

Published in final edited form as:

Nat Med. 2020 December 01; 26(12): 1941–1949. doi:10.1038/s41591-020-1076-0.

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

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 Sloom³¹, 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,∞}

¹Institute for Global Health, University College London, London, UK

²Institute of Health Informatics, University College London, London, UK

³MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

⁴Department of Dermatology, Sozialmedizinisches Zentrum Ost-Donauspital, Vienna, Austria

[∞] i.abubakar@ucl.ac.uk .

Author contributions

R.K.G. and I.A. conceived of the study and led the 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 with interpretation. R.W.A. contributed to data interpretation. All authors critically reviewed and approved the manuscript before 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, Inmed, Janssen, Lucane, Novartis, Oxoid, Berlin Chemie (for participation at sponsored symposia) and Oxford Immunotec (to attend a scientific advisory board meeting), all outside of 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 test kits free of charge from Qiagen for an investigator-initiated research project²⁵. C.E. reports receiving test kits free of charge from Qiagen for investigator-initiated research projects outside of the submitted work. The authors declare no other conflicts of interest.

Peer review information Alison Farrell is the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-1076-0>.

- ⁵Unitat de Tuberculosis, Hospital Universitari Vall d'Hebron-Drassanes, Barcelona, Spain
- ⁶Unitat de TDO de la Tuberculosis 'Servicios Clínicos', Barcelona, Spain
- ⁷Institute for Epidemiology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
- ⁸Lung Clinic Grosshansdorf, Airway Research Center North (ARCN), Großhansdorf, Germany
- ⁹Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
- ¹⁰Department of Respiratory Medicine, Liverpool Hospital, Sydney, Australia
- ¹¹Institut d'Investigació Germans Trias i Pujol, Badalona, Barcelona, Spain
- ¹²CIBER Enfermedades Respiratorias, Badalona, Barcelona, Spain
- ¹³Universitat Autònoma de Barcelona, Badalona, Barcelona, Spain
- ¹⁴Department of Infectious Diseases, The Alfred and Monash University, Melbourne, Australia
- ¹⁵Disease Elimination Program, Burnet Institute, Melbourne, Australia
- ¹⁶KNCV Tuberculosis Foundation, The Hague, The Netherlands
- ¹⁷Institute for Medical Microbiology and Hospital Hygiene, Philipps University of Marburg, Marburg, Germany
- ¹⁸Respiratory Biomedical Research Centre, Institute for Lung Health, Department of Respiratory Sciences, University of Leicester, Leicester, UK
- ¹⁹Hesse State Health Office, Dillenburg, Germany
- ²⁰International Reference Laboratory of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark
- ²¹British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada
- ²²Division of Clinical Infectious Diseases, Research Center Borstel, Borstel, Germany
- ²³German Center for Infection Research (DZIF), Clinical Tuberculosis Center, Borstel, Germany
- ²⁴Tuberculosis Network European Trials Group (TBnet), Borstel, Germany
- ²⁵Department of Medicine, Karolinska Institute, Stockholm, Sweden
- ²⁶Department of Epidemiology, Helmholtz Centre for Infection Research, Braunschweig, Germany
- ²⁷Amsterdam Institute for Global Health and Development, Amsterdam, the Netherlands
- ²⁸Department of Global Health, Amsterdam University Medical Centres, Amsterdam, the Netherlands
- ²⁹Department of Clinical Sciences, University of Barcelona, Barcelona, Spain
- ³⁰Department of Transplant and Infection Immunology, Saarland University, Homburg, Germany
- ³¹Department of Paediatrics and Child Health, Desmond Tutu TB Centre, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

³²Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Sassari, Italy

³³Research Institute of Tuberculosis, Tokyo, Japan

³⁴Swiss Lung Association, Berne, Switzerland

³⁵Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

³⁶Division of Epidemiology and Biostatistics, School of Public Health, University of Cape Town, Cape Town, South Africa

³⁷UCL-TB and UCL Respiratory, University College London, London, UK

³⁸Royal Free London NHS Foundation Trust, London, UK

³⁹Division of Infection & Immunity, University College London, London, UK

These authors contributed equally to this work.

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% confidence interval (CI), 8.0-29.2%) among child contacts, 4.8% (95% CI, 3.0-7.7%) among adult contacts, 5.0% (95% CI, 1.6-14.5%) among migrants and 4.8% (95% CI, 1.5-14.3%) among immunocompromised groups. We confirmed highly variable estimates within risk groups, necessitating an individualized approach to risk stratification. Therefore, we developed a personalized 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 (95%CI, 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 current crude approach to TB risk estimation among people with LTBI in favor of our evidence-based and patient-centered method, in settings aiming for pre-elimination worldwide.

Globally, TB accounts for the greatest number of deaths from a single pathogen, with an estimated 1.5 million deaths and 10 million incident cases in 2018¹. The World Health Organization's End TB Strategy ambitiously aims for a 95% reduction in TB mortality and a 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 per 100,000) by 2035². Preventative antimicrobial treatment for LTBI is considered critical for achieving this objective^{2,3}. In the absence of an assay to detect viable *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 previous treatment^{4,5}. Individuals with LTBI are generally considered to have a lifetime TB risk ranging from 5% to 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 less than 5% over a 2-year period among risk groups, such as adult TB contacts^{7–9}. This might 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 might also undermine the cascade of care, including uptake of preventative treatment among individuals in 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 to 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 might improve predictive ability^{14,15}. However, implementing higher diagnostic thresholds alone does not improve prediction on a population level owing to a marked loss of sensitivity with this approach¹⁶.

In this study, we first sought to characterize the population risk of TB among people tested for LTBI using an individual participant data meta-analysis (IPD-MA). To study progression from LTBI to TB disease more accurately, we focused on settings with low transmission (defined as annual incidence < 20 per 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. Finally, we developed and validated 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 Fig. 1). Of these, we received 18 individual-level data sets, including participants recruited in 20 countries. The pooled data set included a total of 82,360 individual records; of these individuals, 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 Fig. 2). The characteristics of the included study data sets are summarized 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 years). 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 of more than 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 6/6, 6/7 or 7/7 (Supplementary Table 4).

Population-level analysis

In the pooled data set, the 2-year cumulative risk of incident TB was estimated as 4.0% (95% CI, 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 (Fig. 1 and 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% (95% CI, 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 owing 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 owing to immunocompromise. Among recent child contacts, risk was markedly higher among those younger than 5 years old compared to those aged 5–14 years (2-year risk, 26.0% (9.4–60.1) versus 12.4% (5.7–25.6); Fig. 1).

Among child contacts, 85.4% and 93.7% of cumulative risk was accrued in the first 1 and 2 years of follow-up, respectively. Among adult contacts and migrants, the annual risk also declined markedly with time. Of the cumulative 5-year risk, 58.2% and 77.6% were accrued in the first 1 and 2 years of follow-up for adult contacts, with corresponding values among migrants of 66.4% and 81.6%, respectively. 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 Figs. 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 owing to between-study heterogeneity, ranged from 54% to 91% for incidence rates during the 0–2-year interval among untreated people with LTBI, when stratified by indication for screening; forest plots are shown in Extended Data Fig. 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 patients with TB among 31,721 participants from 15 studies (Extended Data Fig. 2). All of these data sets were used for model development and validation, using the internal-external cross-validation (IECV) framework³⁴ described below. Characteristics of the studies included in prediction model development and validation were similar to those that were not (Table 1). Our modeling approach used a flexible parametric survival model with two degrees of freedom on a proportional hazards scale, because this showed the best fit in each imputed data set. From our list of a priori variables of interest, we evaluated nine candidate predictors, of which only previous Bacille Calmette–Guérin (BCG) vaccination and gender were omitted from the final model. The final prediction model included age, a composite ‘TB exposure’ variable (modeled 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 hematological transplant receipt, normalized LTBI test result and preventative treatment commencement. The final model coefficients and standard errors, pooled across multiply imputed data sets, are summarized in Supplementary Table 6, with visual representations of associations between each variable and incident TB risk shown in Fig. 2.

IECV—Next, we used the IECV framework, iteratively discarding one study data set 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 Fig. 3. We assessed discrimination using the C-statistic, which ranged from 0.78 (95% CI, 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 (Fig. 3). Visual calibration plots suggested reasonable calibration in most studies (Extended Data Fig. 6). Because incident TB is an infrequent outcome, predictions were appropriately low, with average predicted risk less than 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 (95% CI, –0.24 to 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 to 0.19)). The pooled random effects meta-analysis calibration slope estimate was 1.11 (0.83–1.38). Slopes appeared heterogeneous, although visual assessment of calibration plots suggested that these

were prone to being extreme owing 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 1) binary LTBI test result and 2) indication for screening (among those with a positive test). The median predicted 2-year TB risk was 2.0% (IQR, 0.8–3.7%) and 0.2% (IQR, 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 (Fig. 4c), with illustrative predicted survival curves for one randomly sampled individual patient per quartile shown in Fig. 4d.

Decision curve analysis

Net benefit quantifies the tradeoff 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% to 20% (reflecting a range of clinician and patient preferences) (Fig. 5).

Sensitivity analyses

We re-examined population-level TB risk without any exclusion of prevalent TB (cases diagnosed <42 d from testing), resulting in markedly higher cumulative risk for each risk group (Extended Data Fig. 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 (95% CI, 0.82–0.93), similar to the primary analysis (Extended Data Fig. 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 owing to the high negative predictive value of LTBI tests when using standard cutoffs (Extended Data Fig. 7b). Finally, 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 Fig. 7c), and net benefit appeared higher when using this model than the strategies of treating either 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 Fig. 7d).

Discussion

In this study, we examined population-level incident TB risk in a pooled data set of more than 80,000 individuals tested for LTBI in 20 countries with low *M. tuberculosis* transmission (annual incidence ~20 per 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 immunocompromized individuals. Most 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 to capture heterogeneity within risk groups and generate personalized risk predictions for incident TB in settings aiming for pre-elimination. Clinical covariates 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 hematological 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 individualized 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 toward 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, parameterized using aggregate data estimates from multiple sources, seeks to estimate TB risk after 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 data set, 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 of 26 (69%) eligible studies. This allowed us to examine progression from LTBI to TB disease using the largest adult and pediatric data set available to date, to our knowledge. We conducted population-level analyses using both one- and two-stage IPD-MA approaches 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 after 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 data sets 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 for 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 might 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. Although increasing data support a role for BCG vaccination in reducing sensitization 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 covariates, 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 to be negative when missing). This sensitivity analysis revealed similar findings to the primary analyses, although effect estimates were noted to be more extreme in the complete case approach, likely owing to a degree of bias in the latter, because complete cases analysis assumes no association between the pattern of missingness and the outcome (that is, 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, because the aim of our prognostic model was to assess the risk of incident TB, after prevalent TB has been clinically ruled out, 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 this study is that its generalizability is restricted to low transmission settings (annual incidence 20 per 100,000 persons). The rationale for limiting to such settings was, first, to examine progression from LTBI to TB disease more accurately by reducing risk of re-infection with *M. tuberculosis* during follow-up. Second, most 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. Because the quantitative LTBI test result is a strong predictor in our model, a different prediction model might, 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, 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) might 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 might evaluate the full health economic effect of programmatic implementation of the model.

A further limitation is that, owing to a lack of data from contributing studies, other potential predictors that might be associated with incident TB risk (including diabetes, malnutrition, fibrotic chest x-ray lesions and other immunosuppression)⁴ were not evaluated. These unmeasured covariates might have contributed to imperfect discrimination and calibration, along with residual heterogeneity in model performance between data sets. 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 might 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 covariate 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, because no data sets 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' modeling approach, with

mathematical parameterization of relative risk for any given biologic agent, might be required to extend its application to these therapies. Because the quantitative LTBI test result is a strong predictor in our model, predictions might also be attenuated in the context of advanced immunosuppression⁷. Reassuringly, performance appeared adequate in a data set of immunocompromised individuals during validation²⁹.

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

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 from January 1, 2002, to December 31, 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 20 per 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 when 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 was examined by comparing original reported results with re-analyzed 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 10-mm cutoff 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 and QuantiFERON-TB Gold Plus; Qiagen), which were assumed

to be equivalent²⁵, and T-SPOT.TB (Oxford Immunotec). 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 d 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 World Health Organization⁵². 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 versus negative) and provision of LTBI treatment (commenced versus not commenced). We further examined progression risk among untreated participants with LTBI, stratified by indication for screening (recent child contacts (<15 years) versus adult contacts versus migrants versus immunocompromised), by separately fitting random effect flexible parametric survival models to each risk group. Child contacts were further stratified by age (<5 years versus 5–14 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 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, because 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 data sets to determine eligibility for inclusion as candidate predictors in the model. We determined that the following predictors were available from a sufficient number of data sets 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 hematological transplant receipt, HIV status and TB preventative treatment commencement.

Variable transformations—Previous data showed that quantitative TST, QuantiFERON Gold-in-Tube and T-SPOT.TB results are associated with risk of incident TB¹⁶. However,

each LTBI test was 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 subpopulation of the entire cohort where all three tests were performed among the same participants in head-to-head studies. We normalized quantitative results for the TST, QuantiFERON Gold-in-Tube and T-SPOT.TB to a percentile scale using this head-to-head population and examined the association between normalized result and risk of incident TB using Cox proportional hazards models with restricted cubic splines. Because TST cutoffs 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. Because 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 normalized percentile results for each test (unadjusted TST, QuantiFERON Gold-in-Tube and T-SPOT.TB) appeared to be associated with similar risk of incident TB (Extended Data Fig. 8). The LTBI tests implemented differed between contributing studies. From this point, all LTBI test results were, therefore, normalized to this percentile scale to enable data harmonization 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). Because most people evaluated for LTBI under routine programmatic conditions have a single test performed, we included only 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 (because this was the most commonly used test in the data set), followed by T-SPOT.TB and then the TST.

Recent contacts were categorized as either 'smear positive and household' or 'other' contacts, because there was no evidence of separation of risk among additional subgroups of the 'other' contacts stratum during exploratory univariable analyses (Extended Data Fig. 8). Because we considered migration from a high TB-burden country (defined as annual TB incidence 100 per 100,000 persons at the year of migration) to be a proxy for previous 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 100 per 100,000 persons) among migrants or when stratified by country of birth among recent contacts (Extended Data Fig. 8).

Age and normalized test result variables were modeled using restricted cubic splines (using a default of five knots placed at recommended intervals⁵⁹) to account for their nonlinear associations with incident TB.

Multiple imputation—A data dictionary and a summary of missingness of candidate predictor variables are 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 owing 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 owing to expected heterogeneity between these groups. We generated ten multiply imputed data sets, 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 data sets; 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 owing to HIV infection with missing data on transplant status, this was assumed to be negative owing 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 data sets using AIC. Variables that were selected in more than 50% of the imputed data sets were included in the final model. T cell responses to *M. tuberculosis* might 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, 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 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 data sets. 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. Because the AIC for the time-varying covariate model was lower across all imputed data sets, this time-varying covariate approach was used for the final model.

IECV—After development of the final model, we used the IECV framework for model validation, allowing concurrent assessment of between-study heterogeneity and generalizability³⁴. In this process, one entire contributing study data set is iteratively discarded from the model training set and used for external validation. This process is repeated until each data set has been used once for validation. The primary outcome for validation was 2-year risk of incident TB. We included data sets with a minimum of five 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 versus Kaplan–Meier-estimated observed 2-year risk in quintiles and using the calibration slope and CITL statistics⁶⁵. Calibration slopes greater than 1 suggest under-fitting (predictions are not varied enough), whereas slopes less than 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 data set. Validation set C-statistics, calibration slopes and CITL metrics were pooled for each study across imputations using Rubin’s rules⁶¹. We then meta-analyzed 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 analyzed 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 0–180 d of recruitment); a complete case approach (for all variables except for HIV status, which was assumed to be 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).

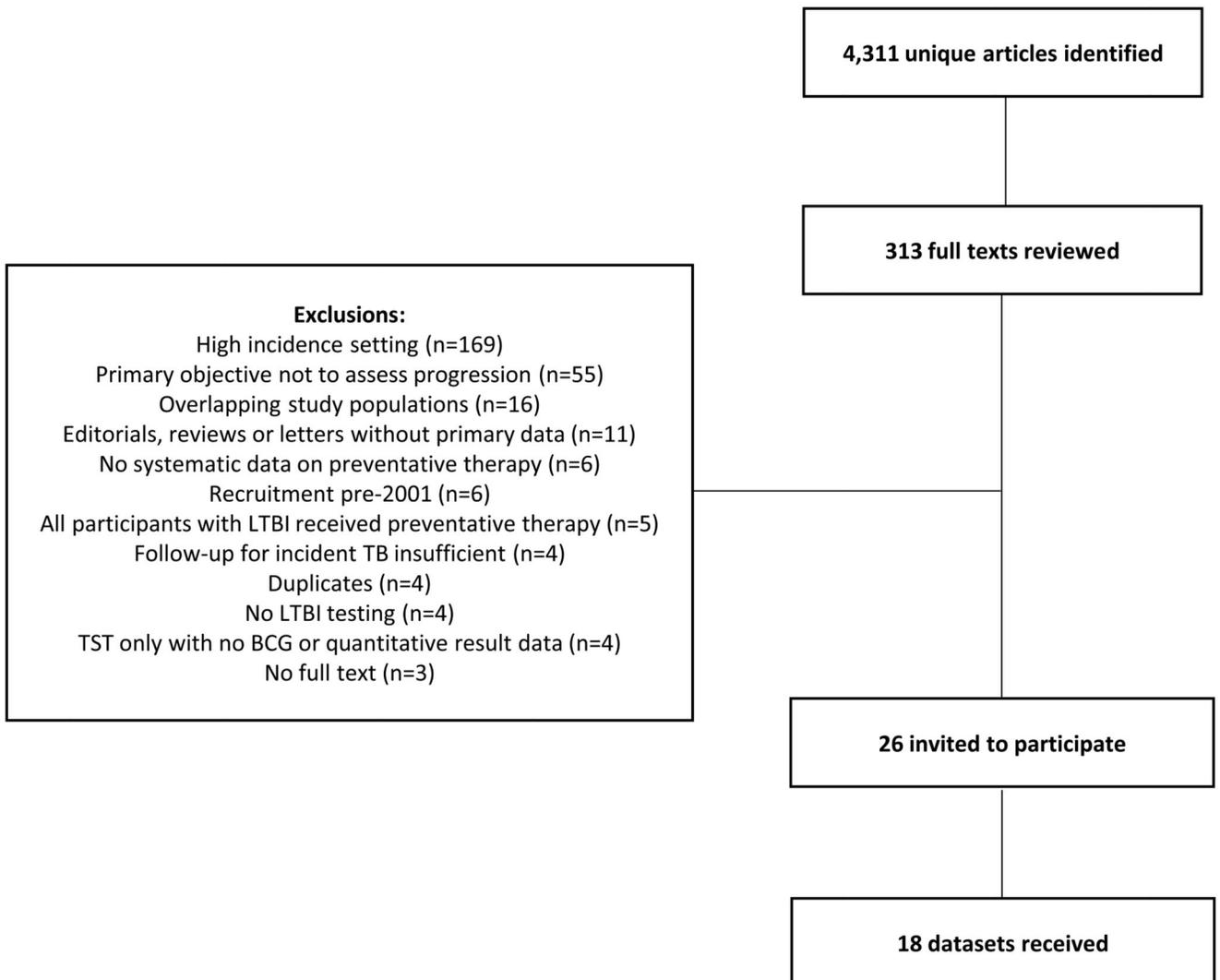
We also examined IECV discrimination parameters for validation data sets when 1) restricted to participants with positive binary LTBI tests; 2) excluding those who received preventative treatment; and 3) imputing an average quantitative positive or negative LTBI

test result (based on the medians among the study population), according to the binary result. The latter analysis was done to assess model performance in situations where the quantitative test result was not available.

Ethics

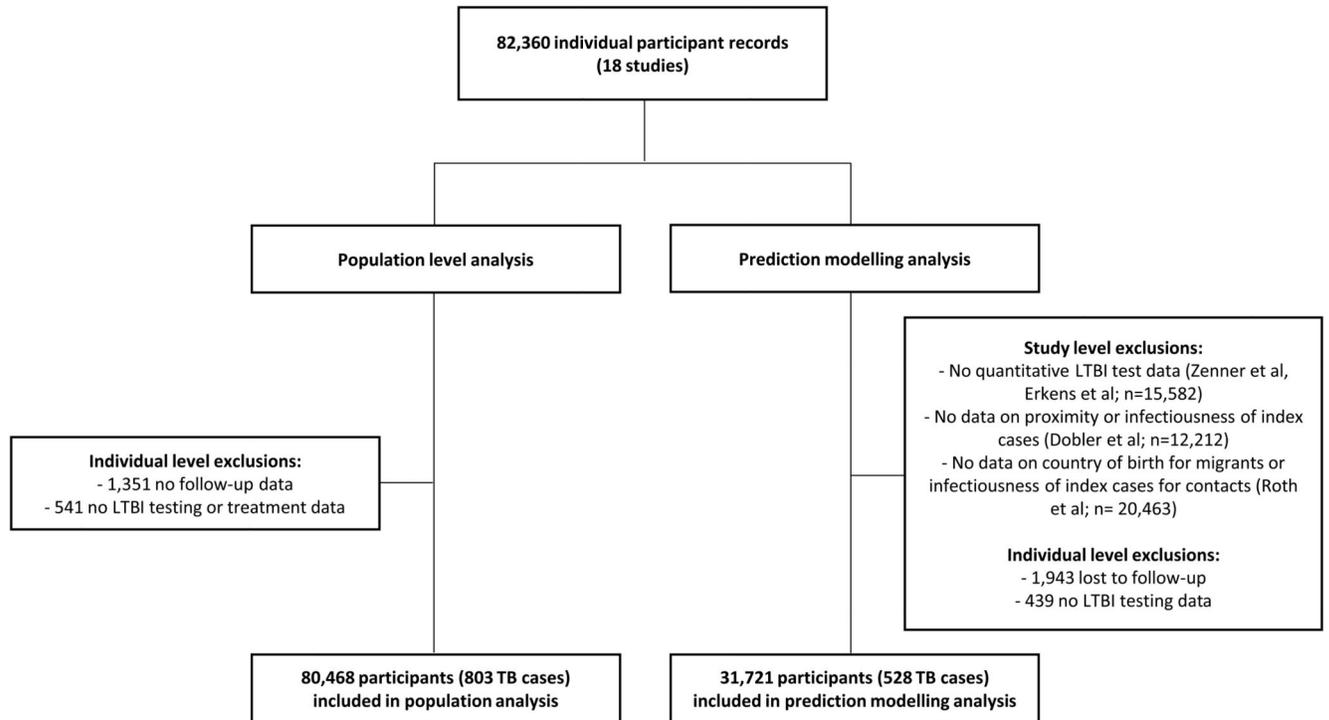
This study involved analyses of fully de-personalized data from previously published cohort studies, with data pooling via a safe haven. Ethical approvals for sharing of data were sought and obtained by contributors of individual participant data, where required.

Extended Data



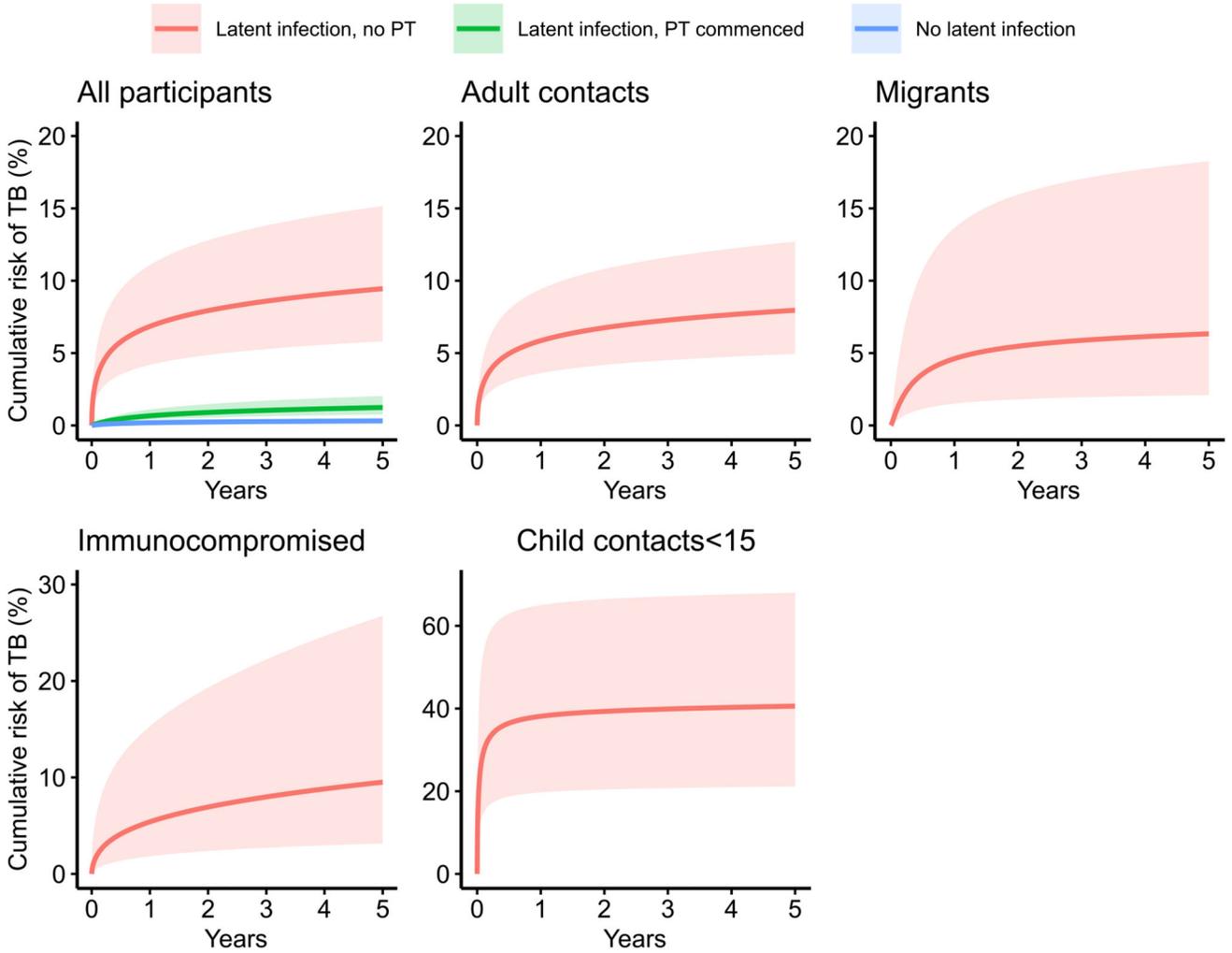
Extended Data Fig. 1. Flow chart outlining systematic review process.

The systematic search strategy and eligibility criteria are shown in Supplementary Tables 8 and 9.

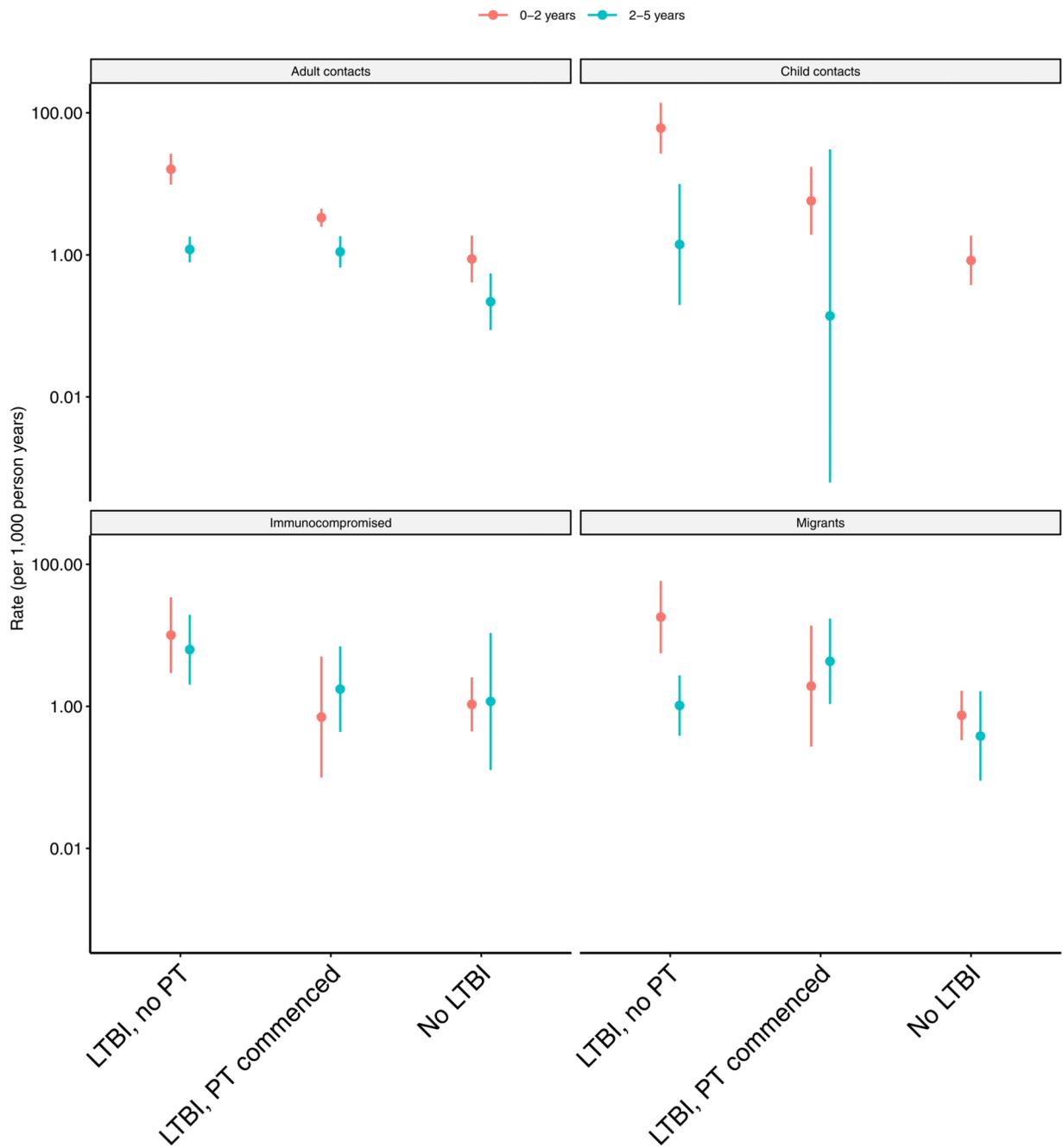


Extended Data Fig. 2. Flow chart showing inclusion of participants in the population-level and prediction modelling analyses.

The systematic search strategy and eligibility criteria are shown in Supplementary Tables 8 and 9.

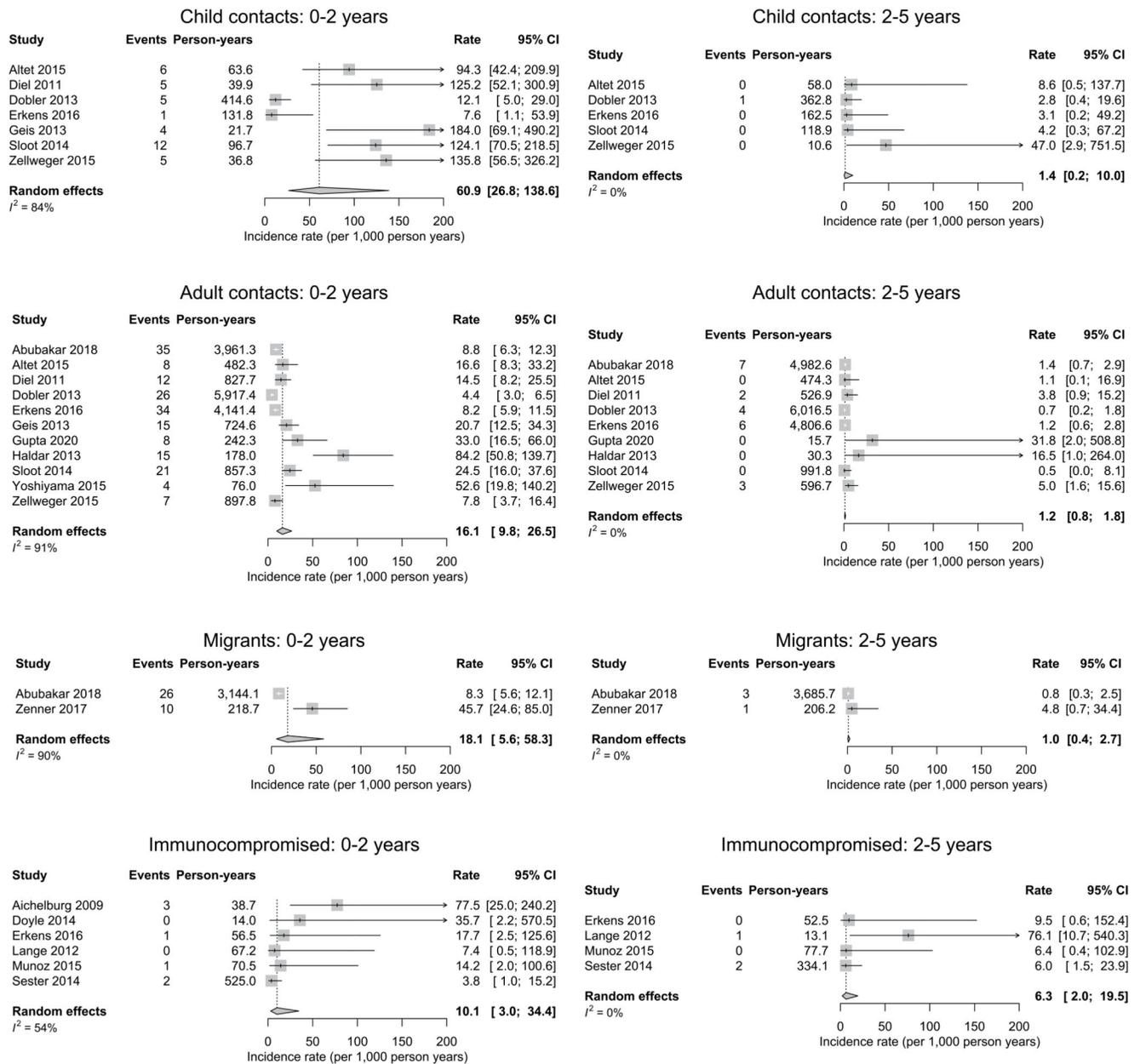


Extended Data Fig. 3. Cumulative risk of prevalent and incident tuberculosis during follow-up. Risk is stratified by binary latent TB test result, provision of preventative treatment, and indication for screening among participants with untreated latent infection (total n = 80,468 participants). Cumulative risk is estimated using flexible parametric survival models with random effects for the intercept by source study, separately fitted to each risk group. Prevalent TB cases (diagnosed within 42 days of recruitment) are included in this sensitivity analysis. Each plot is presented as point estimates (solid line) and 95% confidence intervals (shaded area). PT = preventative treatment.



Extended Data Fig. 4. Pooled TB incidence rates among adults, stratified by risk group. Pooled incidence rates are shown on log₁₀ scale among participants with: latent TB infection (LTBI) with no preventative therapy (PT); LTBI commencing PT; and without evidence of LTBI. Rates are further stratified by follow-up interval (0–2 years vs. 2–5 years) and indication for screening (total n = 52,576 participants). Pooled incidence rate estimates were derived from random intercept Poisson regression models, without continuity correction for studies with zero events. Numeric results are shown for the subgroups with untreated latent TB infection in the forest plots in Extended Data Fig. 5. Plots show point

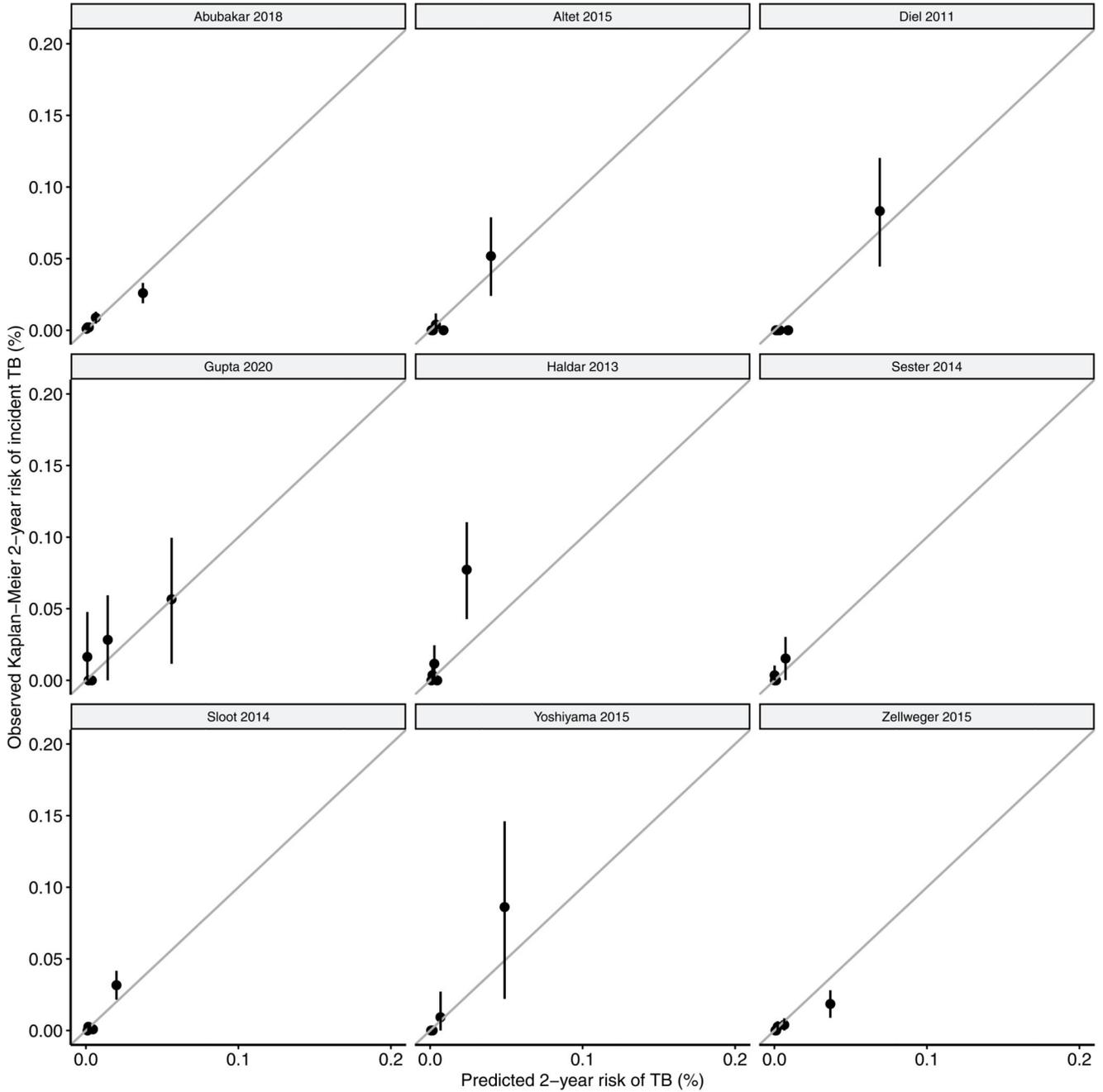
estimates (filled circles) and 95% confidence intervals (vertical error bars). No pooled estimate could be calculated for child contacts without evidence of LTBI for the 2–5 year interval since there were no incident events.



Extended Data Fig. 5. Forest plots showing incidence rates by source study among participants with untreated LTBI.

Forest plots are stratified by follow-up interval (0–2 years vs. 2–5 years) and indication for screening (total n = 52,576 participants). Pooled incidence rate estimates (shown as diamonds) were derived from random intercept Poisson regression models, without continuity correction for studies with zero events. Incidence rates per study are shown with

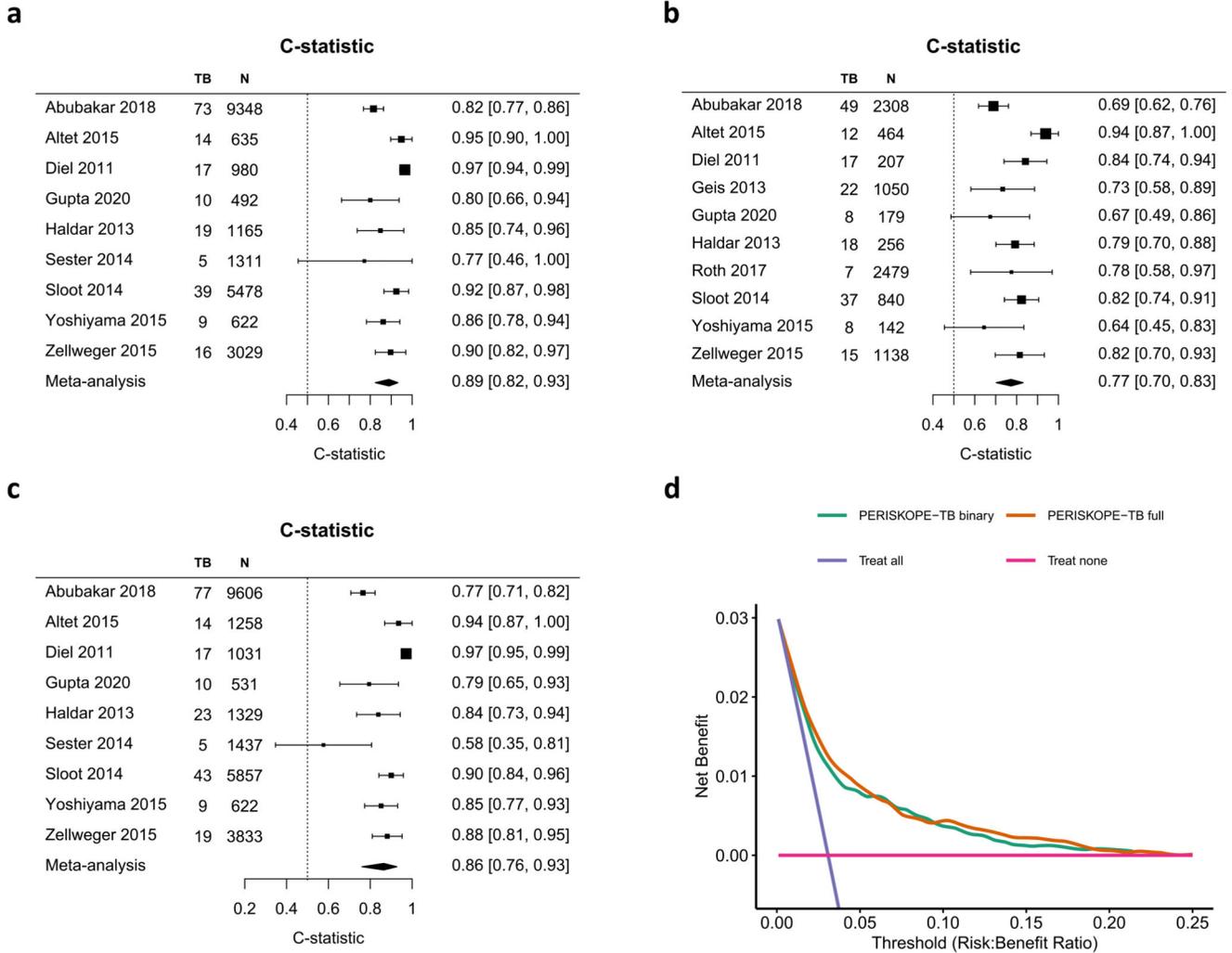
a continuity correction of 0.5 for studies with zero events. Plots show study-level point estimates (grey squares) and 95% confidence intervals (CIs; horizontal error bars).



Extended Data Fig. 6. Calibration plots from internal-external validation of prediction model, stratified by validation study.

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). X-axis shows predicted risk, in quintiles, with corresponding Kaplan Meier 2-year

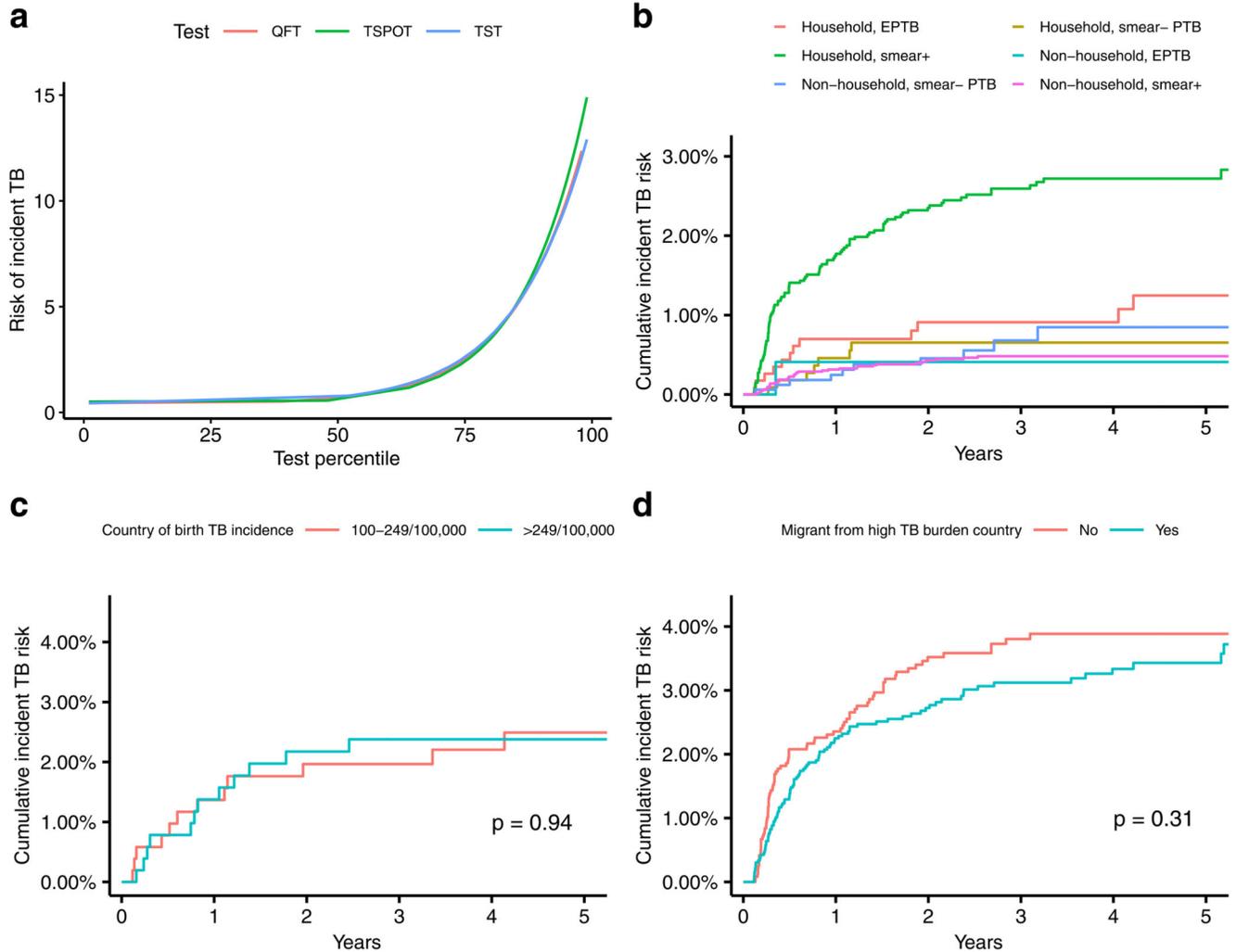
risk of incident TB on the Y-axis (95% confidence intervals are shown by vertical error bars).



Extended Data Fig. 7. Model validation sensitivity analyses.

Forest plots showing recalculation of the C-statistics from internal-external cross validation, limiting validation sets to **a**, participants who did not receive preventative therapy (n = 23,060 participants); **b**, participants with a positive LTBI test (n = 9,063 participants); and **c**, binary LTBI test results (using an average quantitative positive or negative LTBI test result as appropriate, based on the medians among the study population; n = 25,504 participants). ‘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. Panel **d**, shows decision curve analyses (n = 6,418 participants) when using the prediction model using a binary LTBI test result, compared to the full prediction model, ‘treat all’ and ‘treat none’ strategies across a range of threshold probabilities (x-axis). Net benefit appeared higher for the binary model than either the strategies of treating all patients

with evidence of LTBI, or no patients, throughout the range of threshold probabilities. The full model had highest net benefit across most threshold probabilities.



Extended Data Fig. 8. Data supporting assumptions underlying PERISKOPE-TB model.

a, Quantitative results for the tuberculin skin test (TST), QuantiFERON Gold-in-tube (QFT-GIT) and T-SPOT.TB are normalised to a percentile scale using a head-to-head population among whom all three tests were performed from 3 studies including recent TB contacts, migrants and immunocompromised participants ($n = 8,335$; 158 TB cases). We examined the association between normalised test result and risk of incident TB using Cox proportional hazards models with restricted cubic splines. Normalised results for each test appeared to be associated with similar risk of incident TB. **b**, Kaplan Meier plots from pooled dataset showing cumulative risk of incident TB, stratified by proximity and infectiousness of index cases among contacts ($n = 22,231$ participants). There was no evidence of separation of risk of additional subgroups of the ‘other’ (non-smear positive household) contacts stratum. PTB = pulmonary TB; EPTB = extra-pulmonary TB. **c**, Kaplan Meier plots from pooled dataset showing cumulative risk of incident TB among people with positive latent TB

tests, stratified by TB incidence in country of birth among migrants from high TB burden countries (n = 1,031 participants). P value represents Log-rank test. **d**, Kaplan Meier plots from pooled dataset showing cumulative risk of incident TB among people with positive latent TB tests, stratified by country of birth among recent contacts (n = 5,917 participants). P value represents Log-rank test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was funded by the National Institute for Health Research (NIHR) (DRF-2018–11-ST2-004 to R.K.G. and SRF-2011-04-001 and NF-SI-0616-10037 to I.A.), the Wellcome Trust (207511/Z/17/Z to M.N.) and NIHR biomedical research funding to University College London Hospitals. C.L. is funded by the German Center for Infection Research. 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 authors and not necessarily those of the National Health Service, 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 for this analysis.

Data availability

The individual participant data pooled for this analysis are subject to data sharing agreements with the original study authors. The data might be shared with 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 data sets (periskope.org). The code underlying the prediction tool is available at github.com/rishi-k-gupta/PERISKOPE-TB.

References

1. World Health Organization. Global Tuberculosis Report. 2019. https://www.who.int/tb/publications/global_report/en/
2. World Health Organization. The End TB Strategy. 2015. http://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1
3. Getahun H, et al. Management of latent *Mycobacterium tuberculosis* infection: WHO guidelines for low tuberculosis burden countries. *Eur Respir J*. 2015; 46: 1563–1576. [PubMed: 26405286]
4. Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent *Mycobacterium tuberculosis* infection. *N Engl J Med*. 2015; 372: 2127–2135. [PubMed: 26017823]
5. Mack U, et al. LTBI: latent tuberculosis infection or lasting immune responses to *M. tuberculosis*? A TBNET consensus statement. *Eur Respir J*. 2009; 33: 956–973. [PubMed: 19407047]
6. Sterling TR, et al. Guidelines for the treatment of latent tuberculosis infection: recommendations from the National Tuberculosis Controllers Association and CDC, 2020. *MMWR Recomm Rep*. 2020; 69: 1–11.
7. Pai M, et al. Gamma interferon release assays for detection of *Mycobacterium tuberculosis* infection. *Clin Microbiol Rev*. 2014; 27: 3–20. [PubMed: 24396134]

8. Rangaka MX, et al. Predictive value of interferon- γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis.* 2012; 12: 45–55. [PubMed: 21846592]
9. Abubakar I, et al. Prognostic value of interferon- γ release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet Infect Dis.* 2018; 18: 1077–1087. [PubMed: 30174209]
10. Gao J, et al. Knowledge and perceptions of latent tuberculosis infection among Chinese immigrants in a Canadian urban centre. *Int J Fam Med.* 2015; 2015 546042
11. Alsdurf H, Hill PC, 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.* 2016; 16: 1269–1278. [PubMed: 27522233]
12. Campbell JR, Winters N, Menzies D. Absolute risk of tuberculosis among untreated populations with a positive tuberculin skin test or interferon- γ release assay result: systematic review and meta-analysis. *BMJ.* 2020; 368 m549 [PubMed: 32156698]
13. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ.* 2018; 362 k2738 [PubMed: 30139910]
14. Winje BA, et al. Stratification by interferon- γ release assay level predicts risk of incident TB. *Thorax.* 2018; 73: 652–661.
15. Andrews JR, et al. Serial QuantiFERON testing and tuberculosis disease risk among young children: an observational cohort study. *Lancet Respir Med.* 2017; 5: 282–290. [PubMed: 28215501]
16. Gupta RK, et al. Quantitative IFN- γ release assay and tuberculin skin test results to predict incident tuberculosis: a prospective cohort study. *Am J Respir Crit Care Med.* 2020; 201: 984–991. [PubMed: 31825645]
17. Altet N, et al. Predicting the development of tuberculosis with the tuberculin skin test and QuantiFERON testing. *Ann Am Thorac Soc.* 2015; 12: 680–688. [PubMed: 25699406]
18. Diel R, Loddenkemper R, Niemann S, Meywald-Walter K, Nienhaus A. Negative and positive predictive value of a whole-blood interferon- γ release assay for developing active tuberculosis: an update. *Am J Respir Crit Care Med.* 2011; 183: 88–95. [PubMed: 20802162]
19. Dobler CC, Marks GB. Risk of tuberculosis among contacts in a low-incidence setting. *Eur Respir J.* 2013; 41: 1459–1461. [PubMed: 23728408]
20. Geis S, et al. How can we achieve better prevention of progression to tuberculosis among contacts? *Eur Respir J.* 2013; 42: 1743–1746. [PubMed: 23900989]
21. Haldar P, et al. Single-step QuantiFERON screening of adult contacts: a prospective cohort study of tuberculosis risk. *Thorax.* 2013; 68: 240–246. [PubMed: 22956558]
22. Sloot R, Van Der Loeff MFS, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure: a 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med.* 2014; 190: 1044–1052. [PubMed: 25265362]
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.* 2015; 47: 542–549.
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.* 2015; 191: 1176–1184. [PubMed: 25763458]
25. Gupta RK, et al. Evaluation of QuantiFERON-TB Gold Plus for predicting incident tuberculosis among recent contacts: a prospective cohort study. *Ann Am Thorac Soc.* 2020; 17: 646–650. [PubMed: 32083944]
26. Aichelburg MC, et al. Detection and prediction of active tuberculosis disease by a whole-blood interferon- γ release assay in HIV-1-infected individuals. *Clin Infect Dis.* 2009; 48: 954–962. [PubMed: 19245343]
27. Doyle JS, et al. Latent tuberculosis screening using interferon- γ release assays in an Australian HIV-infected cohort: is routine testing worthwhile? *J Acquir Immune Defic Syndr.* 2014; 66: 48–54. [PubMed: 24457631]
28. Lange B, Vavra M, Kern WV, Wagner D. Development of tuberculosis in immunocompromised patients with a positive tuberculosis-specific IGRA. *Int J Tuberc Lung Dis.* 2012; 16: 492–495. [PubMed: 22325468]

29. Sester M, et al. Risk assessment of tuberculosis in immunocompromised patients. A TBNET study. *Am J Respir Crit Care Med.* 2014; 190: 1168–1176. [PubMed: 25303140]
30. Munoz L, et al. Immunodiagnostic tests' predictive values for progression to tuberculosis in transplant recipients: a prospective cohort study. *Transplant Direct.* 2015; 1: e12. [PubMed: 27500217]
31. Roth DZ, et al. Impact of interferon- γ release assay on the latent tuberculosis cascade of care: a population-based study. *Eur Respir J.* 2017; 49 1601546 [PubMed: 28331032]
32. Erkens CGM, et al. Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. *Eur Respir J.* 2016; 48: 1420–1428. [PubMed: 27799389]
33. Zenner D, Loutet MG, Harris R, Wilson S, Ormerod LP. Evaluating 17 years of latent tuberculosis infection screening in north-west England: a retrospective cohort study of reactivation. *Eur Respir J.* 2017; 50 1602505 [PubMed: 28751410]
34. Debray TPA, et al. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med.* 2015; 12 e1001886 [PubMed: 26461078]
35. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res.* 2019; 3: 18. [PubMed: 31592444]
36. Zhang Z, et al. Decision curve analysis: a technical note. *Ann Transl Med.* 2018; 6: 19. [PubMed: 29404365]
37. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016; 69: 245–247. [PubMed: 25981519]
38. Martinez L, et al. The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis. *Lancet.* 2020; 395: 973–984. [PubMed: 32199484]
39. Saunders MJ, 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.* 2017; 17: 1190–1199. [PubMed: 28827142]
40. Saunders MJ, 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.* 2020; 20: 110–122. [PubMed: 31678031]
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.* 2008; 12: 498–505. [PubMed: 18419884]
43. Stewart LA, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data. *JAMA.* 2015; 313: 1657. [PubMed: 25919529]
44. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015; 350 g7594 [PubMed: 25569120]
45. Audigier V, et al. Multiple imputation for multilevel data with continuous and binary variables. *Stat Sci.* 2018; 33: 160–183.
46. Nemes E, et al. Prevention of M. Tuberculosis infection with H4:IC31 vaccine or BCG revaccination. *N Engl J Med.* 2018; 379: 138–149. [PubMed: 29996082]
47. Katelaris AL, et al. Effectiveness of BCG vaccination against Mycobacterium tuberculosis infection in adults: a cross-sectional analysis of a UK-based cohort. *J Infect Dis.* 2020; 221: 146–155. [PubMed: 31504674]
48. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med.* 2011; 30: 377–399. [PubMed: 21225900]
49. Drain PK, et al. Incipient and subclinical tuberculosis: clinical review of early stages and progression of infection. *Clin Microbiol Rev.* 2018; 31 e00021-18
50. Gupta RK, et al. Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet Respir Med.* 2020; 8: 395–406. [PubMed: 31958400]

51. Roe J, et al. Blood transcriptomic stratification of short-term risk in contacts of tuberculosis. *Clin Infect Dis*. 2019; 70: 731–737.
52. World Health Organization. Latent TB Infection: updated and consolidated guidelines for programmatic management. 2018. <http://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/>
53. Groenwold RHH, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol*. 2016; 78: 90–100. [PubMed: 27045189]
54. Wells G, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
55. Zenner D, et al. Treatment of latent tuberculosis infection. *Ann Intern Med*. 2017; 167: 248. [PubMed: 28761946]
56. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019; 22: 153–160. [PubMed: 31563865]
57. Seddon JA, 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*. 2016; 71: 932–939. [PubMed: 27335104]
58. Cobelens FG, et al. Tuberculin skin testing in patients with HIV infection: limited benefit of reduced cutoff values. *Clin Infect Dis*. 2006; 43: 634–639. [PubMed: 16886159]
59. Harrell FE. Biostatistical Modeling. 2004. <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/BioMod/notes.pdf>
60. Audigier V, Resche-Rigon M. micemd: multiple imputation by chained equations with multilevel data. R package, version 1.6.0. 2019.
61. Rubin, DB. Multiple Imputation for Nonresponse In Surveys. Wiley-Interscience; 2004.
62. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002; 21: 2175–2197. [PubMed: 12210632]
63. Clements M, Liu X-R. rstpm2: smooth survival models, including generalized survival models. R package, version 1.5.1. 2019.
64. Bower H, et al. Capturing simple and complex time-dependent effects using flexible parametric survival models: a simulation study. *Commun Stat Simul Comput*. 2019; doi: 10.1080/03610918.2019.1634201
65. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014; 35: 1925–1931. [PubMed: 24898551]
66. Westeneng H-J, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol*. 2018; 17: 423–433. [PubMed: 29598923]
67. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. 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*. 2018; 27: 3505–3522. [PubMed: 28480827]
68. Sjoberg DD. dca: decision curve analysis. R package, version 0.1.0.9000. 2020.

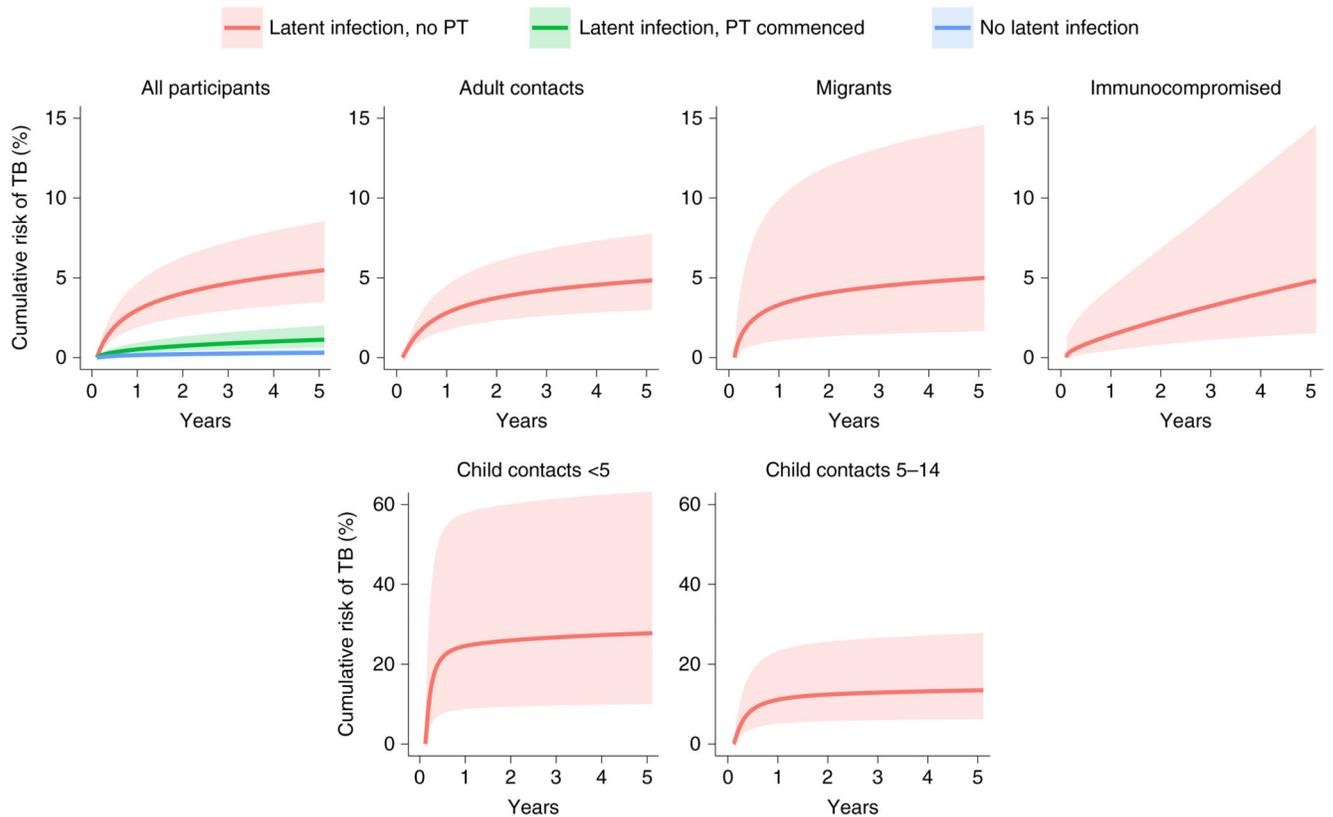


Fig. 1. Population-level cumulative risk of incident TB during follow-up.

Risk is stratified by binary latent TB test result, provision of preventative treatment (PT) 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 d of recruitment) are excluded. Each plot is presented as point estimates (solid line) and 95% CIs (shaded area). Child contacts are shown stratified by age (<5 years and 5-14 years). 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 Fig. 3.

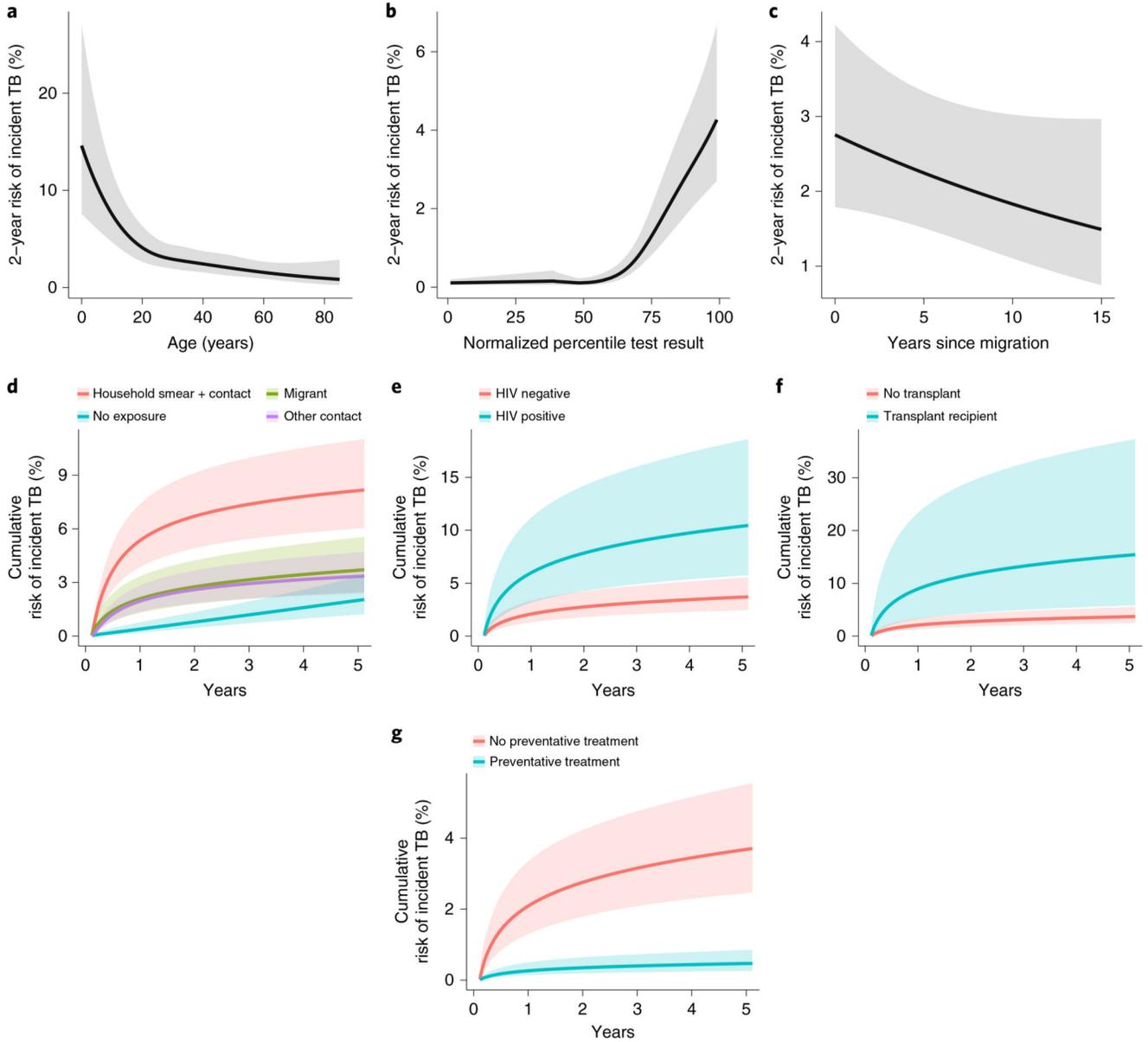


Fig. 2. Visual representations of associations between predictors and incident TB. Illustrative estimates are shown for a 33-year-old migrant from a high TB-burden setting. The example ‘base case’ patient does not commence preventative treatment, is not living with HIV, has not received a previous transplant and has an ‘average’ positive latent TB test. We vary one of these predictors in each plot ((a) age; (b) normalized 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% CIs (shaded area). The model was trained on a pooled data set ($n = 31,090$ 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.

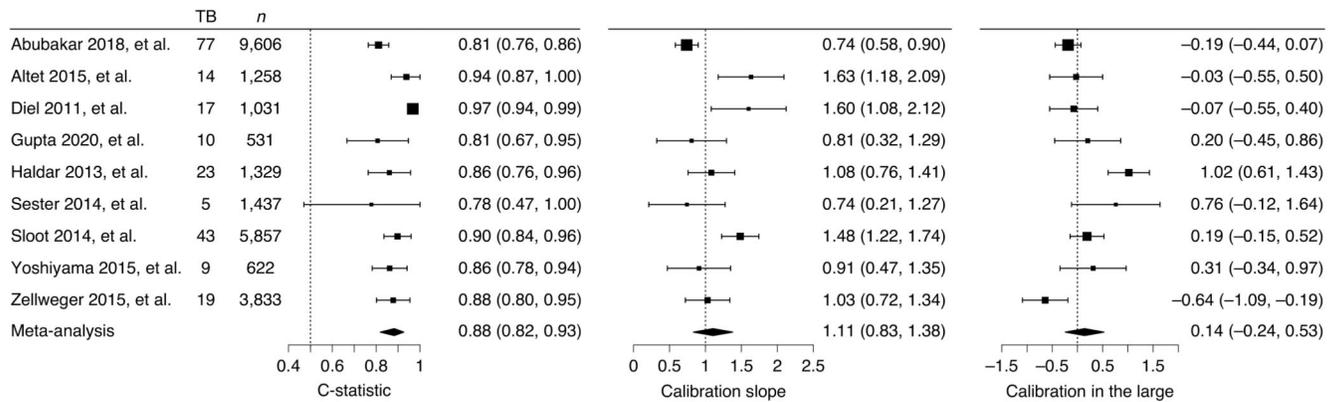


Fig. 3. Forest plots showing model discrimination and calibration metrics for predicting 2-year risk of incident TB.

Discrimination is presented as the C-statistic; calibration is presented as CITL and the calibration slope. Data from nine primary validation studies are shown, from IECV of the model (developed among $n = 31,090$ participants; validated among 25,504 participants 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% CIs (error bars). Pooled estimates are shown as diamonds. Calibration slopes greater than 1 suggest under-fitting (predictions are not varied enough), whereas slopes less than 1 indicate over-fitting (predictions are too extreme). CITL 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.

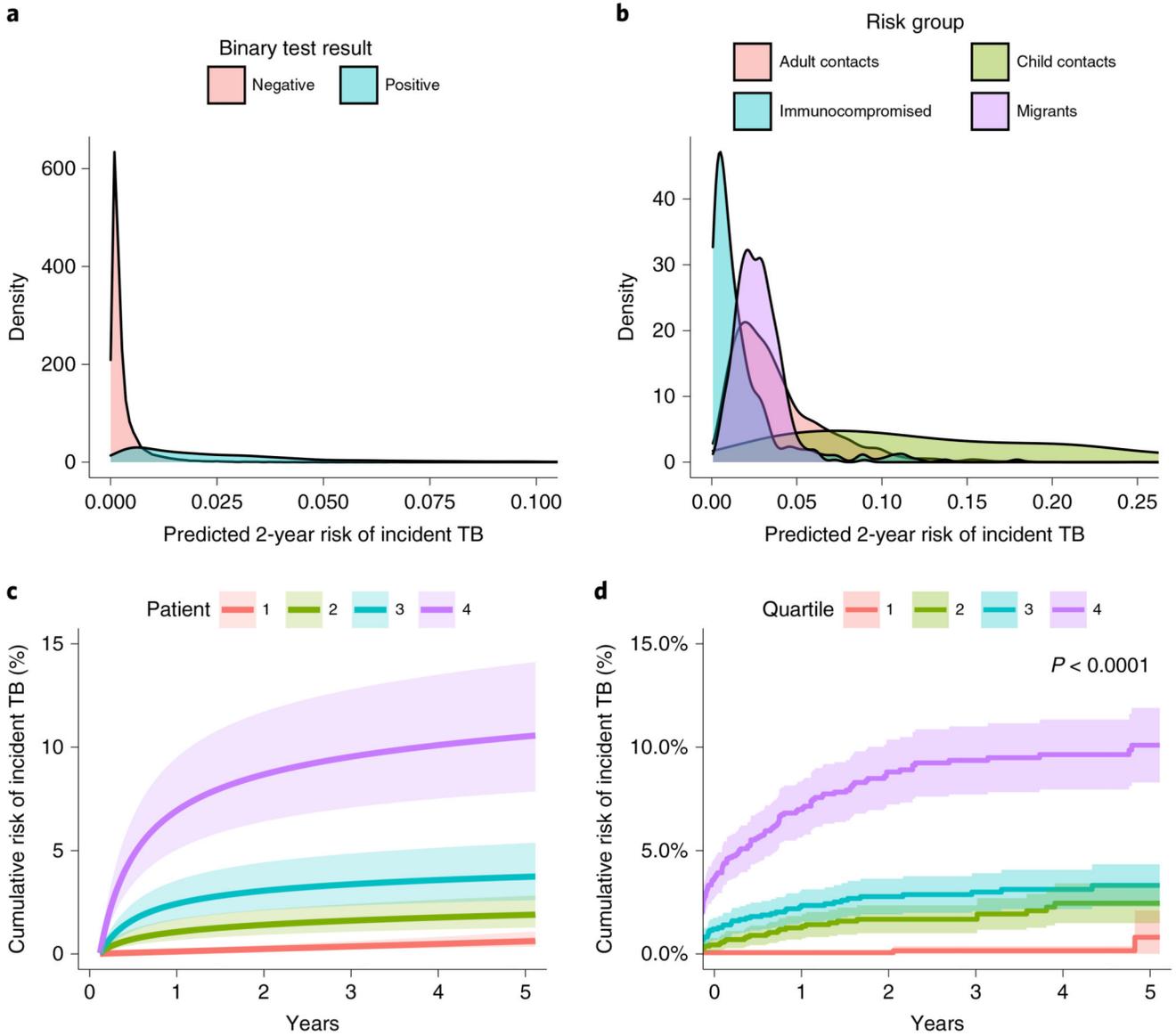


Fig. 4. Distribution of predictions and risk of incident TB in four quartiles of risk for people with positive latent TB tests.

Distribution of risk from prediction model using pooled validation sets of people not receiving preventative therapy from IECV 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 IECV 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 normalized 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 normalized 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 normalized latent TB test result on the 79th percentile; and Patient 4 is a 33-year-old household contact of a smear-positive index TB case with a normalized latent TB test result on the 94th percentile. All four example patients are HIV negative and are not transplant recipients. Equivalent values of normalized percentile test results for QuantiFERON, T-SPOT.TB and TST are shown in Supplementary Table 10. Plots (c, d) are presented as point estimates (solid line) and 95% CIs (shaded area).

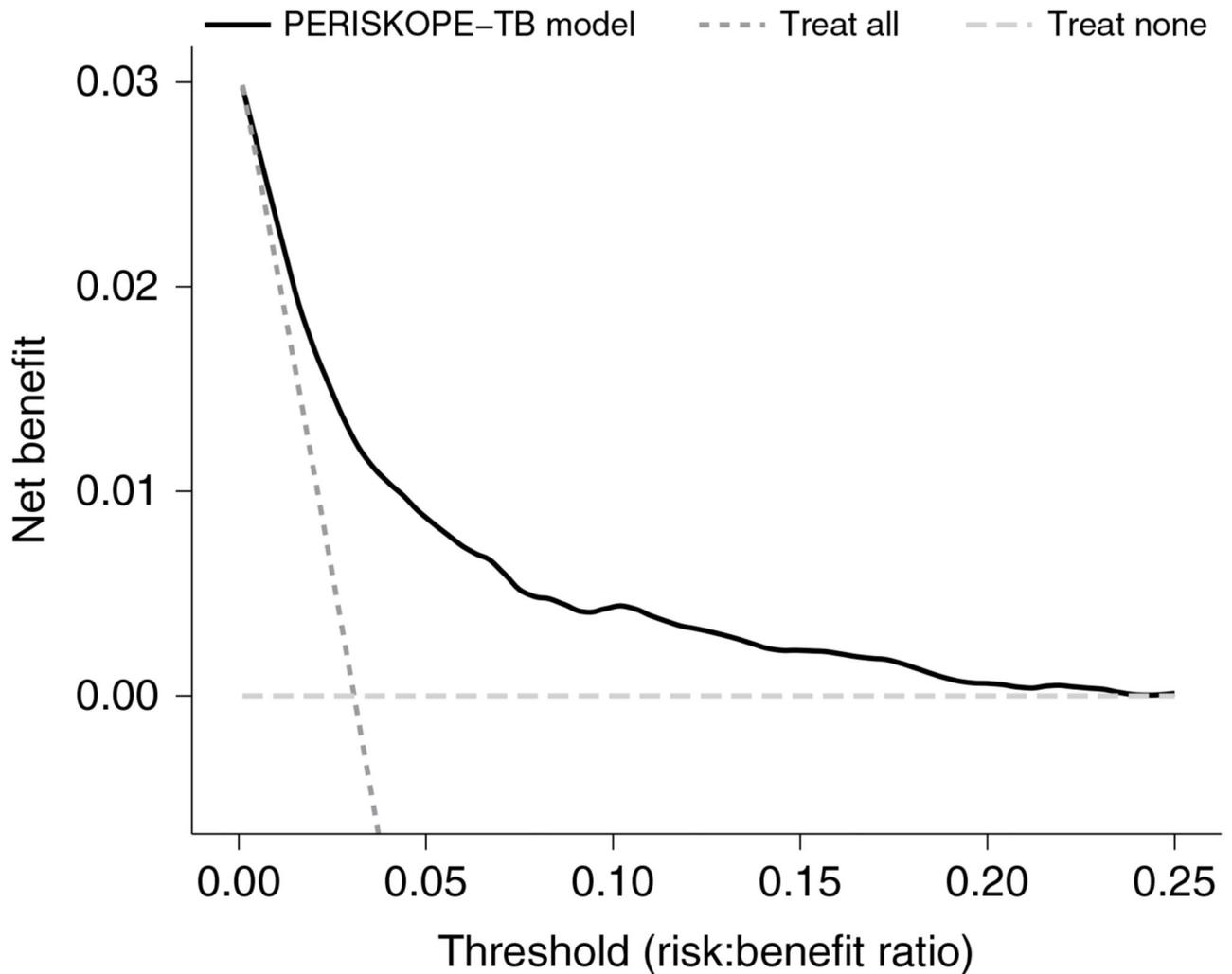


Fig. 5. Decision curve analysis.

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 tradeoff 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 cutoff 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 might have a low threshold probability (for example, 1%, which is equivalent to a risk:benefit ratio of 1:99—that is, 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 might have a higher threshold probability (for example, 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 one patient per 100 tested was appropriately given preventative treatment, as they would otherwise have progressed to incident TB if left untreated.

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

Authors	Publication Year	Country	<i>n</i> (total)	Adults/ children	Population	Follow-up years (median (IQR))	TB cases	Loss to follow-up	Included in prediction modeling	NOS ^a
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 ^b	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 ^b	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 ^b	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

Authors	Publication Year	Country	<i>n</i> (total)	Adults/ children	Population	Follow-up years (median (IQR))	TB cases	Loss to follow-up	Included in prediction modeling	NOS ^a
Zenner et al. ³³	2017	UK	1,341	Adults	Migrants	3.7 (3–4.8)	21	NA	No ^b	7/7
Total			82,360			3.7 (2.1–5.3)	826	2294 (2.8%)		

^aModified version of the Newcastle-Ottawa Scale for cohort studies.

^bNot included in prediction modeling owing to lack of data on proximity or infectiousness of index cases¹⁹ or absent quantitative LTBI test data^{32,33}. A subset of the data set was included in the prediction model for the Roth et al. study³¹; contacts and migrants were excluded owing to no data being available on country of birth or infectiousness of index cases, respectively. Additional study characteristics are shown in Supplementary Table 1.