corresponding 5-year risk of incident TB among these groups was 5.4% (3.5–8.5), 1.1% (0.6–2.0) and 0.3% (0.2–0.5), respectively.

Among untreated people with LTBI, 2-year risk of incident TB was 14.6% (7.5-27.4) among recent child (<15 years) contacts, 3.7% (2.3-6) among adult contacts, 4.1% (1.3-12) among migrants, and 2.4% (0.8-6.8) among people screened due to immunocompromise (without an index exposure). Corresponding 5-year risk was 15.6% (8.0-29.2) among recent child contacts, 4.8% (3.0-7.7) among adult contacts, 5.0% (1.6-14.5) among migrants, and 4.8% (1.5-14.3) among people screened due to immunocompromise. Among recent child contacts, risk was markedly higher among those aged <5 years, compared to those aged 5-14 years (2-year risk 26.0% (9.4-60.1) vs. 12.4% (5.7-25.6); Figure 1).

Among child contacts, 85.4% and 93.7% of cumulative risk was accrued in the first one and two years of follow-up, respectively. Among adult contacts and migrants, the annual risk also declined markedly with time. Of the cumulative 5-year risk, 58.0% and 77.5% was accrued in the first one and two years of follow-up for adult contacts, with corresponding values among migrants of 66.4% and 81.6%. There was a more even distribution of risk during follow-up in the immunocompromised group.

TB incidence rates in years 0-2 and 2-5 of follow-up, stratified by LTBI result, commencement of preventative treatment and indication for screening, are shown in Extended Data Figures 4 and 5. Within each of the risk groups assessed, incidence rates among untreated people with LTBI were markedly higher in the 0-2 year interval, compared to the 2-5 year interval, but were highly heterogeneous across studies (I^2 statistics, representing the proportion of variance that is considered due to between-study heterogeneity, ranged from 54-91% for incidence rates during the 0-2 year interval among untreated people with LTBI, when stratified by indication for screening; forest plots shown in Extended Data Figure 5). These findings suggest highly variable TB risk among people with LTBI, even within risk groups.

Prediction model development

The observed heterogeneity in TB incidence rates across studies, even after stratification by binary LTBI result, commencement of preventative treatment and indication for screening, suggests that an individual level approach to risk stratification is required. We therefore developed a personalized risk prediction model using a subset of the received data (where sufficient individual level variables were available) including 528 TB patients among 31,721 participants from 15 studies (Extended Data Figure 2). All of these datasets were used for model development and validation, using the internal-external cross-validation framework³⁴, described below. Characteristics of the studies included in prediction

model development and validation were similar to those that were not (Table 1). Our modelling approach used a flexible parametric survival model with two degrees of freedom on a proportional hazards scale, since this showed the best fit in each imputed dataset. From our list of *a priori* variables of interest, we evaluated nine candidate predictors, of which only previous BCG and gender were omitted from the final model. The final prediction model included: age, a composite 'TB exposure' variable (modelled with time-varying covariates to account for non-proportional hazards), time since migration for migrants from countries with high TB incidence, HIV status, solid organ or haematological transplant receipt, normalised LTBI test result and preventative treatment commencement. The final model coefficients and standard errors, pooled across multiply imputed datasets, are summarised in Supplementary Table 6, with visual representations of associations between each variable and incident TB risk shown in Figure 2.

Internal-external cross-validation

Next, we used the internal-external cross-validation (IECV) framework, iteratively discarding one study dataset from the model training set and using this for external validation, to concurrently validate the prediction model, explore between-study heterogeneity, and examine generalizability³⁴. Model discrimination and calibration parameters for 2-year risk of incident TB from the primary validation studies are shown in Figure 3. We assessed discrimination using the C-statistic, which ranged from 0.78 (0.47-1.0) in a study of immunocompromised participants with a small number of incident TB cases²⁹ to 0.97 (0.94-0.99) in a study of TB contacts¹⁸. The random-effects meta-analysis estimate of the C-statistic was 0.88 (0.82-0.93).

Calibration assesses agreement between predicted and observed risk. We assessed calibration visually using grouped calibration plots, supplemented by the calibration-in-the-large (CITL) and slope statistics (Figure 3). Visual calibration plots suggested reasonable calibration in most studies (Extended Data Figure 6). Since incident TB is an infrequent outcome, predictions were appropriately low, with average predicted risk <10% in all quintiles of risk. CITL and calibration slopes of 0 and 1 indicate perfect calibration, respectively. The pooled random-effects meta-analysis CITL estimate was 0.14 (-0.24-0.53), with evidence of systematic under-estimation of risk in one study (CITL 1.02 (0.61-1.43)), and over-estimation in one study (CITL -0.64 (-1.09—0.19)). The pooled random-effects meta-analysis calibration slope estimate was 1.11 (0.83-1.38). Slopes appeared heterogeneous, though visual

assessment of calibration plots suggested that these were prone to being extreme due to the skewed distribution of predicted and observed risk, likely reflecting the relatively rare occurrence of incident TB events.

Distribution of predicted risk and individual predictions

Figure 4 shows the distributions of predicted TB risk among participants who did not commence preventative treatment from the pooled IECV validation sets, stratified by: (a) binary LTBI test result; and (b) indication for screening (among those with a positive test). The median predicted 2-year TB risk was 2.0% (interquartile range 0.8-3.7) and 0.2% (0.1-0.3) among participants with positive and negative binary LTBI test results, respectively. We then examined incident TB risk in four quartiles of predicted risk among untreated participants with positive LTBI tests from the pooled validation sets. Kaplan-Meier plots of the four quartiles showed clear separation of observed risk among these four groups (Figure 4c), with illustrative predicted survival curves for one randomly sampled individual patient per quartile shown in Figure 4d.

Decision curve analysis

Net benefit quantifies the trade-off between correctly identifying true positive patients (progressing to incident TB), and incorrectly detecting false-positives, with weighting of each by the threshold probability^{35,36}. The threshold probability corresponds to a measure of both the perceived risk/benefit ratio of initiating preventative treatment, and the threshold of predicted risk above which treatment is recommended. How patients and clinicians weigh the relative costs of drug-related adverse events (as a result of inappropriate treatment) against the benefits of preventing a case of TB can be subjective. Among untreated participants with LTBI from the pooled validation sets in IECV, net benefit for the prediction model was greater than either treating all LTBI patients, or treating none, throughout a range of threshold probabilities from 0-20% (reflecting a range of clinician and patient preferences) (Figure 5).

Sensitivity analyses

We re-examined population-level TB risk without any exclusion of prevalent TB (cases diagnosed <42 days from testing), resulting in markedly higher cumulative risk for each risk group (Extended Data Figure 3). Recalculation of model predictor parameters revealed similar directions and magnitudes of effect to the primary model when using shorter and longer definitions of prevalent TB (baseline risk was expectedly higher with shorter definitions) and when excluding participants who received preventative

treatment (Supplementary Table 7). Model parameters were noted to be more extreme when using a complete case approach (for variables other than HIV, which was assumed negative when missing). The pooled random-effects meta-analysis C-statistic from IECV when limiting to participants who did not receive preventative treatment was 0.89 (0.82-0.93), similar to the primary analysis (Extended Data Figure 7a). The pooled random-effects meta-analysis C-statistic including only participants with a positive binary LTBI test was 0.77 (0.70-0.83). This finding indicates good discrimination even among participants with a conventional diagnosis of LTBI, albeit lower than discrimination when also including participants with a negative binary LTBI test - likely due to the high negative predictive value of LTBI tests when using standard cut-offs (Extended Data Figure 7b). Finally, in order to assess model performance in situations where the quantitative test results are not available, we imputed an average quantitative positive or negative LTBI test result (based on the medians among the study population), according to the binary result in the validation sets. This analysis provided a pooled random-effects meta-analysis C-statistic of 0.86 (0.76-0.93; Extended Data Figure 7c), and net benefit appeared higher when using this model than either the strategies of treating all patients with evidence of LTBI, or no patients, across the range of threshold probabilities. However, the model using a binary test result had a lower C-statistic, and slightly lower net benefit across most threshold probabilities, compared to the full model using quantitative test results (Extended Data Figure 7d).

Discussion

In this study, we examined population-level incident TB risk in a pooled dataset of >80,000 individuals tested for LTBI in 20 countries with low *M.tuberculosis* transmission (annual incidence $\leq 20/100,000$ persons). We found cumulative 5-year risk of incident TB among people with untreated LTBI approaching 16% among child contacts, and approximately 5% among recent adult contacts, migrants from high TB burden settings, and immunocompromised individuals. A majority of cumulative 5-year risk was accrued during the first year among risk groups with an index exposure, supporting previous data suggesting that risk of progressive TB declines markedly with increasing time since infection¹³. However, we noted substantial variation in incidence rates even within these risk groups, suggesting that an individual level approach to risk stratification is required. Therefore, we developed the first directly data-driven model, to our knowledge, to incorporate the magnitude of the T cell response to *M.tuberculosis* with readily available clinical metadata in order to capture heterogeneity within risk groups, and generate personalized risk predictions for incident TB in settings aiming towards pre-elimination. Clinical co-variables in the final model included age, recent contact (including proximity and infectiousness of the index case), migration from high TB burden countries (and time since arrival), HIV status, solid organ or haematological transplant receipt, and commencement of preventative treatment. The model was externally validated by quantifying the meta-analysis C-statistic for predicting incident disease over 2 years, and by evaluating its calibration, using recommended methods³⁷. Most importantly, the model showed clear clinical utility for informing the decision to initiate preventative treatment, compared to treating all or no patients with LTBI.

The personalized predictions from our model will enable more precise delivery of preventative treatment to those at highest risk of TB disease, while concurrently reducing toxicity and costs related to treatment of people at lower risk. Moreover, the model will allow clinicians and patients to make more informed and individualised choices when considering initiation of preventative treatment. The model also challenges the fundamental notion of an arbitrary binary test threshold for diagnosis of LTBI. By incorporating a quantitative measure of immunosensitization to *M.tuberculosis*, we facilitate a shift from the conventional paradigm of LTBI as a binary diagnosis, towards personalized risk stratification for progressive TB. This approach takes advantage of stronger T cell responses being a correlate of risk, while guarding against a loss of sensitivity by arbitrarily introducing higher test thresholds programmatically¹⁶.

The results of our analyses are consistent with, and extend existing evidence. Recent analyses report similar population-level TB incidence rates among adult contacts¹², with markedly higher risk among young children³⁸. Moreover, these recent meta-analyses confirm highly heterogeneous population-level estimates, thus justifying an individual-level approach to risk estimation^{12,38}. Previous models developed and validated in Peru, a high transmission setting, have generated individual or household-level TB risk estimates for TB contacts^{39–41}. Another model, parameterised using aggregate data estimates from multiple sources, seeks to estimate TB risk following LTBI testing in all settings⁴². However, there are currently no publicly available validation data to support its use and the model omits key predictor variables identified in the current study (including the magnitude of the T cell response and infectiousness of index cases)⁴².

Strengths of the current study include the size of the dataset, curated through comprehensive systematic review in accordance with Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data standards⁴³, and with IPD obtained for 18/26 (69%) eligible studies. This allowed us to examine progression from LTBI to TB disease using the largest adult and pediatric dataset available to date, to our knowledge. We conducted population-level analyses using both 1- and 2-stage IPD-MA approaches in order to present both cumulative TB risk and time-stratified incidence rates, respectively, with consistent results from both. We adhered to transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)⁴⁴ standards, using the recommended approach of IECV³⁷, leading to a fully data-driven and validated model for personalized risk estimates following LTBI testing. The coefficients presented in the model are clinically plausible and have been made publicly available to facilitate further independent external validation. Moreover, the contributing datasets included heterogeneous populations of adults, children, recent TB contacts, migrants from high TB burden countries, and immunocompromised groups from 20 countries across Europe, North America, Asia and Oceania, thus making our results generalizable to settings aiming towards pre-elimination globally.

We also used a comprehensive approach to addressing missing data by using multi-level multiple imputation in the primary analysis, assuming missingness at random and in keeping with recent guidance^{34,45}. This approach facilitated imputation of variables that were systematically missing from some included studies. Previous BCG vaccination and HIV status were noted to be missing from a large

proportion of participants. This missingness may have reduced our power to detect an association between these variables and incident TB, and BCG vaccination was notably not included in the final prognostic model. While increasing data support a role for BCG vaccination in reducing sensitisation to *M. tuberculosis*^{46,47}, additional data are required to further assess the association between BCG vaccination and incident TB risk, after adjustment for other co-variables including quantitative T cell responses. We supported our primary multiple imputation approach using a complete case sensitivity analysis (for variables other than HIV, which was assumed negative when missing). This sensitivity analysis revealed similar findings to the primary analyses, though effect estimates were noted to be more extreme in the complete case approach, likely owing to a degree of bias in the latter, since complete cases analysis assumes no association between the pattern of missingness and the outcome (i.e. incident TB) after adjusting for all other covariates⁴⁸. Given that TB incidence and predictor missingness both varied according to contributing study, this assumption is unlikely to be valid in the current context.

We also used a range of arbitrary definitions of prevalent TB in the primary and sensitivity analyses, since the aim of our prognostic model is to assess the risk of incident TB, after prevalent TB has been clinically ruled out, in order to inform risk/benefit decisions regarding preventative treatment initiation. With increasing recognition of the continuum of *M.tuberculosis* infection using novel diagnostics (including incipient and/or subclinical phases)⁴⁹, the distinction between prevalent and incident disease is becoming increasingly blurred. Future studies could consider integration of our prognostic model with next generation biomarkers, such as blood transcriptional signatures for incipient TB^{50,51}.

A limitation of the study is that its generalisability is restricted to low transmission settings (annual incidence $\leq 20/100,000$ persons). The rationale for limiting to such settings was, firstly, to examine progression from LTBI to TB disease more accurately, by reducing risk of re-infection with *M.tuberculosis* during follow-up. Secondly, the majority of the population in high transmission settings are likely to have a positive LTBI test result, further undermining test specificity for progression to TB disease⁵². Since the quantitative LTBI test result is a strong predictor in our model, a different prediction model may therefore be required in such settings. For example, a recent study developing a prediction model for TB among close contacts in Peru found that the TST result added no value to the model³⁹. Future studies could test our model for use in high transmission settings, updating the parameters as

necessary, in order to extend its application to these settings. A second limitation of the current study is that model calibration was observed to be imperfect during external validation. However, conventional metrics (such as the calibration slope) may not be entirely appropriate in this context, which has a highly skewed distribution of predicted and observed risk, reflecting the rare occurrence of incident TB events. Reassuringly, in decision curve analysis, which accounts for both discrimination and calibration performance in quantifying net benefit, the model showed clinical utility³⁵. Future studies may evaluate the full health economic impact of programmatic implementation of the model.

A further limitation is that, due to a lack of data from contributing studies, other potential predictors that may be associated with incident TB risk (including diabetes, malnutrition, fibrotic chest x-ray lesions and other immunosuppression)⁴ were not evaluated. These unmeasured covariates may have contributed to imperfect discrimination and calibration, along with residual heterogeneity in model performance between datasets. As additional studies are published, the prognostic model can be prospectively evaluated and updated as required. We also note that offer and acceptance of preventative treatment may be more likely among people at higher risk of TB. We therefore accounted for preventative treatment provision in the model by including it as a co-variate along with our other predictors of interest, as widely recommended⁵³. However, residual confounding by indication cannot be excluded in observational studies. In addition, the present model is not applicable for patients commencing biologic agents since no datasets were identified that examined the natural history of LTBI in the context of biologic therapy, in the absence of preventative treatment for TB. A 'hybrid' modelling approach, with mathematical parameterisation of relative risk for any given biologic agent, may be required to extend its application to these therapies. Since the quantitative LTBI test result is a strong predictor in our model, predictions may also be attenuated in the context of advanced immunosuppression⁷. Reassuringly, performance appeared adequate in a dataset of immunocompromised individuals during validation²⁹.

In summary, we present a freely available and directly data-driven personalized risk predictor for incident TB (PERISKOPE-TB; periskope.org). This tool will allow a programmatic paradigm shift for TB prevention services in settings aiming towards pre-elimination globally, by facilitating shared decision-making between clinicians and patients for preventative treatment initiation.

Acknowledgements

The study was funded by National Institute for Health Research (DRF-2018-11-ST2-004 to R.K.G.; SRF-2011-04-001 and NF-SI-0616-10037 to I.A.), the Wellcome Trust (207511/Z/17/Z to M.N.) and by NIHR Biomedical Research Funding to and University College London Hospitals. C.L. is funded by the German Center for Infection Research (DZIF). J.S.D. receives salary support from the National Health and Medical Research Council (Australia). This paper presents independent research supported by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The study funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript. The authors would like to thank all of the research teams involved in the primary studies that contributed data towards this analysis.

Author contributions

R.K.G. and I.A. conceived the study and led pooling of data. R.K.G., M.X.R., A.C, M.L., M.N. and I.A. wrote the study protocol and developed the analysis plan. R.K.G. conducted the analyses and wrote the first draft of the manuscript. R.K.G., C.J.C. and M.K. performed the systematic literature review. M.Q. and A.C. provided statistical and multiple imputation expertise. A.Y. and R.K.G. developed the website interface for the risk predictor tool. M.C.A., N.A., R.D., C.C.D., J.D., J.S.D., C.E., S.G., P.H., A.M.H., T.H., J.C.J., C.L., B.L., F.v.L., L.M., C.R., K.R., D.R., M.S., R.S., G.S., G.W., T.Y., J.-P.Z. and D.Z. contributed primary data and assisted interpretation. R.W.A contributed to data interpretation. All authors critically reviewed and approved the manuscript prior to submission.

Competing interests

J.S.D.'s institution receives investigator-initiated research grants and consultancy income from Gilead Sciences, AbbVie, Bristol Myers Squibb and Merck. The Burnet Institute receives funding from the Victorian Government Operational Infrastructure Fund. C.L. reports honoraria from Chiesi, Gilead, Insmad, Janssen, Lucane, Novartis, Oxoid, Berlin Chemie (for participation at sponsored symposia) and from Oxford Immunotec (to attend a scientific advisory board meeting), all outside the submitted work. M.S. reports receipt of test kits free of charge from Qiagen and from Oxford Immunotec for investigator-initiated research projects. I.A. reports receiving free test kits from Qiagen for an

investigator-initiated research project²⁵. C.E. reports receiving free test kits from Qiagen for investigator-initiated research projects, outside the submitted work. The authors declare no other conflict of interest.

References

1. World Health Organization. WHO | Global tuberculosis report 2019. (2019). Available at: https://www.who.int/tb/publications/global_report/en/.
2. World Health Organization. The End TB Strategy. (2015). Available at: http://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1. (Accessed: 1st October 2017)
3. Getahun, H. *et al.* Management of latent Mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries. *Eur. Respir. J.* **46**, 1563–1576 (2015).
4. Getahun, H., Matteelli, A., Chaisson, R. E. & Raviglione, M. Latent Mycobacterium tuberculosis Infection. *N. Engl. J. Med.* **372**, 2127–2135 (2015).
5. Mack, U. *et al.* LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. *Eur. Respir. J.* **33**, 956–973 (2009).
6. Sterling, T. R. *et al.* Guidelines for the Treatment of Latent Tuberculosis Infection: Recommendations from the National Tuberculosis Controllers Association and CDC, 2020. *MMWR. Recomm. Reports* **69**, 1–11 (2020).
7. Pai, M. *et al.* Gamma Interferon Release Assays for Detection of Mycobacterium tuberculosis Infection. *Clin. Microbiol. Rev.* **27**, 3–20 (2014).
8. Rangaka, M. X. *et al.* Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet. Infect. Dis.* **12**, 45–55 (2012).
9. Abubakar, I. *et al.* Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet. Infect. Dis.* **18**, 1077–1087 (2018).
10. Gao, J. *et al.* Knowledge and Perceptions of Latent Tuberculosis Infection among Chinese Immigrants in a Canadian Urban Centre. *Int. J. Family Med.* **2015**, 546042 (2015).
11. Alsdurf, H., Hill, P. C., Matteelli, A., Getahun, H. & Menzies, D. The cascade of care in diagnosis and treatment of latent tuberculosis infection: a systematic review and meta-analysis. *Lancet*

- Infect. Dis.* **16**, 1269–1278 (2016).
12. Campbell, J. R., Winters, N. & Menzies, D. Absolute risk of tuberculosis among untreated populations with a positive tuberculin skin test or interferon-gamma release assay result: systematic review and meta-analysis. *BMJ* **368**, (2020).
 13. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Revisiting the timetable of tuberculosis. *BMJ* **362**, k2738 (2018).
 14. Winje, B. A. *et al.* Stratification by interferon- γ release assay level predicts risk of incident TB. *Thorax* **73**, 652–661 (2018).
 15. Andrews, J. R. *et al.* Serial QuantiFERON testing and tuberculosis disease risk among young children: an observational cohort study. *Lancet Respir. Med.* **5**, 282–290 (2017).
 16. Gupta, R. K. *et al.* Quantitative Interferon Gamma Release Assay and Tuberculin Skin Test Results to Predict Incident Tuberculosis: A Prospective Cohort Study. *Am. J. Respir. Crit. Care Med.* rccm.201905-0969OC (2019). doi:10.1164/rccm.201905-0969OC
 17. Altet, N. *et al.* Predicting the Development of Tuberculosis with the Tuberculin Skin Test and QuantiFERON Testing. *Ann. Am. Thorac. Soc.* **12**, 680–688 (2015).
 18. Diel, R., Loddenkemper, R., Niemann, S., Meywald-Walter, K. & Nienhaus, A. Negative and positive predictive value of a whole-blood interferon-gamma release assay for developing active tuberculosis: an update. *Am. J. Respir. Crit. Care Med.* **183**, 88–95 (2011).
 19. Dobler, C. C. & Marks, G. B. Risk of tuberculosis among contacts in a low-incidence setting. *Eur. Respir. J.* **41**, 1459–1461 (2013).
 20. Geis, S. *et al.* How can we achieve better prevention of progression to tuberculosis among contacts? *Eur. Respir. J.* **42**, 1743–1746 (2013).
 21. Haldar, P. *et al.* Single-step QuantiFERON screening of adult contacts: a prospective cohort study of tuberculosis risk. *Thorax* **68**, 240–246 (2013).
 22. Sloot, R., Van Der Loeff, M. F. S., Kouw, P. M. & Borgdorff, M. W. Risk of tuberculosis after

- recent exposure: A 10-year follow-up study of contacts in Amsterdam. *Am. J. Respir. Crit. Care Med.* **190**, 1044–1052 (2014).
23. Yoshiyama, T., Harada, N., Higuchi, K., Saitou, M. & Kato, S. Use of the QuantiFERON-TB Gold in Tube test for screening TB contacts and predictive value for active TB. *Infect. Dis. (London, England)* **47**, 542–549 (2015).
 24. Zellweger, J.-P. *et al.* Risk Assessment of Tuberculosis in Contacts by IFN- γ Release Assays. A Tuberculosis Network European Trials Group Study. *Am. J. Respir. Crit. Care Med.* **191**, 1176–1184 (2015).
 25. Gupta, R. K. *et al.* Evaluation of QuantiFERON-TB Gold Plus for Predicting Incident Tuberculosis among Recent Contacts: A Prospective Cohort Study. *Ann. Am. Thorac. Soc.* AnnalsATS.201905-407RL (2020). doi:10.1513/AnnalsATS.201905-407RL
 26. Aichelburg, M. C. *et al.* Detection and prediction of active tuberculosis disease by a whole-blood interferon-gamma release assay in HIV-1-infected individuals. *Clin. Infect. Dis.* **48**, 954–962 (2009).
 27. Doyle, J. S. *et al.* Latent Tuberculosis screening using interferon-gamma release assays in an Australian HIV-infected cohort: is routine testing worthwhile?. *J. Acquir. Immune Defic. Syndr.* **66**, 48–54 (2014).
 28. Lange, B., Vavra, M., Kern, W. V & Wagner, D. Development of tuberculosis in immunocompromised patients with a positive tuberculosis-specific IGRA. *Int. J. Tuberc. Lung Dis.* **16**, 492–495 (2012).
 29. Sester, M. *et al.* Risk assessment of tuberculosis in immunocompromised patients. A TBNET study. *Am. J. Respir. Crit. Care Med.* **190**, 1168–1176 (2014).
 30. Munoz, L. *et al.* Immunodiagnostic Tests' Predictive Values for Progression to Tuberculosis in Transplant Recipients: A Prospective Cohort Study. *Transplant. direct* **1**, e12 (2015).
 31. Roth, D. Z. *et al.* Impact of interferon-gamma release assay on the latent tuberculosis cascade of care: A population-based study. *Eur. Respir. J.* **49**, 1601546 (2017).

32. Erkens, C. G. M. *et al.* Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. *Eur. Respir. J.* **48**, 1420–1428 (2016).
33. Zenner, D., Loutet, M. G., Harris, R., Wilson, S. & Ormerod, L. P. Evaluating 17 years of latent tuberculosis infection screening in north-west England: a retrospective cohort study of reactivation. *Eur. Respir. J.* **50**, (2017).
34. Debray, T. P. A. *et al.* Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLOS Med.* **12**, e1001886 (2015).
35. Vickers, A. J., van Calster, B. & Steyerberg, E. W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic Progn. Res.* **3**, 18 (2019).
36. Zhang, Z. *et al.* Decision curve analysis: a technical note. *Ann. Transl. Med.* **6**, 19 (2018).
37. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal–external, and external validation. *J. Clin. Epidemiol.* **69**, 245–247 (2016).
38. Martinez, L. *et al.* The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis. *Lancet* **395**, 973–984 (2020).
39. Saunders, M. J. *et al.* A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. *Lancet Infect. Dis.* **17**, 1190–1199 (2017).
40. Saunders, M. J. *et al.* A household-level score to predict the risk of tuberculosis among contacts of patients with tuberculosis: a derivation and external validation prospective cohort study. *Lancet Infect. Dis.* **20**, 110–122 (2020).
41. Li, R. *et al.* Two Clinical Prediction Tools to Improve Tuberculosis Contact Investigation. *Clin. Infect. Dis.* (2020). doi:10.1093/cid/ciz1221
42. Menzies, D., Gardiner, G., Farhat, M., Greenaway, C. & Pai, M. Thinking in three dimensions: a web-based algorithm to aid the interpretation of tuberculin skin test results. *Int. J. Tuberc. Lung Dis.* **12**, 498–505 (2008).

43. Stewart, L. A. *et al.* Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data. *JAMA* **313**, 1657 (2015).
44. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, (2015).
45. Audigier, V. *et al.* Multiple imputation for multilevel data with continuous and binary variables. *Stat. Sci.* **33**, 160–183 (2018).
46. Nemes, E. *et al.* Prevention of M. Tuberculosis infection with H4:IC31 vaccine or BCG revaccination. *N. Engl. J. Med.* **379**, 138–149 (2018).
47. Katelaris, A. L. *et al.* Effectiveness of BCG Vaccination Against Mycobacterium tuberculosis Infection in Adults: A Cross-sectional Analysis of a UK-Based Cohort. *J. Infect. Dis.* **221**, 146–155 (2020).
48. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **30**, 377–399 (2011).
49. Drain, P. K. *et al.* Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. *Clin. Microbiol. Rev.* **31**, e00021-18 (2018).
50. Gupta, R. K. *et al.* Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet. Respir. Med.* **0**, (2020).
51. Roe, J. *et al.* Blood transcriptomic stratification of short-term risk in contacts of tuberculosis. *Clin. Infect. Dis.* (2019). doi:10.1093/cid/ciz252
52. World Health Organization. WHO | Latent TB Infection : Updated and consolidated guidelines for programmatic management. *WHO* (2018). Available at: <http://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/>. (Accessed: 15th August 2018)
53. Groenwold, R. H. H. *et al.* Explicit inclusion of treatment in prognostic modeling was

recommended in observational and randomized settings. *J. Clin. Epidemiol.* **78**, 90–100 (2016).

Figure Legends

Figure 1: Population-level cumulative risk of incident tuberculosis during follow-up.

Risk is stratified by binary latent TB test result, provision of preventative treatment, and indication for screening among participants with untreated latent infection (total n=80,468 participants). Cumulative risk is estimated using flexible parametric survival models with random effects intercepts by source study, separately fitted to each risk group. Prevalent TB cases (diagnosed within 42 days of recruitment) are excluded. Each plot is presented as point estimates (solid line) and 95% confidence intervals (shaded area). PT = preventative treatment. Numbers of participants, TB cases and numeric cumulative risk estimates for each plot are presented in Supplementary Table 5. Cumulative TB risk including prevalent TB cases is presented in Extended Data Figure 3.

Figure 2: Visual representations of associations between predictors and incident tuberculosis.

Illustrative estimates are shown for a 33-year old migrant from a high TB burden setting. The example 'base case' patient does not commence preventative treatment, is not living with HIV, has not received a previous transplant, and has an 'average' positive latent TB test. We vary one of these predictors in each plot ((a) age; (b) normalised latent TB test result; (c) years since migration; (d) exposure to *M. tuberculosis*; (e) HIV status; (f) transplant receipt; and (g) preventative treatment). Each plot is presented as point estimates (solid line) and 95% confidence intervals (shaded area). The model was trained on a pooled dataset (n=31,090 participants). Model parameters are provided in Supplementary Table 6. 'Household smear+ contact' = household contact of sputum smear-positive index case; 'Other contact' = contact of non-household or smear-negative index case; 'Migrant' = migrant from high TB incidence country, without recent contact.

Figure 3: Forest plots showing model discrimination and calibration metrics for predicting 2-year risk of incident tuberculosis.

Discrimination is presented as the C-statistic; calibration is presented as calibration-in-the-large (CITL) and the calibration slope. Data from nine primary validation studies are shown, from internal-external cross-validation of the model (developed among n=31,090 participants; validated among 25,504 in this analysis). 'TB' column indicates number of incident TB cases within 2 years of study entry and 'N' indicates total participants per study included in analysis. Each forest plot shows point estimates (squares) and 95% confidence intervals (error bars). Pooled estimates are shown as diamonds. Calibration slopes >1 suggest under-fitting (predictions are not varied enough), while slopes <1 indicate

over-fitting (predictions are too extreme). Calibration-in-the-large indicates whether predictions are systematically too low (CITL >0) or too high (CITL <0). Dashed lines indicate line of no discrimination (C-statistic), and perfect calibration (CITL and slope), respectively.

Figure 4: Distribution of predictions and risk of incident tuberculosis in four quartiles of risk for people with positive latent TB tests.

Distribution of risk from prediction model using pooled validation sets of people not receiving preventative therapy from internal-external cross-validation of the model (n=27,511 participants), stratified by (a) binary latent TB test result and (b) indication for screening among untreated people with positive LTBI tests. (c) Kaplan-Meier plots for quartile risk groups (1=lowest risk) of untreated individuals with positive LTBI tests (n=6,418 participants). Quartiles represent four equally sized groups based on predicted risk of incident TB, from the pooled validation sets derived from internal-external cross-validation of the prediction model. P value represents Log-rank test ($p=1.137 \times 10^{-40}$). (d) Randomly sampled individual patients from each risk quartile. Patient 1 is a 22-year-old with no TB exposure and a normalised latent TB test result on the 68th percentile; Patient 2 is a 41-year-old migrant from a high TB burden country (3.8 years since migration) with normalised latent TB test result on the 80th percentile; Patient 3 is a 51-year-old household contact of a smear positive index TB case with a normalised latent TB test result on the 79th percentile; Patient 4 is a 33-year-old household contact of a smear positive index TB case with a normalised latent TB test result on the 94th percentile. All four example patients are HIV negative and are not transplant recipients. Equivalent values of normalised percentile test results for QuantiFERON, T-SPOT.TB and tuberculin skin test are shown in Supplementary Table 10. Plots (c) and (d) are presented as point estimates (solid line) and 95% confidence intervals (shaded area).

Figure 5: Decision curve analysis.

Shown as net benefit of the prediction model among untreated participants from the pooled validation sets with positive binary latent TB tests (n=6,418 participants), compared to 'treat all' and 'treat none' strategies across a range of threshold probabilities (x-axis). Net benefit quantifies the trade-off between correctly identifying true positive progressors to incident TB, and incorrectly detecting false-positives, with weighting of each by the threshold probability³⁵. The threshold probability corresponds to a measure of both the perceived risk/benefit ratio of initiating preventative treatment, and the percentage cut-off for the prediction model, above which treatment is recommended. Net benefit appeared higher

than either the strategies of treating all patients with evidence of LTBI, or no patients, throughout the range of threshold probabilities, suggesting clinical utility. For illustration, a patient who is very concerned about developing TB disease, but not concerned regarding side-effects of preventative treatment, may have a low threshold probability (e.g. 1%, which is equivalent to a risk:benefit ratio of 1:99, i.e. the outcome of developing TB is considered to be 99 times worse than taking unnecessary preventative treatment). In contrast, a patient who is less concerned about developing TB but is very concerned about side-effects of preventative treatment may have a higher threshold probability (e.g. 10% which is equivalent to a risk:benefit ratio of 1:9). The unit of net benefit is 'true positives'³⁵. For instance, a net benefit of 0.01 would be equivalent to a strategy where 1 patient per 100 tested was appropriately given preventative treatment, as they would otherwise have progressed to incident TB if left untreated.

Tables

Table 1: Characteristics of contributing studies included in individual participant data meta-analysis.

Additional study characteristics are shown in Supplementary Table 1. IQR = interquartile range.

Authors	Publication Year	Country	N (total)	Adults / children	Population	Follow-up years (median (IQR))	TB cases	Loss to follow-up	Included in prediction modelling	NOS [^]
Abubakar et al ⁹	2018	UK	10,045	Adults	Contacts & migrants	4.7 (3.7-5.5)	147	10 (0.1%)	Yes	7/7
Aichelburg et al ²⁶	2009	Austria	830	Adults	People with HIV	1.2 (0.7-1.4)	11	25 (3%)	Yes	7/7
Altet et al ¹⁷	2015	Spain	1,339	Adults & children	Contacts	4 (4-4)	95	0 (0%)	Yes	7/7
Diel et al ¹⁸	2011	Germany	1,414	Adults & children	Contacts	3.5 (2.5-4.2)	19	381 (26.9%)	Yes	7/7
Dobler & Marks ¹⁹	2013	Australia	12,212	Adults & children	Contacts	4.2 (2-6.9)	94	351 (2.9%)	No*	7/7
Doyle et al ²⁷	2014	Australia	919	Adults	People with HIV	2.9 (1.7-3.6)	2	47 (5.1%)	Yes	7/7
Erkens et al ³²	2016	Netherlands	14,241	Adults & children	Mixed population screening	5.5 (3-7.4)	134	NA	No*	6/6
Geis et al ²⁰	2013	Germany	1,283	Adults & children	Contacts	0.8 (0.4-1.1)	33	62 (4.8%)	Yes	6/6
Gupta et al ²⁵	2020	UK	623	Adults	Contacts	1.9 (1.6-2.2)	13	0 (0%)	Yes	7/7
Haldar et al ²¹	2013	UK	1,411	Adults & children	Contacts	1.9 (1.3-2.4)	37	30 (2.1%)	Yes	7/7
Lange et al ²⁸	2012	Germany	456	Adults	Immunocompromised	2.8 (2-3.1)	1	42 (9.2%)	Yes	7/7
Munoz et al ³⁰	2015	Spain	76	Adults	Transplant recipients	4.3 (3.6-4.8)	2	0 (0%)	Yes	7/7
Roth et al ³¹	2017	Canada	22,949	Adults & children	Mixed population screening	3 (1.8-4.3)	58	NA	Subset*	6/6
Sester et al ²⁹	2014	Multiple European countries	1,464	Adults	Immunocompromised	2.7 (1.5-3.5)	11	7 (0.5%)	Yes	7/7
Sloot et al ²²	2014	Netherlands	5,895	Adults & children	Contacts	5.9 (3.6-7.7)	81	NA	Yes	7/7
Yoshiyama et al ²³	2015	Japan	625	Adults & children	Contacts	1.8 (1.4-2)	12	0 (0%)	Yes	6/7
Zellweger et al ²⁴	2015	Multiple European countries	5,237	Adults & children	Contacts	2.6 (1.9-3.5)	55	1339 (25.6%)	Yes	7/7
Zenner et al ³³	2017	UK	1,341	Adults	Migrants	3.7 (3-4.8)	21	NA	No*	7/7
Total			82,360			3.7 (2.1-5.3)	826	2294 (2.8%)		

*Not included in prediction modelling due to lack of data on proximity or infectiousness of index cases¹⁹, or absent quantitative LTBI test data^{32,33}. A subset of the dataset was included in the prediction model for the Roth et al study³¹; contacts and migrants were excluded due to no data being available on country of birth or infectiousness of index cases, respectively.

[^]Modified version of the Newcastle Ottawa Scale for cohort studies.

Methods

Systematic review and pooling of individual participant data

We conducted a systematic review and IPD-MA, in accordance with Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data standards⁴³, to investigate the risk of progression to TB disease among people tested for LTBI in low transmission settings. The study is registered with PROSPERO (CRD42018115357). We searched Medline and Embase for studies published 01/01/2002 – 31/12/2018 using comprehensive MeSH and keyword terms for 'TB', 'IGRA', 'TST', 'latent TB', and 'predictive value', without language restrictions. Longitudinal studies that primarily aimed to assess the risk of progression to TB disease among individuals tested for LTBI and that were conducted in a low TB transmission setting (defined as annual incidence $\leq 20/100,000$ persons at the midpoint of the study) were eligible for inclusion. The full search strategy and eligibility criteria are provided in Supplementary Tables 8 and 9. Titles and abstracts underwent a first screen; relevant articles were selected for the second screen, which included full text review. Both first and second screens were performed by two independent reviewers, with disagreements resolved through discussion and arbitration by a third reviewer where required. Corresponding authors of eligible studies were invited to contribute IPD. Received data were mapped to a master variables list, and the integrity of the IPD were examined by comparing original reported results with re-analysed results using contributed data. Quality assessment was performed using a modified version of the Newcastle-Ottawa scale for cohort studies⁵⁴.

Definitions

Participants entered the cohort on the day of LTBI screening or diagnosis, and exited on the earliest of censor date (last date of follow-up), active TB diagnosis date, date of death, or date of loss to follow-up (where available). LTBI was defined as any positive LTBI test (TST or commercial IGRA), using TST thresholds as defined by the contributing study (a 10mm cut-off was used for studies that assessed multiple thresholds). Quantitative IGRA thresholds were calculated according to standard manufacturer guidelines. IGRAs included three generations of QuantiFERON TB assays (QuantiFERON Gold-In-Tube, QuantiFERON Gold, QuantiFERON-TB Gold Plus; Qiagen, Hilden, Germany), which were assumed to be equivalent²⁵, and T-SPOT.TB (Oxford Immunotec, UK). Microbiologically confirmed and/or clinically diagnosed TB cases were included, as per contributing study definitions. In the absence of a widely accepted temporal distinction between prevalent and incident disease, prevalent TB at the

time of screening was arbitrarily defined as a TB diagnosis within 42 days of enrolment; these cases were omitted from the primary analysis. Alternative shorter and longer temporal definitions were tested as sensitivity analyses. Participants with missing outcomes or durations of follow-up were considered lost to follow-up. 'Preventative treatment' was defined as any LTBI treatment regimen recommended by the WHO⁵². All contributing studies included regimens consistent with this guidance; the effectiveness of each regimen was assumed to be equivalent⁵⁵.

Population-level analysis

Survival analysis

In a one-stage IPD-MA approach, we used flexible parametric survival models, with a random effect intercept by source study to account for between study heterogeneity, to examine population level risk of incident TB, stratified by LTBI screening result (positive vs negative) and provision of LTBI treatment (commenced vs. not commenced). We further examined progression risk among untreated participants with LTBI, stratified by indication for screening (recent child contacts (<15 years) vs adult contacts vs migrants vs immunocompromised), by separately fitting random-effect flexible parametric survival models to each risk group. Child contacts were further stratified by age (<5 vs. 5 to 14 years).

Incidence rates

We also calculated TB incidence rates (per 1,000 person-years) in a two-stage IPD-MA approach, stratified by LTBI screening result, provision of LTBI treatment, and indication for screening. Rates were calculated separately for the 0-2 year and 2-5 year follow-up intervals. Pooled incidence rate estimates for each risk group and follow-up interval were derived using random intercept Poisson regression models, without continuity correction for studies with zero events, in the meta package in R⁵⁶.

Prediction model analysis

Variables of interest

We then developed and validated a personalized prediction model for incident TB, in accordance with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidance⁴⁴. For this analysis, we included studies that reported quantitative LTBI test results, proximity and infectiousness (based on sputum smear status) of index cases for contacts, and country of birth and time since entry for migrants, since we considered these variables fundamental *a priori*. Using this subset of the data, we examined the availability of a range of variables of interest, specified *a priori*, in

the contributing datasets to determine eligibility for inclusion as candidate predictors in the model. We determined that the following predictors were available from a sufficient number of datasets for further evaluation: age, gender, quantitative LTBI test result, previous BCG vaccination, recent contact (including proximity and infectiousness of index case), migration from a high TB incidence setting, time since migration, solid organ or haematological transplant receipt, HIV status and TB preventative treatment commencement.

Variable transformations

Previous data have shown that quantitative TST, QuantiFERON Gold-in-tube (QFT-GIT) and T-SPOT.TB results are associated with risk of incident TB¹⁶. However, each LTBI test is reported using different scales, and it has hitherto been unclear whether quantitative values of each test are equivalent with respect to incident TB risk. To assess this further, we examined a sub-population of the entire cohort where all three tests were performed among the same participants in head-to-head studies. We normalised quantitative results for the TST, QFT-GIT and T-SPOT.TB to a percentile scale using this head-to-head population, and examined the association between normalised result and risk of incident TB using Cox proportional hazards models with restricted cubic splines. Since TST cut-offs are frequently stratified by BCG vaccination and HIV status^{57,58}, we also examined whether these variables modified the association between quantitative TST measurement and incident TB risk in the head-to-head subpopulation. Since there was no evidence that including interaction terms for either BCG or HIV improved model fit (based on Akaike Information Criteria (AIC)), we used unadjusted TST measurements. This analysis revealed that the normalised percentile results for each test (unadjusted TST, QFT-GIT and T-SPOT.TB) appeared to be associated with similar risk of incident TB (Extended Data Figure 8). The LTBI tests implemented differed between contributing studies. From this point, all LTBI test results were therefore normalised to this percentile scale to enable data harmonisation across studies, by transforming raw quantitative results to the relevant percentile using look-up tables derived from the head-to-head population (Supplementary Table 10). Since most people evaluated for LTBI under routine programmatic conditions have a single test performed, we only included one test result per participant in the prediction model. We preferentially included tests where quantitative results were available. Where quantitative results were available for more than one test, we preferentially included the QuantiFERON result (since this was the most commonly used test in the dataset), followed by T-SPOT.TB, and then the TST.

Recent contacts were categorised as either 'smear positive and household' or 'other' contacts, since there was no evidence of separation of risk among additional subgroups of the 'other' contacts stratum during exploratory univariable analyses (Extended Data Figure 8). Since we considered migration from a high TB burden country (defined as annual TB incidence $\geq 100/100,000$ persons at the year of migration) to be a proxy for prior TB exposure, we included this in a composite 'TB exposure' variable, which included four mutually-exclusive levels: household contact of smear-positive index case; 'other' contact; migrant from country with high TB incidence, without recent contact; and no exposure. There was no evidence of separation of incident TB risk when stratified by TB incidence in country of birth above the binary country of birth threshold (TB incidence $\geq 100/100,000$ persons) among migrants, or when stratified by country of birth among recent contacts (Extended Data Figure 8).

Age and normalised test result variables were modelled using restricted cubic splines (using a default of 5 knots placed at recommended intervals⁵⁹) to account for their non-linear associations with incident TB.

Multiple imputation

A data dictionary and summary of missingness of candidate predictor variables is provided in Supplementary Table 11. We performed multi-level multiple imputation to account for sporadically and systematically missing data (assuming missingness at random⁴⁸), while respecting clustering by source study, in accordance with recent guidance⁴⁵ using the *micemd* package in R⁶⁰. We used predictive mean matching for continuous variables, due to their skewed distributions. We included all variables (including transformations) assessed in the downstream prediction model in the imputation model, along with auxiliary variables, to ensure congeniality. Multi-level imputation was done separately for contacts and non-contacts due to expected heterogeneity between these groups. We generated ten multiply imputed datasets, with 25 between-imputation iterations. Model convergence was assessed by visually examining plots of imputed parameters against iteration number. All downstream analyses were done in each of the ten imputed datasets; model coefficients and standard errors were combined using Rubin's rules⁶¹. No imputation was done for participants missing binary LTBI test results, or for those lost to follow-up; these individuals were excluded. For recent TB contacts or people screened due to HIV infection with missing data on transplant status, this was assumed negative due to the very low prevalence of transplant receipt when observed among these risk groups (<0.5%).

Variable selection and final model development

We performed backward selection of the nine candidate predictors in each of the pooled imputed datasets, using AIC. Variables that were selected in more than 50% of the imputed datasets were included in the final model. T cell responses to *M.tuberculosis* may be impaired in the context of immunosuppression (including among people with HIV or transplant recipients)⁷. We therefore also tested whether there was a significant interaction between HIV or transplant and the normalized percentile test result variable, in order to assess whether the association between the quantitative test result and incident TB risk varied according to HIV or transplant status. This analysis showed no evidence of effect modification, based on AIC, thus these interaction terms were not included in the final model.

We used flexible parametric survival models in order to facilitate estimation of baseline risk throughout the duration of follow-up⁶², using the *rstpm2* package⁶³. We examined a range of degrees of freedom for the baseline hazard, using proportional hazards and odds scales, and selected the final model parameters based on the lowest AIC across the imputed datasets. Visual inspection of survival curves suggested non-proportional hazards for the composite exposure category; we therefore assessed whether including this variable as a time-varying covariate (by including an interaction between the composite exposure covariate of interest and time) improved model fit⁶⁴. Since the AIC for the time-varying covariate model was lower across all imputed datasets, this time-varying covariate approach was used for the final model.

Internal-external cross-validation

Following development of the final model, we used the internal-external cross-validation (IECV) framework for model validation, allowing concurrent assessment of between-study heterogeneity and generalisability³⁴. In this process, one entire contributing study dataset is iteratively discarded from the model training set and used for external validation. This process is repeated until each dataset has been used once for validation. The primary outcome for validation was 2-year risk of incident TB. We included datasets with a minimum of 5 incident TB cases, and where participants had been included regardless of LTBI test result, as the primary validation sets. We assessed model discrimination using the C-statistic for 2-year TB risk. Model calibration was assessed by visually examining calibration plots of predicted risk vs. Kaplan Meier estimated observed 2-year risk in quintiles, and using the calibration

slope and calibration-in-the-large (CITL) statistics⁶⁵. Calibration slopes >1 suggest under-fitting (predictions are not varied enough), while slopes <1 indicate over-fitting (predictions are too extreme). Slopes were calculated by fitting survival models with the model linear predictor as the sole predictor; the calculated coefficient for the linear predictor provides the calibration slope. CITL indicates whether predictions are systematically too low (CITL >0) or too high (CITL <0). We calculated CITL for each validation set by fixing all model coefficients from model development (including the baseline hazard terms), and re-estimating the intercept. The difference between the development model and recalculated validation model intercepts provided the CITL statistic⁶⁶.

Pooling of IECV parameters and random-effects meta-analysis

IECV was performed on each imputed dataset. Validation set C-statistics, calibration slopes and CITL metrics were pooled for each study across imputations using Rubin's rules⁶¹. We then meta-analysed these metrics across validation studies with random-effects, using logit-transformed C-statistics as previously recommended⁶⁷, to derive pooled discrimination and calibration estimates. The IECV validation sets were also pooled, with averaging of the predicted 2-year risk of TB for each individual in the validation sets across imputations, for downstream decision curve analyses as described below.

Decision curve analysis

Decision curve analysis complements model validation parameters by assessing the potential clinical utility of a prediction model^{35,36}. Net benefit quantifies the proportion of true positive cases detected minus the proportion of false positives, with weighting of each by the 'threshold probability'³⁵. The 'threshold probability' reflects both the risk/benefit ratio of initiating preventative treatment, and the percentage cut point for the prediction model, above which treatment is recommended. We calculated net benefit across a range of clinically relevant threshold probabilities (to account for a range of clinician and patient preferences) in comparison to the default strategies of treating either all or no patients with a positive LTBI test. We analysed net benefit using the `stdca` command from the `ddsjoberg/dca` package in R⁶⁸, using the stacked validation sets of untreated participants with positive LTBI tests from IECV (to ensure that each individual for whom a prediction was generated had not been included in the model training set used to derive that prediction).

Sensitivity analyses

First, we re-examined population-level TB risk without exclusion of prevalent TB cases. Second, we recalculated prediction model parameters using: alternative definitions of prevalent TB (ranging from diagnosis within 0 to 180 days of recruitment); a complete case approach (for all variables except for HIV status, which was assuming negative where this was missing); and exclusion of participants who received preventative treatment. Parameters for each of these models were compared with the primary model (without time-varying covariates to facilitate interpretation).

We also examined IECV discrimination parameters for validation datasets when: (a) restricted to participants with positive binary LTBI tests; (b) excluding those who received preventative treatment; and (c) imputing an average quantitative positive or negative LTBI test result (based on the medians among the study population), according on the binary result. The latter analysis was done to assess model performance in situations where the quantitative test result is not available.

Ethics

This study involved analyses of fully depersonalized data from previously published cohort studies, with data pooling via a safe haven. Ethical approvals for sharing of data were sought and obtained by contributors of individual participant data, where required.

Data availability statement

The individual participant data pooled for this analysis are subject to data sharing agreements with the original study authors. The data may be shared to interested parties by the corresponding authors of the original studies, subject to data sharing agreements.

Code availability

The final prognostic model developed in this study has been made freely available, to enable immediate implementation in clinical practice and independent external validation in new datasets (periskope.org).

The code underlying the prediction tool is available at github.com/rishi-k-gupta/PERISKOPE-TB.

Methods-only References

54. Wells, G. *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. (Accessed: 3rd June 2019)
55. Zenner, D. *et al.* Treatment of Latent Tuberculosis Infection. *Ann. Intern. Med.* **167**, 248 (2017).
56. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Heal.* **22**, 153–160 (2019).
57. Seddon, J. A. *et al.* The impact of BCG vaccination on tuberculin skin test responses in children is age dependent: evidence to be considered when screening children for tuberculosis infection. *Thorax* **71**, 932–939 (2016).
58. Cobelens, F. G. *et al.* Tuberculin Skin Testing in Patients with HIV Infection: Limited Benefit of Reduced Cutoff Values. *Clin. Infect. Dis.* **43**, 634–639 (2006).
59. Frank E Harrell. Biostatistical Modeling. (2004). Available at: <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/BioMod/notes.pdf>. (Accessed: 7th February 2020)
60. Audigier, V. & Resche-Rigon, M. micemd: Multiple Imputation by Chained Equations with Multilevel Data. (2019). R package version 1.6.0.
61. Rubin, D. B. *Multiple imputation for nonresponse in surveys*. (Wiley-Interscience, 2004).
62. Royston, P. & Parmar, M. K. B. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* **21**, 2175–2197 (2002).
63. Clements, M. & Liu, X.-R. rstpm2: Smooth Survival Models, Including Generalized Survival Models. (2019). R package version 1.5.1.
64. Bower, H. *et al.* Capturing simple and complex time-dependent effects using flexible parametric survival models: A simulation study. *Commun. Stat. - Simul. Comput.* 1–17 (2019).

doi:10.1080/03610918.2019.1634201

65. Steyerberg, E. W. & Vergouwe, Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur. Heart J.* **35**, 1925–31 (2014).
66. Westeneng, H.-J. *et al.* Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol.* **17**, 423–433 (2018).
67. Snell, K. I., Ensor, J., Debray, T. P., Moons, K. G. & Riley, R. D. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C -statistic and calibration measures? *Stat. Methods Med. Res.* **27**, 3505–3522 (2018).
68. Sjoberg, D. D. dca: Decision Curve Analysis. (2020). R package version 0.1.0.9000.