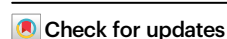


Development and validation of prediction algorithm to identify tuberculosis in two large California health systems

Received: 25 October 2024

Accepted: 31 March 2025

Published online: 10 April 2025



Heidi Fischer¹✉, Lei Qian¹, Zhuoxin Li¹, Katia Bruxvoort^{1,2}, Jacek Skarbinski^{3,4}, Yuching Ni³, Jennifer H. Ku¹, Bruno Lewin^{5,6}, Saadiq Garba⁷, Parag Mahale¹, Sally F. Shaw¹, Brigitte Spence¹ & Sara Y. Tartof^{1,7}

California data demonstrate failures in latent tuberculosis screening to prevent progression to tuberculosis disease. Therefore, we developed a clinical risk prediction model for tuberculosis disease using electronic health records. This study included Kaiser Permanente Southern California and Northern California members ≥ 18 years during 2008–2019. Models used Cox proportional hazards regression, Harrell's C-statistic, and a simulated TB disease outcome accounting for cases prevented by current screening which includes both observed and simulated cases. We compared sensitivity and number-needed-to-screen for model-identified high-risk individuals with current screening. Of 4,032,619 and 4,051,873 Southern and Northern California members, tuberculosis disease incidences were 4.1 and 3.3 cases per 100,000 person-years, respectively. The final model C-statistic was 0.816 (95% simulation interval 0.805–0.824). Model sensitivity screening high-risk individuals was 0.70 (0.68–0.71) and number-needed-to-screen was 662 (646–679) persons-per tuberculosis disease case, compared to a sensitivity of 0.36 (0.34–0.38) and number-needed-to-screen of 1632 (1485–1774) with current screening. Here, we show our predictive model improves tuberculosis screening efficiency in California.

Tuberculosis (TB) disease, caused by *Mycobacterium tuberculosis* (Mtb), results in substantial morbidity and mortality¹. California has the largest absolute number of people with TB disease in the United States (US), with 22% of cases in 2023 occurring in California². As a US state meeting thresholds for the highest level of TB incidence and cases, California has been highlighted by the Centers for Disease Control and Prevention as a state where prevention efforts must be expanded³. More than 2 million Californians are estimated to be living with latent tuberculosis

infection (LTBI), and 87% of TB disease cases in 2021 were attributed to progression from LTBI⁴.

Since most TB disease in the US results from progression from LTBI, domestic TB control is focused on LTBI screening and treatment efforts to prevent progression to TB disease.

Until 2024, the California Department of Public Health (CDPH) recommended screening asymptomatic adults for LTBI based on any of the following risk factors: 1) birth, travel, or residence in a country with high TB incidence ("HTBIC"; definition used by the CDPH that

¹Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, USA. ²Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. ³Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. ⁴Department of Infectious Diseases, Oakland Medical Center, Kaiser Permanente Northern California, Oakland, CA, USA. ⁵Department of Family Medicine, Kaiser Permanente Los Angeles Medical Center, Los Angeles, CA, USA. ⁶Department of Clinical Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, USA. ⁷Department of Health Systems Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, USA. ✉e-mail: Heidi.Fischer@kp.org

includes all individuals born outside of the US, Northern or Western Europe, Canada, or Australia/New Zealand) for >1 month; 2) immunosuppression; and 3) close contact to someone with TB disease^{5,6}. However, guideline implementation has been suboptimal^{6–9}. Screening and treatment guidelines following these risk factors are difficult to implement^{10–12}. Many of these risk factors are not readily identifiable in patients' electronic health records (EHRs), and providers have competing priorities during the limited time of a medical encounter. Our prior study in a large California health system also identified implementation gaps, where LTBI screening practices often missed individuals with the highest risk of LTBI, and only one-fifth of patients with TB disease were screened for LTBI prior to their diagnosis⁸. Using methods which conservatively estimated the total number of TB disease cases possibly prevented due to current LTBI screening and treatment practices added to actual TB cases, 62% of TB disease patients were never screened for LTBI. This indicates a failure to screen those at highest risk of progression to TB disease¹³.

In this work, we examined whether prevention of TB disease through LTBI screening would improve if health systems were able to more successfully implement CDPH guidelines by focusing on the CDPH screening suggestions identifiable in the EHR, or if a different screening strategy would lead to more efficient screening^{14,15}. We compare these strategies by first developing and validating a prognostic multivariable predictive model to identify individuals at highest risk of TB disease in the next 10 years using information available in the EHR. Next, we compare the sensitivity, specificity, and number needed to screen (NNS) of the final predictive model to: 1.) current real-world screening practices and 2.) a screening program that adheres as closely as possible to CDPH recommendations using only information available in the EHR. We find that our newly developed model using readily-available EHR data greatly improves the efficiency of current screening, and can also be a tool for healthcare institutions and providers who require more focused, efficient screening to that currently recommended by the CDPH.

Results

Participants

The KPSC cohort was used as the development and internal validation dataset and included 4,032,619 people, of whom 925 developed TB disease during a median follow-up time of 4.6 years (4.1 per 100,000 person-years). Patients developing TB disease were more likely to be Asian or Hawaiian/Pacific Islander, male, older, born in a HTBIC, have lower BMI, and have more comorbidities (Table 1, Supplementary Table 1).

The KPNC cohort was used as the external validation dataset and had 4,051,873 patients, of which 871 developed TB disease during a median follow-up time of 6.9 years (3.3 per 100,000 person-years). The patient populations for KPSC and KPNC were largely similar on the predictors of interest (Supplementary Table 2), except KPSC had a higher proportion of patients identifying as Hispanic, while KPNC had a higher proportion of patients identifying as Asian and White, and there was a larger proportion of patients with diabetes, higher Charlson scores, and other comorbidities at KPSC. Characteristics of those developing TB disease at KPNC were similar to those identified for KPSC (Table 2). There were 12,566 patients treated for LTBI at KPSC and 13,639 treated at KPNC during the study period, making up 21.6% and 24.1% of 58,167 and 56,491 patients testing positive for LTBI in each system during the study period, respectively.

Model development, specification, and performance

Predictors were ranked in 3 tiers of importance to be included in Cox proportional hazards predictive models, with tier 1 as most important; the most complex model included variables in tiers 1, 2 and 3, while the most parsimonious included only tier 1 variables (Supplementary Table 3). Categorical, linear, and spline forms of relevant tier 1 variables

with continuous forms (BMI, age, Charlson score) were also explored across models. Table 3 compares model performance, measured by Harrell's C-statistic, the time-varying area under the precision (positive predictive value) and recall (sensitivity) curve (time-varying AUPRC) at 10 years following index date, mean calibration, and calibration slope^{16–19}, for all candidate models using the training data, as well as for final models in the internal and external testing datasets. Performance metrics averaged across 5 cross validations in the training dataset were similar for all candidate models, achieving Harrell C-statistics between 0.827 and 0.830, prAUCs between 0.00407 and 0.00470 (compared to a baseline prevalence of 0.00039) and mean calibration and calibration slopes near 1. Since our preference was for the most parsimonious, interpretable model in the event of similar model fit, we chose a model containing only Level 1 variables in their categorical form. Validation on the internal and external datasets showed similar model performance, with Harrell's C statistic of 0.829 and 0.858 and prAUCs of 0.00485 and 0.00294 for KPSC and KPNC, respectively. Calibration metrics varied more in the validation datasets, with mean calibration of 1.10 and a calibration slope of 0.97 for KPSC, indicating slight systematic under prediction of risks. For KPNC, mean calibration was 0.84 and the calibration slope was 1.11, indicating a slight over-prediction of risks, where lower predicted risks may be estimated to be too high, and higher predicted risks are too low (underfitting).

Since current LTBI screening and treatment practices may have resulted in the prevention of TB disease, causing underestimation of effectiveness of current screening, measures of model fit in internal and external test datasets were compared using an additional simulated outcome accounting for TB disease cases possibly prevented due to current LTBI treatment practices ("hypothetical outcome")¹³. Validating the final model using 100 simulations of the hypothetical outcome, the prediction model recorded a median Harrell's C of 0.793 (95% SI 0.770, 0.809) for KPSC and 0.817 (95% SI 0.805, 0.824) for KPNC, and comparable prAUCs or both sites. Mean calibration and calibration slope for KPSC were 1.38 (95% SI 1.29, 1.45) and 0.86 (95% SI 0.81, 0.89), respectively, indicating increased systematic under prediction of risks and a move toward overfitting of low and high risks. Mean calibration and calibration slope for KPNC were 1.10 (95% SI 1.07, 1.13) and 0.94 (95% SI 0.91, 0.96), respectively, indicating a move toward slight systematic underprediction of risks across all risk profiles using the hypothetical outcome.

Hazard ratios (HR) for the final model are presented in Table 4 and showed similar patterns to those observed in univariate data: higher hazards of TB disease were observed for lower BMI, older age, immunosuppression, diabetes, higher Charlson score, birth in a HTBIC, and Asian/Pacific Islander, Hispanic, and Black race/ethnicity.

Cut-points and comparison with current and EHR-based CDPH recommended screening

To compare the final predictive model with current real-world screening and ideal EHR-based CDPH recommendations, 10-year risk cut-points were chosen so that the same percentage of patients would be screened as those that had been screened for LTBI in the year prior to index (5.2% of patients screened), during the full study period (17.7% of patients screened), and those who met criteria for ideal EHR-based CDPH screening (29.7% patients could be hypothetically screened) in the KPSC population in the year prior to index. Supplementary Tables 4 and 5 give detailed instructions for calculating 10-year risk using the final model. Using the hypothetical TB disease outcome, Fig. 1 compares the estimates and 95% SI on the receiver operator curve of actual screening in the year before, actual screening during the study period, and ideal EHR-based CDPH screening (all marked with a circle), with the results that would be expected if a similar portion of the population was screened in each of these categories by selecting the patients with the highest risk scores in the developed prediction model (marked with a triangle). Specificity estimates were

Table 1 | Current screening, ideal electronic health record-based California Department of Public Health screening, and predictors for new prediction model for tuberculosis disease by tuberculosis disease status for Kaiser Permanente Southern California population, *N* = 4,032,619

Characteristic ^a	No TB Disease, <i>N</i> = 4,031,694 ^b	TB Disease, <i>N</i> = 925 ^b
Current Screening in year before index	207,817 (5.2%)	45 (4.9%)
Current Screening in full study period	715,828 (18%)	139 (15%)
Ideal EHR-based CDPH screening	1,199,035 (30%)	695 (75%)
Sex		
Female	2,093,599 (52%)	410 (44%)
Male	1,938,095 (48%)	515 (56%)
Age (years)		
18–<35 years	1,508,564 (37%)	161 (17%)
35–<50 years	1,117,798 (28%)	187 (20%)
50–<65 years	964,189 (24%)	288 (31%)
65–<75 years	292,823 (7.3%)	194 (21%)
75+ years	148,320 (3.7%)	95 (10%)
Body mass index category		
Less than 18.5: underweight	48,840 (1.2%)	29 (3.1%)
18.5–25: healthy weight	1,044,269 (26%)	396 (43%)
25–30: overweight	1,236,753 (31%)	273 (30%)
30–35: moderately obese	730,474 (18%)	109 (12%)
Over 35: severely obese	519,323 (13%)	36 (3.9%)
Missing	452,035 (11%)	82 (8.9%)
Born in a HTBIC (actual and imputed)		
Yes	1,165,980 (29%)	686 (74%)
No	2,865,714 (71%)	239 (26%)
Race/Ethnicity		
White	1,457,624 (36%)	103 (11%)
Asian/Pacific Islander	456,019 (11%)	467 (50%)
Black	326,617 (8.1%)	48 (5.2%)
Hispanic	1,487,866 (37%)	297 (32%)
Other/Multiple/Unknown	303,568 (7.5%)	10 (1.1%)
Diabetes	340,270 (8.4%)	239 (26%)
Immunosuppressed	43,827 (1.1%)	35 (3.8%)
Weighted Charlson Comorbidity Score, continuous	1 (1, 1)	1 (1, 3)
0	3,180,297 (79%)	511 (55%)
1–3	738,726 (18%)	317 (34%)
4+	112,671 (2.8%)	97 (10%)

^aAdditional candidate predictors considered presented in Supplementary Table 1.

^b*n* (%).

CDPH California Department of Public Health, EHR electronic health record, HTBIC High tuberculosis incidence country.

Table 2 | Current screening, ideal electronic health record-based California Department of Public Health screening, and predictors for new prediction model for tuberculosis disease by tuberculosis disease status for Kaiser Permanente Northern California population, *N* = 4,051,873

Characteristic	No TB disease, <i>N</i> = 4,051,002 ^a	TB disease, <i>N</i> = 871 ^a
Screened in year before index	237,290 (5.9%)	29 (3.3%)
Screened during study period	762,060 (19%)	121 (14%)
Ideal EHR-based CDPH screening	1,218,076 (30%)	712 (82%)
Sex		
Female	2,092,561 (52%)	363 (42%)
Male	1,958,441 (48%)	508 (58%)
Age (years)		
18–<35 years	1,509,842 (37%)	181 (21%)
35–<50 years	1,106,953 (27%)	180 (21%)
50–<65 years	962,535 (24%)	275 (32%)
65–<75 years	282,738 (7.0%)	147 (17%)
75+ years	188,934 (4.7%)	88 (10%)
Body mass index category		
Less than 18.5: underweight	46,109 (1.1%)	37 (4.2%)
18.5–<25: healthy weight	1,084,953 (27%)	382 (44%)
25–<30: overweight	1,130,508 (28%)	232 (27%)
30–<35: moderately obese	608,912 (15%)	62 (7.1%)
35+: severely obese	419,428 (10%)	26 (3.0%)
Missing	761,092 (19%)	132 (15%)
Born in a HTBIC (actual and imputed)		
Yes	2,862,167 (71%)	167 (19%)
No	1,188,835 (29%)	704 (81%)
Race/Ethnicity		
White	1,923,077 (47%)	71 (8.2%)
Asian/Pacific Islander	766,971 (19%)	620 (71%)
Black	266,613 (6.6%)	39 (4.5%)
Hispanic	754,447 (19%)	115 (13%)
Other/Multiple/Unknown	339,894 (8.4%)	26 (3.0%)
Diabetes	80,431 (2.0%)	35 (4.0%)
Immunosuppressed	36,370 (0.9%)	27 (3.1%)
Weighted Charlson Score		
0	3,461,947 (85%)	595 (68%)
1–3	520,899 (13%)	229 (26%)
4+	68,156 (1.7%)	47 (5.4%)

HTBIC high tuberculosis incidence country, EHR electronic health records, CDPH California Department of Public Health.

^a*n* (%).

unchanged across simulations so 95% SI appear only for sensitivity, which did vary across simulations. Points in the upper left quadrant of Fig. 1 reflect preferred models with the highest sensitivity and specificity.

In the year before index date, 5.2% of patients at KPSC were screened, giving a sensitivity of 0.06 (95% SI 0.04, 0.08), while 5.9% of patients were screened at KPNC, giving a sensitivity of 0.04 (95% SI 0.03, 0.05). Choosing a 10-year risk cut-point from the final model where approximately 5.2% of individuals are screened in the KPSC training dataset (corresponding to a 10-yr TB disease risk cut-point of 0.15%), the predictive model gave a sensitivity of 0.35 (95% SI 0.32,

0.37) in the KPSC internal test dataset, screening 5.4% of the population, and a sensitivity of 0.38 (95% SI 0.37, 0.39) in the KPNC external test dataset, screening 6.3% of the population. Specificity was 0.94 for all testing strategies.

Looking at screening across the full study period, 17.7% of patients at KPSC were screened, giving a sensitivity of 0.33 (95% SI 0.27, 0.36), while 18.8% of patients were screened at KPNC, giving a sensitivity of 0.36 (95% SI 0.34, 0.38). Choosing a 10-year risk cut-point where approximately 17.7% of individuals are screened in the KPSC training dataset (corresponding to a 10-yr TB disease risk cut-point of 0.05%), the predictive model gave a sensitivity of 0.64 (95% SI 0.61–0.67) in the KPSC internal test dataset, screening 17.9% of the population, and a

Table 3 | Measures of model fit on training, internal test, and external test datasets

Model name	Harrell's C-statistic	prAUC ^b	Mean calibration	Calibration Slope
Training data, averaged across 5 cross validations				
Level 1,2,3 Variables, categorical ^a	0.828	0.00459	1.00	0.97
Level 1,2,3 Variables Continuous, spline	0.829	0.00407	1.00	0.97
Level 1,2,3 Variables Continuous, linear	0.828	0.00415	1.00	0.97
Level 1,2 Variables, categorical	0.829	0.00470	1.00	0.98
Level 1,2 Variables Continuous, spline	0.830	0.00413	1.00	0.98
Level 1,2 Variables Continuous, linear	0.829	0.00423	1.00	0.98
Level 1 Variables, categorical (Final Model)	0.827	0.00451	1.00	0.99
Level 1 Variables Continuous, spline	0.828	0.00412	1.00	0.99
Level 1 Variables Continuous, linear	0.827	0.00426	1.00	0.99
Final model, KPSC internal test				
Actual TB Disease	0.829	0.00485	1.10	0.97
Actual and Hypothetical TB Disease (95% SI)	0.793 (0.770,0.809)	0.0045 (0.0043,0.0051)	1.38 (1.29,1.45)	0.86 (0.81,0.89)
Final model, KPNC external test				
Actual TB Disease	0.858	0.00294	0.84	1.11
Actual and Hypothetical TB Disease (95% SI)	0.817 (0.805,0.824)	0.0028 (0.0027,0.0029)	1.10 (1.07,1.13)	0.94 (0.91,0.96)

Bold fonts indicate change in dataset for model evaluations.
prAUC area under the precision recall curve, KPSC Kaiser Permanente Southern California, KPNC Kaiser Permanente Northern California, 95% SI 95 percent simulation interval.
^aLevel 1 Variables (included in all models) are gender, age, race/ethnicity, BMI, diabetes, Charlson Score, birth in a HTBIC (with imputation for missing values), immunosuppression, Level 2 Variables are end stage renal disease, hepatitis B, hepatitis C, smoking. Level 3 Variables are Neighborhood deprivation index (census), median household income (census), education(census), insurance status, and language. Age, BMI, Charlson Score are modeled continuously. All continuous versions use restricted cubic splines with 3 knots, and are also explored linearly. Measures of fit averaged across 5 cross validations on training data.
^bCompared to an event rate (adjusted for censoring) of 0.00042 and 0.00031 at KPSC and KPNC respectively. Event rates under simulations for actual and hypothetical TB disease ranged from 0.00048 to 0.00053 in the KPSC internal test dataset and from 0.00039 to 0.0004 in KPNC external test dataset.

sensitivity of 0.70 (95% SI 0.68–0.71) in the KPNC external test dataset, screening 20.3% of the population. Specificity was 0.76 for actual screening at both KPSC and KPNC, and 0.81 for the screening proposed using model cut-points in both datasets.

Following ideal EHR-based CDPH screening, 29.8% of patients at KPSC would be screened, giving a sensitivity of 0.73 (95% SI 0.69–0.75), while 30.1% of patients would be screened at KPNC, giving a sensitivity of 0.79 (95% SI 0.77, 0.80). Choosing a 10-year risk cut-point where approximately 29.8% of individuals are screened in the KPSC training dataset (corresponding to a 10-yr TB disease risk cut-point of 0.02%), the predictive model gave a sensitivity of 0.76 (95% SI 0.72–0.79) in the KPSC internal test dataset, screening 29.9% of the population, and a sensitivity of 0.81 (95% SI 0.80, 0.83) in the KPNC external test dataset, screening 30.1% of the population. Specificity was 0.72 and 0.73 for ideal EHR-based CDPH screening at KPSC and KPNC, respectively, and 0.68 and 0.70 for screening proposed using model cut-points in both the internal and external datasets.

Table 5 shows estimates of sensitivity and specificity, along with the NNS (number of patients needed to be screened to find one additional person with TB disease) for each screening strategy under both observed and hypothetical TB disease outcomes. Improvements in NNS show similar patterns, with much lower NNS observed using the model-derived risk cut-points in the year before index date and full study period (361; 95% SI [351, 369] and 662; 95% SI [646, 679], respectively) compared with actual screening (3222; 95% SI [2727, 4026] and 1632; 95% SI [1485, 1774], respectively) at KPNC, and similar patterns observed at KPSC. More comparable NNS was observed using model-derived risk cut-points screening a similar percentage of the population compared with ideal EHR-based CDPH screening.

Discussion

Our newly developed tuberculosis risk prediction model designed to utilize information available from a patient's EHR demonstrated marked improvements in screening performance and efficiency

compared with current screening strategies, increasing sensitivity from 0.33 (95% SI: 0.27, 0.36) and 0.36 (95% SI: 0.34, 0.38) for actual screening during the study period at KPSC and KPNC, respectively, to 0.64 (95% SI: 0.61–0.67) and of 0.70 (95% SI: 0.68–0.71) by screening a similar percentage of people determined to be of highest risk by our newly developed prediction model, with similar gains in NNS. Improvements were even more substantial when comparing screening practices in the year prior to index date but were not as strong when comparing the model-based screening approach to ideal EHR-based CDPH screening. This suggests the model is most valuable in situations when the goal is to screen a smaller percentage of the population more efficiently compared to the wider ideal EHR-based CDPH screening recommendations, which may be challenging for many health systems.

A well-known limitation of the full CDPH screening recommendations is that they do not lend themselves easily to EHR-based best-practice alerts or EHR-based screening guidance as the screening criteria are not consistently documented in EHRs^{6–8} Screening criteria such as travel or residence in a HTBIC country may be asked more frequently by physicians during TB screening among minors in many health systems, but are not part of current practice for adults²⁰. Indeed, the vast majority of people selected for ideal EHR-based CDPH screening as evaluated in the current work would be patients born in HTBICs, since only this and immunosuppression information can be extracted from the EHR, and immunosuppression is rare among the general patient population. Given 27% of individuals in California are born outside the United States primarily in HTBICs¹⁴, focusing broadly on birth in a HTBIC for LTBI screening may not be logistically feasible or be cost prohibitive. Since TB disease progresses from LTBI in about 5–10% of LTBI-positive individuals over their lifetime¹⁵, institutions may instead prefer to focus LTBI screening for individuals with the highest likelihood of progression to TB disease. Our clinical risk prediction model identifies individuals at highest risk for TB disease using readily available EHR data to maximize cost savings and operational efficiency of this proposed LTBI screening approach.

Table 4 | Hazard ratios estimated in final predictive model

Variable	Hazard ratio (95% CI)	
	Fully adjusted (final prediction model)	Unadjusted
Male	1.63 (1.40, 1.90)	1.47 (1.36, 1.59)
Age		
18–<35 years (reference)	1.00	1.00
35–50 years	1.09 (0.85, 1.39)	1.21 (1.07, 1.37)
50–<65 years	1.62 (1.28, 2.04)	1.94 (1.74, 2.17)
65–<75 year	2.93 (2.27, 3.79)	4.15 (3.69, 4.68)
75+ years	3.19 (2.32, 4.38)	3.99 (3.44, 4.64)
Race/Ethnicity		
White (reference)	1.00	1.00
Asian/Pacific Islander	5.53 (4.03, 7.57)	15.01 (13.27, 16.97)
Black	2.34 (1.6, 3.44)	2.03 (1.67, 2.47)
Hispanic	2.38 (1.77, 3.19)	3.18 (2.80, 3.62)
Mult/Other/Unknown	0.72 (0.34, 1.52)	0.87 (0.60, 1.27)
Body Mass Index		
18.5–25: normal weight (reference)	1.00	1.00
Less than 18.5: underweight	1.89 (1.24, 2.90)	1.77 (1.42, 2.20)
25–30: overweight	0.55 (0.46, 0.66)	0.55 (0.51, 0.60)
30–35: moderately obese	0.43 (0.34, 0.56)	0.36 (0.32, 0.41)
Over 35: severely obese	0.25 (0.17, 0.38)	0.17 (0.14, 0.21)
Missing	0.94 (0.71, 1.23)	0.58 (0.51, 0.67)
Charlson score		
Charlson score 0 (reference)	1.00	1.00
Charlson score 1–3	1.67 (1.35, 2.06)	2.39 (2.20, 2.59)
Charlson score 4+	2.3 (1.65, 3.20)	4.69 (4.14, 5.32)
Diabetes	1.48 (1.17, 1.87)	3.33 (3.06, 3.62)
Born in a HTBIC	2.99 (2.37, 3.78)	7.23 (6.65, 7.87)
Immunosuppressed	2.87 (1.99, 4.14)	3.98 (3.32, 4.77)

HTBIC high tuberculosis incidence country.

The ideal percentage of patients to be screened using the newly developed risk prediction equation will be health system dependent. Health systems may, for example, prefer to screen the upper 5% of high-risk individuals each year. Using the model cut-point that screens around 5–6% of model-identified higher risk individuals, we were able to attain similar sensitivity and much lower NNS than the real-world screening that took place during the entire study period, which screened around 18 percent of the population.

It is important to note that current screening practices may show poor performance due to current low-risk screening requirements, including hospital-based requirements for clinicians, school-based policies, or required testing in congregate settings such as nursing homes. However, as of May 17, 2019, routine screening of healthcare workers is no longer recommended by the CDC or CDPH, and this policy was officially adopted by KPNC in 2023 and KPSC in 2024. These changes will likely free up testing resources and provide an opportunity to adapt testing to focus on higher risk individuals^{21,22}.

Calibration measures were not as robust in validation datasets compared to the training datasets, but this did not seem to practically make a difference for screening improvements with our choices of cut-points. It could be the case, however, that our cut-point approach may screen a different percentage of patients in other health systems. Because percentages screened were so similar in the external test

dataset using cut-points chosen in the training dataset, we do not believe this will be a major problem when the model is used on larger health care systems with EHR data in California. Health providers and systems interested in using this risk prediction model could also consider defining their own cut-points more closely aligned with targeted screening percentages in their populations. Given California's large population and burden of TB disease, there are already many opportunities to use this prediction tool to improve LTBI screening efficiency. However, it could provide further benefit being used outside of California, and its validation in geographic regions outside of California is a valuable direction for future research.

A recent TB prediction tool with the goal of focused, efficient TB disease screening was developed using immigration and health administrative data in Canada²³. The prediction tool developed in Canada focused on the population of recent immigrants to Canada, which is one of the highest-risk subgroups for progression to TB disease. In contrast, the model developed in the current manuscript aims to better target LTBI screening among the larger population living in low TB incidence countries, including patients born in a HTBIC who immigrated less recently and other higher risk populations, such as immunosuppressed persons. In recent years, more TB disease diagnoses among foreign born individuals in the U.S. have occurred in the 10 years after arrival in the U.S. than among those in the U.S. <10 years²⁴. As such, effective LTBI screening requires moving the focus beyond using the public health sector to screen recent immigrants to the U.S., to expand partnerships with health care providers. This new risk prediction algorithm is a valuable tool for improving the efficiency of LTBI screening in the healthcare setting using information available in EHRs, which do not have access to information supplied to immigration authorities used in the Canadian model. Thus, we believe the Canadian model and the model developed in this manuscript can work in tandem with the common goal of decreasing TB disease incidence overall.

Study limitations include the inability to limit the TB disease outcome to those cases progressing from LTBI in this population. However, given biases identified in current LTBI testing practices and large identified gaps in LTBI screening among observed TB disease cases^{8,13}, developing this risk prediction model using only the population of patients tested for LTBI would miss important characteristics of our patient population at high risk of progression to TB disease from LTBI, the vast majority of whom were never screened. Though building a risk prediction model among the entire patient population will result in the inclusion of patients with TB disease due to recent exposure, this will account for a small minority of TB disease cases in our patient population given almost 90% of TB disease cases in California were the result of progression from LTBI⁴. Additionally, many characteristics of patients contracting TB disease from recent exposure may have high overlap with patients at highest risk of progression to TB disease from LTBI, such as immunosuppression and certain health conditions. Given evidence that many patients with symptomatic TB disease may delay care²⁵, there may also be benefit in targeting LTBI screening for high-risk individuals contracting TB disease from both these pathways to prevent the additional spread of TB disease.

Finally, though our simulated hypothetical TB disease is a conservative accounting of TB disease cases avoided due to current treatment and screening practices, it is not precise. To more rigorously understand the usefulness of our new predictive model and compare it with ideal EHR-based CDPH screening and current screening practices, a full, rigorous simulation model incorporating a range of inputs for LTBI risk, TB disease progression rates, and TB due to recent exposure by important risk groups could be developed. This work is beyond the scope of the current paper and is an important direction for future research.

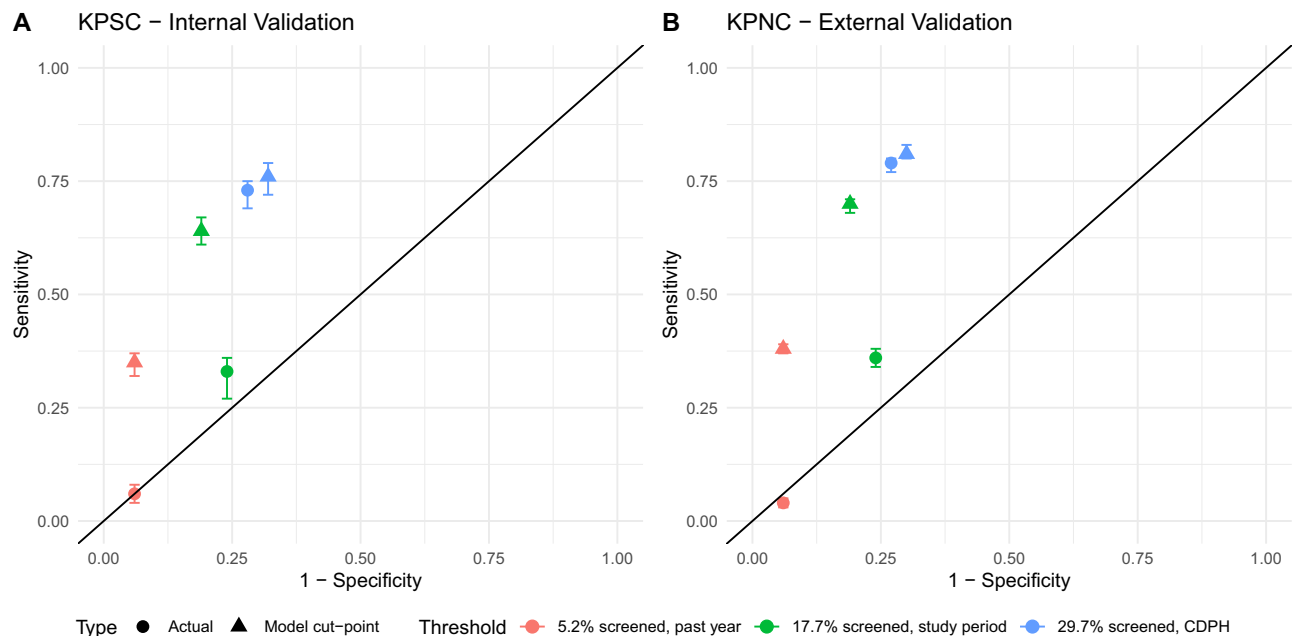


Fig. 1 | Comparisons of performance using model cut-points vs with various screening regimens. Comparison of current screening, ideal electronic health record-based California Department of Public Health hypothetical screening, and model-based screening on the receiver operator curve for hypothetical

tuberculosis outcome using simulation medians with 95 percent simulation intervals across 100 simulations for KPSC (**A**) and KPNC (**B**). KPSC Kaiser Permanente Southern California, KPNC Kaiser Permanente Northern California, CDPH California Department of Public Health.

In conclusion, we have developed and validated a prognostic multivariable predictive model to identify individuals at highest risk of TB disease in the next 10-years. This model can be used to vastly improve performance and efficiency of LTBI screening programs, which are currently poorly adhered to due to myriad challenges in implementation. Our proposed approach uses data readily available in an EHR to create a prioritized list of patients for targeted outreach, directing resources to persons at highest risk for TB disease. It can be used to develop automated prompts for providers to consider LTBI screening during primary care visits, or as part of a larger, targeted external outreach program in health care organizations.

Methods

Study setting

We predicted TB disease in a cohort of patients within Kaiser Permanente Southern California (KPSC) and Kaiser Permanente Northern California (KPNC) between January 1, 2009 and December 31, 2019. KPSC and KPNC are large, integrated health systems serving over 9.2 million racially and socioeconomically diverse members across 489 medical offices^{26,27}. We used the KPSC population as the development dataset and the KPNC dataset for external validation. While both part of Kaiser Permanente, KPSC and KPNC are separate systems, with separate workforces, best practices, governing boards, and data collection and storage processes. Both these differences and their similarities as large healthcare systems in California, make KPSC and KPNC ideal for developing a prediction algorithm for use in California.

Population

The study population consisted of KPSC and KPNC members aged ≥ 18 years with membership for a minimum of 24 consecutive months at any point between January 2008 through December 2019, allowing for a 45-day gap. To allow adequate time to capture information on baseline predictors, we defined the index date as the later of the first day after fulfilling 1-year of the membership criterion, or January 1, 2009. Patients with TB disease prior to the index date were excluded from the study population.

Outcome

We defined our outcome of TB disease by a positive culture or nucleic acid amplification for *Mtb*. Patients were followed from index date until the first positive test for TB disease, end of enrollment allowing for a 45-day gap, death, initiation of any medication for LTBI (isoniazid [INH], rifampin, or rifamycin-INH short-course combinations [INH + rifampin or INH + rifapentine]), or December 31, 2019, whichever came first.

Current LTBI screening and treatment practices may have resulted in the prevention of TB disease, causing underestimations of their effectiveness. To understand how this may impact comparisons between current real-world implementation of screening practices, the final predictive model, and a hypothetical screening program that adheres as closely as possible to CDPH recommendations using only information available in the EHR, measures of model fit in internal and external test datasets were compared using an additional simulated outcome accounting for TB disease cases possibly prevented due to these practices (“hypothetical outcome”)¹³. Specifically, we assumed those testing positive for LTBI who ever filled an LTBI treatment prescription before testing positive for TB disease (“prescribed LTBI treatment”) would have a 2% chance of progression to TB disease during the study period had they not been treated. We estimated this assuming the upper bound of 95 percentile estimates for age-specific lifetime risks of TB progression from non-conversion positive skin tests as described by Horsburgh²⁸, adjusted for maximum follow-up time of 10 years¹³. The hypothetical outcome then combines these hypothetical TB disease cases with observed TB cases during the study period.

Predictors

All predictors were defined in the year prior to index date. They were ranked in 3 tiers of importance (with tier 1 as most important) and classified by whether they might affect initial exposure to TB disease (“exposure”) or progression to TB disease from TB infection (“progression”), or both/unknown. Certain tier 1 variables could be modeled either categorically, continuously linearly, or continuously with a

Table 5 | Sensitivity, specificity, and number needed to screen for various cut-points and screening strategies

	Sensitivity	Specificity	NNS
Internal validation, KPSC			
Actual TB disease			
Actually screened in year before index date (5.2% of the population)	0.05	0.94	2745
Model based screening of 5.2% of the population	0.41	0.94	320
Actually screened during full study period (17.7% of the population)	0.15	0.76	3632
Model based screening of 17.7% of the population	0.71	0.81	611
Ideal EHR-based CDPH screening (29.7% of the population)	0.76	0.72	852
Model based screening of 29.7% of the population	0.82	0.68	917
Actual and hypothetical TB disease with 95% SI			
Actually screened in year before index date (5.2% of the population)	0.06 (0.04, 0.08)	0.94	1806 (1287, 2577)
Model based screening of 5.2% of the population	0.35 (0.32, 0.37)	0.94	300 (282, 315)
Actually screened during full study period (17.7% of the population)	0.33 (0.27, 0.36)	0.76	1369 (1165, 1752)
Model based screening of 17.7% of the population	0.64 (0.61–0.67)	0.81	548 (515, 576)
Ideal EHR-based CDPH screening (29.7% of the population)	0.73 (0.69–0.75)	0.72	709 (662, 762)
Model based screening of 29.7% of the population	0.76 (0.72–0.79)	0.68	788 (728, 839)
External validation, KPNC			
Actual TB disease			
Actually screened in year before index date (5.2% of the population)	0.03	0.94	5580
Model based screening of 5.2% of the population	0.47	0.94	393
Actually screened during full study period (17.7% of the population)	0.14	0.76	5362
Model based screening of 17.7% of the population	0.78	0.81	790
Ideal EHR-based CDPH screening (29.7% of the population)	0.88	0.70	1079
Model based screening of 29.7% of the population	0.82	0.73	1059
Actual and hypothetical TB disease with 95% SI			
Actually screened in year before index date (5.2% of the population)	0.04 (0.03, 0.05)	0.94	3222 (2727, 4026)
Model based screening of 5.2% of the population	0.38 (0.37, 0.39)	0.94	361 (351, 369)
Actually screened during full study period (17.7% of the population)	0.36 (0.34, 0.38)	0.76	1632 (1485, 1774)
Model based screening of 17.7% of the population	0.70 (0.68–0.71)	0.81	662 (646, 679)
Ideal EHR-based CDPH screening (29.7% of the population)	0.79 (0.77, 0.80)	0.73	832 (805, 855)
Model based screening of 29.7% of the population	0.81 (0.80, 0.83)	0.70	877 (850, 901)

Bold fonts indicate change in dataset and/or outcome definition for model evaluations.

KPSC Kaiser Permanente Southern California, KPNC Kaiser Permanente Northern California, 95% SI 95 percent simulation interval, CDPH California Department of Public Health, TB tuberculosis, EHR electronic health record.

restricted cubic spline with 3 knots, so all three functional forms were considered in the variable selection process. All predictors were either already available in the EHR, or available through linkage to information available in the EHR. These included demographic and clinical characteristics, social determinants of health, and known risk factors for LTBI and TB disease including birth in a HTBIC or immunosuppression. Supplementary Table 3 in the Supplementary Appendix describes and defines these predictors in detail. Due to high missingness (>50% in both health systems), birth in a HTBIC was imputed based on a previously published prediction algorithm, which used preferred language, race/ethnicity, and percentage of the population born outside the U.S. at the census level in logistic regression, all of which can also be linked to information available in the EHR²⁹.

To evaluate a screening program adhering as closely as possible to CDPH guidelines using information available in the EHR (“ideal EHR-based CDPH screening”), patients were considered to have met criteria to be ideal EHR-based CDPH screened if they were born in a HTBIC or were immunocompromised at index date. Variables measuring residence or travel for more than one month to a HTBIC were not included as part of the ideal EHR-based CDPH screening definition because they are not accurately captured in all follow-up years nor easily linked with the EHR. Similarly, close contact with TB disease was not included in the ideal EHR-based CDPH screening, as the purpose of CDPH screening criteria for use in health systems is to identify LTBI positive

patients who should be proactively contacted for LTBI screening, and a diagnosis code for recent exposure to TB disease assigned during a medical encounter would not serve this purpose.

To evaluate current screening practices, patients were defined as being previously screened for LTBI if they received a LTBI test, defined as either a tuberculin skin test (TST) or an interferon- γ release assay (IGRA) at least 60 days before a positive test for TB. Due to the slow nature of TB progression, a 60-day period was selected to allow for diagnostic workup of suspected TB disease^{25,30}.

Missing data

Patients missing variables necessary to impute birth in a HTBIC (preferred language or address information necessary to predict percent born outside the US living in patient’s census tract) were excluded from the analyses due to the small percentage of patients missing this information (3.5%). Following previous work, patients missing information on race/ethnicity were grouped with those that have Other or Multiple race/ethnicities, while those missing body mass index (BMI) and smoking status were classified into other/unknown groups^{8,13,29}.

Statistical analysis

The KPSC dataset was split into training and testing components in an 80/20 split. To maximize interpretability and facilitate implementation in the clinical setting, we derived a prediction model using Cox

proportional hazards regression. The relationships between predictors and outcomes, as well as selection of final model predictors and estimates for parameter values were assessed using only the training dataset at KPSC and variable selection was performed using 5-fold cross validation. Variables were removed from the model according to 3 tiers defined in the predictor section, with model fit evaluated using variables in tiers 1–3 first, then tiers 1–2, and then only tier 1.

We assessed and compared model performance by evaluating the following 2 metrics for each model candidate on the KPSC training dataset: Harrell's C-statistic and the time-varying area under the precision (positive predictive value) and recall (sensitivity) curve (time-varying AUPRC) at 10 years following index date^{16,17}. Random performance for Harrell's C was indicated by a score of 0.5. Random performance for AUPRC is equal to sample prevalence of the outcome of interest, with scores higher than the sample prevalence (up to a score of 1) indicating superior model performance.

We additionally assessed calibration by considering mean calibration (exponentiated calibration-in-the-large) and calibration slope, with values near one for both indicating a well-calibrated model¹⁹. Mean calibration, or exponentiated calibration-in-the-large, is a calibration measure approximating the ratio of observed events in the new data set to the number predicted by the risk model; models of this form have long been used in epidemiology to compare observed to expected numbers of events and are referred to as standardized incidence ratios¹⁸. Mean calibration less than one indicates systematic underestimation of actual risks and calibration slope estimates less than one indicate that lower predicted risks are too low compared with actual risks, while higher predicted risks are too high (overfitting). Harrell's C, AUPRC, mean calibration and calibration slope were jointly considered as primary metrics. Once a final predictive model was chosen using the primary metrics, we explored the efficacy of various cut-points using 10-year risk calculated from the final prediction model on the KPSC training dataset. To allow comparison of the final predictive model with current real-world screening and ideal EHR-based CDPH screening, cut-points were chosen so that the same percentage of patients would be screened as those that had been screened for LTBI in the year prior to index (5.2% of the population screened), during the full study period (17.7% of the population screened), and those meeting criteria for ideal EHR-based CDPH screening (29.7% would be hypothetically screened) in the KPSC population. The percentage of patients screened in these three categories at KPNC during the study period was similar. We then compared sensitivity, specificity and number needed to screen (NNS, defined as number needed to be screened to find an additional person TB disease, the inverse of positive predictive value) measures of these three testing strategies, accounting for censoring³¹.

Once the final model and cut-points were chosen, performance was evaluated using the KPSC test data (internal validation), and then the KPNC data (external validation) using both observed TB disease cases, and the hypothetical outcome which additionally accounts for TB disease cases possible prevented due to current screening practices. Measures of fit using the hypothetical outcome are summarized over 100 simulations using the simulation median and 95% simulation intervals (95% SI). Predicted probabilities for these datasets were calculated using methods described in Supplementary Tables S3 and 4.

The study adhered to the updated TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) + AI guidelines for reporting³². All models were developed by using the R software version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria)³³.

This study was approved by the KPSC Institutional Review Board, with a waiver of informed consent, as this data-only study posed minimal risk to study participants. A study protocol was not prepared. There was no patient or public involvement in the design, conduct, reporting, interpretation, or dissemination of this study.

Reporting summary

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

Data availability

The raw data used to develop models in this manuscript are protected and are not available due to data privacy laws. Data contain potentially identifiable information (e.g. dates of diagnoses and exams) that cannot be shared openly without appropriate human subjects approval and data use agreements. The Kaiser Permanente Southern California institutional policy requires a data transfer agreement be executed between KPSC and the individual recipient entity prior to transmittal of patient-level data outside KPSC. This is a legal requirement. Requests for data can be addressed to the Central Business Office of the Department of Research and Evaluation (contact via Diane.C.Robinson@kp.org).

Code availability

Code for analyses conducted in this work is available as Supplementary Code 1.

References

1. Tuberculosis (TB): Clinical Overview of Tuberculosis. <https://www.cdc.gov/tb/hcp/clinical-overview/> (2025). Accessed April 8, 2025.
2. Williams, P. M. et al. Tuberculosis United States, 2023. *MMWR Morb. Mortal. Wkly Rep.* **73**, 265–270 (2024).
3. Reported Tuberculosis in the United States, 2021: State and Local Data. https://www.cdc.gov/tb/statistics/reports/2021/state_local_data.htm (2021). Accessed September 18, 2023.
4. California Department of Public Health Tuberculosis Control Branch. TB in California: 2021 Snapshot. <https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/TBCB-TB-Snapshot-2021.pdf> (2022). Accessed November 11, 2023.
5. Latent Tuberculosis Infection Guidance for Preventing Tuberculosis in California, CDPH CTCA Joint Guideline, 2018 Revision; California Department of Public Health. <https://ctca.org/wp-content/uploads/2018/12/LTBI-Guidelines-2018-Revision-FINAL.pdf#page=6> (2018). Accessed September 8, 2021.
6. Lewinsohn, D. M. et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clin. Infect. Dis.* **64**, 111–115 (2017).
7. Bibbins-Domingo, K. et al. Screening for Latent Tuberculosis Infection in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* **316**, 962–969 (2016).
8. Ku, J. H. et al. Latent tuberculosis infection testing practices in a large US integrated healthcare system. *Clin. Infect. Dis.* **78**(5), 1304–1312 (2024).
9. *Report on Latent Tuberculosis Infection in California, 2022* (California Department of Public Health, Tuberculosis Branch, 2022).
10. LoBue, P. A. & Mermin, J. H. Latent tuberculosis infection: the final frontier of tuberculosis elimination in the USA. *Lancet Infect. Dis.* **17**, e327–e333 (2017).
11. Shah, M. & Dorman, S. E. Latent Tuberculosis Infection. *N. Engl. J. Med.* **385**, 2271–2280 (2021).
12. Goletti, D., Delogu, G., Matteelli, A. & Migliori, G. B. The role of IGRA in the diagnosis of tuberculosis infection, differentiating from active tuberculosis, and decision making for initiating treatment or preventive therapy of tuberculosis infection. *Int. J. Infect. Dis.* **124**, S12–S19 (2022).
13. Fischer, H. et al., editors. *Prior Screening for Latent Tuberculosis Among Patients Diagnosed With Tuberculosis Disease: Missed Opportunities? Open Forum Infectious Diseases* (Oxford University Press US, 2023).

14. Johnson, H. & Sanchez, S. Just the Facts: Immigrants in California, <https://www.ppic.org/publication/immigrants-in-california/> (2019). Accessed August 20, 2021.
15. Tuberculosis (TB): TB Risk Factors. <https://www.cdc.gov/tb/risk-factors/> (2025). Accessed April 8, 2025.
16. Harrell, F. E. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis* (Springer, 2015).
17. Yuan, Y. et al. A threshold-free summary index of prediction accuracy for censored time to event data. *Stat. Med.* **37**, 1671–1681 (2018).
18. Crowson, C. S., Atkinson, E. J. & Therneau, T. M. Assessing calibration of prognostic risk scores. *Stat. Methods Med. Res.* **25**, 1692–1706 (2016).
19. McLernon, D. J. et al. Assessing Performance and Clinical Usefulness in Prediction Models With Survival Outcomes: Practical Guidance for Cox Proportional Hazards Models. *Ann. Intern. Med.* **176**, 105–114 (2023).
20. Finnell, S. M., Stanton, J. L. & Downs, S. M. Actionable recommendations in the Bright Futures child health supervision guidelines. *Appl. Clin. Inf.* **5**, 651–659 (2014).
21. Updated Centers for Disease Control Tuberculosis Screening Recommendations for Health Care Personnel (HCP) and Nationwide Shortage of Tuberculin Skin Test Antigens. <https://www.cdph.ca.gov/Programs/CHCQ/LCP/Pages/AFL-19-28.aspx> (2021).
22. *Latent tuberculosis infection: updated and consolidated guidelines for programmatic management* (World Health Organization; 2018). Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications/i/item/9789241550239>.
23. Puyat, J. H. et al. Predicting risk of tuberculosis disease in people migrating to a low-TB incidence country: development and validation of a multivariable dynamic risk prediction model using health administrative data. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciae561> (2024).
24. Tsang, C. A. L. A., Navin, T. R. & Armstrong, L. R. Tuberculosis Among Foreign-Born Persons Diagnosed ≥ 10 Years After Arrival in the United States, 2010–2015. *MMWR Morb. Mortal. Wkly Rep.* **66**, 295–298 (2017).
25. Asch, S., Leake, B., Anderson, R. & Gelberg, L. Why do symptomatic patients delay obtaining care for tuberculosis? *Am. J. Respir. Crit. Care Med.* **157**, 1244–1248 (1998).
26. Koebnick, C. et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. *Perm. J.* **16**, 37–41 (2012).
27. Gordon, N. P. Similarity of the adult Kaiser Permanente membership in Northern California to the insured and general population in Northern California: statistics from the 2011 California Health Interview Survey. https://divisionofresearch.kaiserpermanente.org/projects/memberhealthsurvey/SiteCollectionDocuments/chis_non_kp_2011.pdf (2015). Accessed September 10, 2019.
28. Horsburgh, Jr. C. R. Priorities for the treatment of latent tuberculosis infection in the United States. *N. Engl. J. Med.* **350**, 2060–2067 (2004).
29. Fischer, H. et al. Development and validation of a prediction algorithm to identify birth in countries with high tuberculosis incidence in two large California health systems. *PLoS One* **17**, e0273363 (2022).
30. Sarmiento, K., Hirsch-Moverman, Y., Colson, P. & El-Sadr, W. Help-seeking behavior of marginalized groups: a study of TB patients in Harlem, New York. *Int. J. Tuberculosis Lung Dis.* **10**, 1140–1145 (2006).
31. Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).
32. Collins, G. S. et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385.q902.
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (2021).

Acknowledgements

This study was funded by the National Institute of Health (grant number: 5R01 AI151072). The authors thank the patients of Kaiser Permanente Southern California and Kaiser Permanente Northern California for helping to improve care through information collected through our electronic health record systems.

Author contributions

Authors H.F., L.Q., J.S., S.G., K.B., B.L., and S.Y.T. worked on concept and design. Authors H.F., L.Q., Z.L., J.S., K.B., B.L., S.M., J.H.K., S.Y.T., and Y.N. worked on acquisition, analysis, and/or interpretation of data. Author H.F. worked on the drafting of the manuscript. Authors H.F., L.Q., Z.L., J.S., S.F.S., K.B., J.H.K., B.L., P.M., and S.Y.T. working on critical revisions of the manuscript for important intellectual content. Authors H.F., Z.L., L.Q., and Y.N. worked on statistical analyses. Author S.Y.T. obtained funding. Authors S.F.S. and B.S. worked on administrative, technical, or material support. Author S.Y.T. worked on supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58775-6>.

Correspondence and requests for materials should be addressed to Heidi Fischer.

Peer review information *Nature Communications* thanks Rishi Gupta, who co-reviewed with James Greenan-Barrett, and Christine Ho for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025