

Area-Based Sociodemographic Factors Associated with Latent Tuberculosis Infection in a Low-Prevalence Setting

Jeffrey I. Campbell,^{1,2*} Mary Tabatneck,³ Grete E. Wilt,⁴ Mingwei Sun,⁵ Wei He,⁶ Nicholas Musinguzi,⁷ Bethany Hedt-Gauthier,⁸ Gabriella S. Lamb,⁹ Don Goldmann,⁹ Vishakha Sabharwal,^{1,2} Thomas J. Sandora,⁹ and Jessica E. Haberer¹⁰

¹Section of Pediatric Infectious Diseases, Boston Medical Center, Boston, Massachusetts; ²Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; ³Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts; ⁴Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; ⁵Center for Research Information Technology, Boston Children's Hospital, Boston, Massachusetts; ⁶Center for Research Information Science and Computing, Massachusetts General Hospital, Boston, Massachusetts; ⁷Global Health Collaborative, Mbarara University of Science and Technology, Mbarara, Uganda; ⁸Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts; ⁹Division of Infectious Diseases, Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts; ¹⁰Center for Global Health, Massachusetts General Hospital, Boston, Massachusetts

Abstract. Area-based sociodemographic markers, such as census tract foreign-born population, have been used to identify individuals and communities with a high risk for tuberculosis (TB) infection in the United States. However, these markers have not been evaluated as independent risk factors for TB infection in children. We evaluated associations between census tract poverty, crowding, foreign-born population, and the CDC's Social Vulnerability Index (CDC-SVI) ranking and TB infection in a population of children tested for TB infection in Boston, Massachusetts. After adjustment for age, crowding, and foreign-born percentage, increasing census tract poverty was associated with increased odds of TB infection (adjusted odds ratio [aOR] per 10% increase in population proportion living in poverty: 1.20 [95% CI, 1.04–1.40]; $P = 0.01$), although this association was attenuated after further adjustment for preferred language. In separate models, increasing CDC-SVI ranking was associated with increased odds of TB infection, including after adjustment for age and language preference (aOR per 10-point increase in CDC-SVI rank: 1.08 [95% CI, 1.02–1.15]; $P = 0.01$). Our findings suggest area-based sociodemographic factors may be valuable for characterizing TB infection risk and defining the social ecology of pediatric TB infection in low-burden settings.

More than 1 million children and adolescents in the United States are estimated to have latent tuberculosis (TB) infection.¹ For the most part, efforts to identify pediatric TB infection in the United States rely upon contact investigations and routine screening for TB risk factors,² although ongoing incident pediatric TB disease in the United States highlights the limitations of these approaches. Researchers and public health authorities are increasingly interested in using administrative data that can be extrapolated from the electronic health record (EHR) to identify individuals at risk for TB infection.³ In particular, area-based markers of risk for TB infection may bypass some challenges with self-reported risk factors, which may be stigmatizing or purposefully omitted (e.g., immigration history)⁴ and subject to social desirability bias (e.g., language preference).⁵ Prior TB epidemiologic studies have used the proportion of foreign-born individuals in a census tract as a proxy for foreign-born status^{3,6} and have found pediatric TB disease prevalence to be associated with area poverty and demographic composition.⁷ However, the association between area-based sociodemographic factors and pediatric TB infection risk in the United States has not been vetted. In this study, we aimed to determine the relationship between area-based sociodemographic factors and the likelihood of TB infection in a population of children tested for TB infection in an urban, low-burden setting.

We conducted a cross-sectional study nested within a retrospective cohort study of children < 18 years old tested for TB using a tuberculin skin test (TST) or interferon gamma release assay (IGRA) between January 2017 and May 2019

in Boston, Massachusetts.⁸ We excluded confirmatory positive tests, tests associated with addresses outside Massachusetts, and tests for which area-based sociodemographic data were not available.

The primary outcome was TB infection diagnosis after a positive TB infection test (TST or IGRA) per clinician documentation. Tests that treating clinicians determined to be falsely positive were classified as negative. Patients with positive tests who did not complete the diagnostic steps to establish a TB infection diagnosis (i.e., chest radiograph or physical exam) were considered to have TB infection. When patients had multiple positive tests, only the first positive was included. We excluded TSTs that were not read and IGRAs that were borderline/invalid/indeterminate, as well as patients diagnosed with mycobacterial disease.

We selected three census tract-level exposures: 1) the estimated percentage of individuals living below the federal poverty line⁹; 2) the estimated percentage of individuals living in crowded households (i.e., more people than rooms in a household)¹⁰; and 3) the estimated percentage of the population who were foreign born.^{11,12} In a secondary analysis, we assessed the relationship between TB infection prevalence and the CDC's Social Vulnerability Index (CDC-SVI) statewide census tract ranking.¹³ Notably, the CDC-SVI includes poverty and crowding; the CDC-SVI also uses non-English language preference as a proxy for foreign-born population. We used 2018 5-year American Community Survey data to provide stable estimates of area-based exposures. Prespecified covariates were age and language preference,^{11,14} which were recorded in the EHR.

We used descriptive statistics to summarize exposures by outcome. We assessed pairwise correlation between poverty, crowding, and foreign-born percentage using Pearson correlations; we reviewed variables were considered to be highly correlated if the pairwise correlation coefficient was ≥ 0.7 .¹⁵ We

*Address correspondence to Jeffrey I. Campbell, Section of Pediatric Infectious Diseases, Boston Medical Center, 670 Albany St., 6th Fl., Boston, MA 02115-5724. E-mail: jeffrey.campbell@bmc.org

created univariable mixed-effects logistic regression models to account for clustering at the census tract (to account for the clustering of TB infection cases within small areas, e.g., due to local transmission within households or close communities) and individual levels (because individuals could receive multiple tests). We then constructed two sets of multivariable models. Model 1 included age and all sociodemographic factors but excluded language (given the limitations of measuring language in EHRs^{5,16}). Model 2 included language as well as variables from model 1. We repeated this modeling sequence for CDC-SVI ranking.

In these models, sociodemographic variables are represented as 10% increases in the census tract population with the characteristic. We performed a sensitivity analysis in which census tracts were ranked against all census tracts in Massachusetts by poverty, crowding, and foreign-born population; census tract rank was then used as the exposure variable in models.

Address geocoding was conducted using ArcGIS Pro version 10.3 (Esri, Redlands, CA). Statistical analysis was conducted in Stata version 17 (StataCorp, College Station, TX). This study received ethics approval at Boston Children's Hospital and Mass General Brigham.

As shown in Figure 1, 13,353 TB tests were obtained, of which 11,190 tests among 10,090 patients were included; 190 (1.7%) were positive and 11,000 were negative. Table 1 summarizes patient characteristics and Table 2 shows the distribution of TB infection diagnoses by area-based marker, here presented by quartile ranking of an individual's census tract among all census tracts in Massachusetts. Patients resided in 1,123 of Massachusetts's 1,478 census tracts. To illustrate the distribution of sociodemographic factors across the state, Supplemental Figure 1 shows census tracts in which patients resided, colored by poverty, foreign-born population, crowding, and SVI. Table 2 shows the percentage of the census tract population with each sociodemographic characteristic by quartile. The pairwise correlation coefficients between poverty, crowding, and foreign-born percentage were < 0.7.

In the univariable analyses, Spanish language and other non-English language preference and census tract poverty, crowding, and foreign-born percentage were associated with increased odds of TB infection (Table 1). In multivariable

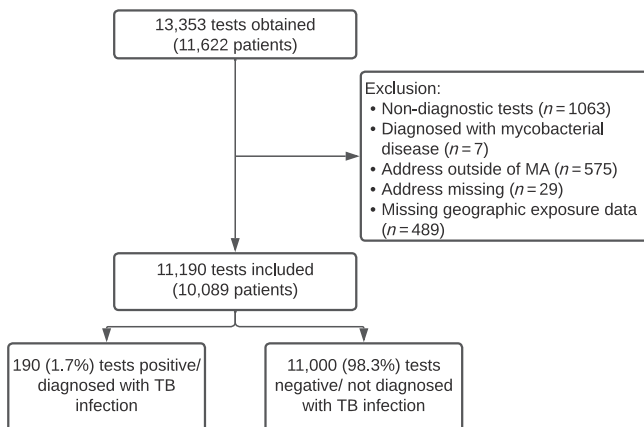


FIGURE 1. Tests obtained and diagnoses of TB infection in the study population. MA = Massachusetts; TB = tuberculosis.

TABLE 1
Characteristics of included patients and univariable and multivariable analyses of predictors of TB infection

Variable (N = 11,190)	TB infection (N = 190)	No TB infection (N = 11,000)	Univariable models OR (95% CI)	Multivariable model 1* aOR (95% CI)	Multivariable model 2† aOR (95% CI)	Multivariable model 3‡ aOR (95% CI)
Age (years)						
< 5	38 (1.4%)	2,744 (98.6%)	REF	REF	REF	REF
5–11	64 (1.8%)	3,464 (98.2%)	1.40 (0.86–2.29)	1.41 (0.88–2.25)	1.31 (0.82–2.09)	1.31 (0.81–2.11)
12–17	88 (1.8%)	4,792 (98.2%)	1.46 (0.89–2.39)	1.54 (0.96–2.48)	1.64 (1.01–2.67)§	1.65 (1.01–2.68)§
Language¶						
English	78 (1.1%)	6,777 (98.9%)	REF	–	REF	REF
Spanish	82 (2.5%)	3,145 (97.5%)	2.53 (1.70–3.76)	–	2.23 (1.48–3.36)	2.11 (1.39–3.21)
Other	30 (2.8%)	1,047 (97.2%)	2.79 (1.64–4.77)	–	2.80 (1.60–4.90)	2.82 (1.61–4.92)
Percentage living in poverty, median (IQR)	16.8% (10–21.7%)	13.9% (6.2–20.1%)	1.27# (1.10–1.45)	1.20# (1.04–1.40)§	1.13# (0.97–1.32)	–
Percentage living in crowded quarters, median (IQR)	3.7% (1.8–5.9%)	2.8% (1.0–5.4%)	1.47# (1.09–1.97)§	0.90# (0.54–1.51)	0.81# (0.47–1.39)	–
Percentage foreign born, median (IQR)	31.2% (22.6–41.7%)	28.5% (16.6–38.5%)	1.20# (1.06–1.36)	1.18# (0.97–1.43)	1.11# (0.92–1.35)	–
SVI census tract rank, median (IQR)	69 (47–90)	62 (28–95)	1.13** (1.06–1.20)	–	–	1.08** (1.02–1.15)

aOR = adjusted odds ratio; IQR = interquartile range; SVI = Social Vulnerability Index; TB = tuberculosis.
 * Includes age, poverty, crowding, and foreign-born percentage.
 † Includes age, poverty, crowding, foreign-born percentage, and language preference.
 ‡ Includes age, language preference, and CDC-SVI rank.
 § P < 0.05.
 || P < 0.01.
 ¶ Excludes 31 patients with missing data.
 # Per 10% increase in census tract population with characteristic.
 ** Per 10-point increase in CDC-SVI rank (increasing rank corresponds to increasing social vulnerability).

TABLE 2
Distribution of total, positive, and negative tests by area-based sociodemographic variable

Characteristic	No TB infection (N = 11,000)	TB infection (N = 190)	Total (N = 11,190)
Poverty quartile			
First (least impoverished)	3,098 (18.7%)	23 (0.7%)	3,121 (27.9%)
Second	2,012 (18.0%)	40 (2.0%)	2,052 (18.3%)
Third	3,442 (30.8%)	79 (2.2%)	3,521 (31.5%)
Fourth (most impoverished)	2,448 (21.9%)	48 (1.9%)	2,496 (22.3%)
Crowding quartile			
First (lowest proportion in crowded quarters)	2,177 (19.5%)	22 (1.0%)	2,199 (19.7%)
Second	2,041 (18.2%)	35 (1.7%)	2,076 (18.6%)
Third	2,921 (26.1%)	51 (1.7%)	2,972 (26.6%)
Fourth (highest proportion in crowded quarters)	3,861 (34.5%)	82 (2.1%)	3,943 (35.2%)
Foreign-born quartile			
First (lowest proportion foreign born)	637 (5.7%)	7 (1.1%)	644 (5.8%)
Second	1,257 (11.2%)	16 (1.3%)	1,273 (11.4%)
Third	2,182 (19.5%)	27 (1.2%)	2,209 (19.7%)
Fourth (highest proportion foreign born)	6,924 (61.9%)	140 (2.0%)	7,064 (63.1%)
SVI rank			
First (lowest quartile of SVI census tracts)	2,414 (21.6%)	19 (0.8%)	2,433 (21.8%)
Second	2,243 (20.0%)	32 (1.4%)	2,275 (20.3%)
Third	2,760 (24.7%)	68 (2.4%)	2,828 (25.3%)
Fourth (highest quartile of SVI census tracts)	3,583 (32.0%)	71 (1.9%)	3,654 (32.7%)

SVI = Social Vulnerability Index; TB = tuberculosis.

analysis that included age and all area-based variables (model 1), an increasing percentage of living in poverty was significantly associated with TB infection diagnosis. In the analysis that also included language (model 2), Spanish language and other language were significantly associated with increased odds of TB infection, as was older age, but not any area-based sociodemographic factors. In the secondary analysis, increasing SVI ranking was significantly associated with increased odds of TB infection when adjusting for age (adjusted odds ratio [aOR] for a 1-point increase in CDC-SVI rank: 1.13 [95% CI, 1.07–1.20]; $P < 0.001$) and age plus language (aOR for a 10-point increase in CDC-SVI rank: 1.08 [95% CI, 1.02–1.15]; $P = 0.01$) (model 3). Effect sizes and significance were largely unchanged in the sensitivity analysis in which area-based sociodemographic variables were represented as census tract statewide rank rather than as the proportion of the population (results not shown).

In this cross-sectional study of children tested for TB in Massachusetts, census tract poverty was associated with increased odds of TB infection diagnosis after adjusting for other area-based sociodemographic markers and age, but this association did not persist after adjusting for preferred language. Increasing CDC-SVI was associated with increased odds of TB infection in models that included age and language. Taken together, these findings suggest that area-based demographic factors may have a role in identifying children at increased risk for TB infection, particularly when language cannot be determined.

Our analysis adds evidence that area-based poverty is an important factor for understanding the distribution of TB in the United States. County-level poverty was found to be associated with countywide TB disease incidence in the continental United States,⁹ and census tract poverty has been associated with pediatric TB disease prevalence in California.⁷ Neighborhood poverty was also associated with the identification of a source case among children with TB infection in New York.¹⁷ In our study, poverty likely reflected other processes associated with TB transmission. For instance, census tract poverty may reflect immigrant populations from

high-TB burden countries or countries associated with the most TB cases in the United States¹⁸ more effectively than census tract foreign-born percentage. Additionally, marked disparities by race and ethnicity have been documented among U.S.-born individuals with TB infection.¹⁹ Area-based poverty may be a proxy for these disparities in our analysis.

Foreign-born percentage and crowding were not associated with TB infection diagnosis in multivariable models. TB epidemiologic studies from North Carolina and California used neighborhood foreign-born population proportion as a proxy for patients' foreign birth status.^{3,6} In the study from California, census tract foreign-born population was significantly associated with positive TB infection screening among adults, including in models that contained language preference. Despite the lack of association in our study, the high absolute numbers of patients with TB infection living in census tracts with high crowding and foreign-born status suggest that these factors remain important proxies for TB infection risk. Future studies using population prevalence or incidence of TB infection as an outcome measure might reflect the effect of these variables. Such an approach has been applied to study ecological associations with TB disease incidence in California and Washington State, which have found similar relationships between area-based sociodemographic markers (including poverty, crowding, and education) and reported TB incidence.^{20,21}

Inclusion of preferred language in multivariable models attenuated the association between poverty and TB infection diagnosis. Non-English language preference has been used as a marker for foreign birth and thus TB infection risk.^{3,14} However, although highly specific, this variable may lack sensitivity for foreign-born status.²² Moreover, language preference may not always be reported in the EHR or may be subject to social desirability biases. Our findings suggest that area-based markers may have utility for identifying TB infection risk when language cannot be reliably measured, for example as components of future EHR-based risk scores that should be the subject of future research. Our data also support the combined use of individual- and community-level

data to understand TB prevalence, as others have used to describe TB disease incidence,²³ although further investigation is needed to identify specific combinations that could be used for risk prediction.

Social vulnerability, as measured by the CDC-SVI, was associated with increased TB infection risk. To our knowledge, our study is the first to examine the association between CDC-SVI and pediatric TB infection, although prior literature has evaluated other composite area-based indices to examine associations between community vulnerability and TB prevalence.²⁰ Our findings reinforce a consistent conclusion across studies that TB infection remains a disease of socially vulnerable communities in the United States and also suggest that those communities most affected by TB also face social and structural barriers to care.

Our study has limitations. Most patients lived in a single urban area, hampering generalizability. Our findings may not reflect recent TB trends in the context of COVID-19. Analysis of area-based factors may incorrectly assign risk based on census tract attributes that do not reflect a child's immediate lived ecosystem, introducing potential ecological fallacies. Because census tract variables were measured cross-sectionally, our study was not designed to assess causal links between sociodemographic characteristics (e.g., determine whether increasing census tract foreign-born status led to increasing census tract poverty or crowding); future mediation analysis using directly measured social drivers of health variables (e.g., surveys) could yield important insights into how social drivers lead to TB infection in children in the United States. We were unable to geocode or measure one or more sociodemographic factors for 4% of tests. Finally, we were unable to account for testing indication or adherence to targeted testing practices, which represents unmeasured confounding in our study.

In conclusion, census tract poverty and social vulnerability are area-based sociodemographic factors associated with pediatric TB infection diagnosis in a low-prevalence setting. The preferred language documented in the EHR was also associated with TB infection and attenuated the association with poverty but not with social vulnerability. Our findings suggest that area-based markers merit additional investigation in prevalence studies to understand the ecosystem of pediatric TB in the United States and hold promise as candidate variables for prediction scores to identify children at risk for TB infection. Specifically, research is needed to understand relationships between ecological sociodemographic data and TB infection prevalence in different settings in the United States. Studies should investigate how individual- and area-level variables can be incorporated into predictive risk scores to identify those at highest risk for TB infection.

Received December 20, 2022. Accepted for publication June 8, 2023.

Published online August 14, 2023.

Note: Supplemental material appears at www.ajtmh.org.

Funding support: J. I. C. was supported by Agency for Healthcare Research and Quality grant number T32 HS000063 as part of the Harvard-wide Pediatric Health Services Research Fellowship Program. J. E. H. was supported by K24MH114732.

Authors' addresses: Jeffrey I. Campbell and Vishakha Sabharwal, Section of Pediatric Infectious Diseases, Boston Medical Center, Boston, MA, and Boston University Chobanian & Avedisian School of Medicine, Boston, MA. E-mails: jeffrey.campbell@bmc.org and

Vishakha.Sabharwal@bmc.org. Mary Tabatneck, Department of Pediatrics, Boston Children's Hospital, Boston, MA, E-mail: Mary.Tabatneck@childrens.harvard.edu. Grete E. Wilt, Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA, E-mail: gwilt@fas.harvard.edu. Mingwei Sun, Center for Research Information Technology, Boston Children's Hospital, Boston, MA, E-mail: sunmingwei09@gmail.com. Wei He, Center for Research Information Science and Computing, Massachusetts General Hospital, Boston, MA, E-mail: WHE3@mgh.harvard.edu. Nicholas Musunguzi, Global Health Collaborative, Mbarara University of Science and Technology, Mbarara, Uganda, E-mail: nmusunguzi@gmail.com. Bethany Hedt-Gauthier, Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, E-mail: bethhedt@gmail.com. Gabriella S. Lamb, Don Goldmann, and Thomas J. Sandora, Division of Infectious Diseases, Department of Pediatrics, Boston Children's Hospital, Boston, MA, E-mails: gabriella.sophia.lamb@gmail.com, Don.Goldmann@childrens.harvard.edu, and Thomas.Sandora@childrens.harvard.edu. Jessica E. Haberer, Center for Global Health, Massachusetts General Hospital, Boston, MA, E-mail: jhaberer@partners.org.

REFERENCES

- Mancuso JD, Diffenderfer JM, Ghassemieh BJ, Home DJ, Kao TC, 2016. The prevalence of latent tuberculosis infection in the United States. *Am J Respir Crit Care Med* 194: 501–509.
- Kimberlin DW, Barnett ED, Lynfield R, Sawyer MH, eds., 2021. *Red Book: 2021–2024 Report of the Committee on Infectious Diseases*, 32nd edition. Itasca, IL: American Academy of Pediatrics.
- Fischer H et al., 2022. 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.
- Kim G, Molina US, Saadi A, 2019. Should immigration status information be included in a patient's health record? *AMA J Ethics* 21: E8–E16.
- Klinger EV, Carlini SV, Gonzalez I, St Hubert S, Linder JA, Rigotti NA, Kontos EZ, Park ER, Marinacci LX, Haas JS, 2015. Accuracy of race, ethnicity, and language preference in an electronic health record. *J Gen Intern Med* 30: 719–723.
- Bonnewell JP, Farrow L, Dicks KV, Cox GM, Stout JE, 2020. Geographic analysis of latent tuberculosis screening: a health system approach. *PLoS One* 15: e0242055.
- Myers WP, Westenhouse JL, Flood J, Riley LW, 2006. An ecological study of tuberculosis transmission in California. *Am J Public Health* 96: 685–690.
- Campbell JI et al., 2023. Multicenter analysis of attrition from the pediatric TB infection care cascade in Boston. *J Pediatr* 253: 181–188.
- Mollalo A, Mao L, Rashidi P, Glass GE, 2019. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. *Int J Environ Res Public Health* 16: 157.
- Baker M, Das D, Venugopal K, Howden-Chapman P, 2008. Tuberculosis associated with household crowding in a developed country. *J Epidemiol Community Health* 62: 715–721.
- Mirzazadeh A, Kahn JG, Haddad MB, Hill AN, Marks SM, Readhead A, Barry PM, Flood J, Mermin JH, Shete PB, 2021. State-level prevalence estimates of latent tuberculosis infection in the United States by medical risk factors, demographic characteristics and nativity. *PLoS One* 16: e0249012.
- Haddad MB, Raz KM, Lash TL, Hill AN, Kammerer JS, Winston CA, Castro KG, Gandhi NR, Navin TR, 2018. Simple estimates for local prevalence of latent tuberculosis infection, United States, 2011–2015. *Emerg Infect Dis* 24: 1930–1933.
- Flanagan BE, Hallisey EJ, Adams E, Lavery A, 2018. Measuring community vulnerability to natural and anthropogenic hazards: the Centers for Disease Control and Prevention's Social Vulnerability Index. *J Environ Health* 80: 34–36.
- Vonnahme LA, Todd J, Puro J, Oakley J, Jones M, Rivera P, Langer AJ, Ayers T, 2020. 1651. Describing the tuberculosis infection cascade of care based on electronic health record data. *Open Forum Infect Dis* 7: S813–S814.

15. Dormann CF et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27–46.
16. Rajaram A, Thomas D, Sallam F, Verma AA, Rawal S, 2020. Accuracy of the preferred language field in the electronic health records of two Canadian hospitals. *Appl Clin Inform* 11: 644–649.
17. Slutsker JS, Trieu L, Crossa A, Ahuja SD, 2018. Using reports of latent tuberculosis infection among young children to identify tuberculosis transmission in New York City, 2006–2012. *Am J Epidemiol* 187: 1303–1310.
18. Tsang CA, Langer AJ, Kammerer JS, Navin TR, 2020. US tuberculosis rates among persons born outside the United States compared with rates in their countries of birth, 2012–2016. *Emerg Infect Dis* 26: 533–540.
19. Kim S, Cohen T, Horsburgh CR, Miller JW, Hill AN, Marks SM, Li R, Kammerer JS, Salomon JA, Menzies NA, 2022. Trends, mechanisms, and racial/ethnic differences of tuberculosis incidence in the US-born population aged 50 years or older in the United States. *Clin Infect Dis* 74: 1594–1603.
20. Bakhsh Y, Readhead A, Flood J, Barry P, 2023. Association of area-based socioeconomic measures with tuberculosis incidence in California. *J Immigr Minor Health* 25: 643–652.
21. Olson NA, Davidow AL, Winston CA, Chen MP, Gazmararian JA, Katz DJ, 2012. A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996–2005. *BMC Public Health* 12: 365.
22. Readhead A, Flood J, Barry P, 2022. Health insurance, health-care utilization and language use among populations who experience risk for tuberculosis, California 2014–2017. *PLoS One* 17: e0268739.
23. Oren E, Koepsell T, Leroux BG, Mayer J, 2012. Area-based socio-economic disadvantage and tuberculosis incidence. *Int J Tuberc Lung Dis* 16: 880–885.